Charlotte Gooskens

Mutual Intelligibility between Closely Related Languages

Language Contact and Bilingualism

Editor
Yaron Matras

Volume 30

Charlotte Gooskens

Mutual Intelligibility between Closely Related Languages

DE GRUYTER
MOUTON

This book has been published open access thanks to the financial support of the Open Access Book Fund of the University of Groningen.

ISBN 978-3-11-113165-8
e-ISBN (PDF) 978-3-11-113469-7
e-ISBN (EPUB) 978-3-11-113548-9
ISSN 2190-698X
DOI https://doi.org/10.1515/9783111134697

Library of Congress Control Number: 2024939654

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.dnb.de.

Cover image: Anette Linnea Rasmus/Fotolia
Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Preface

When I was ten years old, I made my first trip away from Denmark, accompanied by my grandmother, to visit distant relatives in Norway. I vividly remember how overwhelmed I was when I began to play with my Norwegian cousins and their friends. Initially, I felt restricted by the differences between my Danish and their Norwegian, but before long, we overcame these language barriers. We discovered that we could each speak our own language and still understand each other well enough to communicate. I had a wonderful time with my new friends in this beautiful country for the rest of the summer.

This childhood memory has profoundly influenced my scientific career. At first, my curiosity revolved around the communication dynamics among Scandinavians (Danes, Norwegians, and Swedes) when using their native languages. Over time, I broadened my perspective to include communication between speakers of closely related languages outside Scandinavia. Unlike Chinese and English, which are vastly dissimilar, many languages worldwide share genetic roots and enough resemblance for speakers to understand each other to varying degrees, even upon first encounter. This is what is referred to as mutual intelligibility. Examples include the Scandinavian languages or Spanish and Portuguese.

Linguists have been interested in mutual intelligibility between closely related language varieties since the middle of the previous century, and recently, developments of methodologies have taken place in various fields of research that are valuable for research on mutual intelligibility. These developments include advancements in dialectology, sociolinguistics, psycholinguistics, applied linguistics, and neuro-linguistics. Additionally, the availability of personal computers, the internet, and advanced linguistic research tools have contributed to new developments in the field. The findings and ideas of the past and these recent developments make it possible to measure the level of mutual intelligibility between speakers of many language varieties and relate it to linguistic and extra-linguistic factors that may explain intelligibility levels.

For the past two decades I have used these methods to conduct intelligibility research at the University of Groningen as the principal investigator in two large projects supported by the Dutch Research Council (NWO). The first project, *Linguistic determinants of mutual intelligibility in Scandinavia*[1], focused on the mutual intelligibility between Scandinavian and West-Germanic languages. The second project, *Mutual intelligibility of closely related languages in Europe:*

1 Project number 276-75-005, 2006-2011, PI Charlotte Gooskens, other project members Nanna Hilton, Sebastian Kürschner, Jens Moberg, Anja Schüppert, and Renée van Bezooijen.

https://doi.org/10.1515/9783111134697-202

linguistic and non-linguistic determinants[2], the so-called MICReLa project, expanded my scope to include Romance and Slavic languages spoken in Europe. These projects resulted in three doctoral theses (Schüppert 2011; Golubović 2016; Swarte 2016).

Moreover, I have collaborated with research groups investigating intelligibility in other regions worldwide, such as between Finnish and Estonian, Kurdish dialects, languages in Vanuatu, Turkey, Iran, Ethiopia, and Indonesia. Through these collaborations and personal interactions with students and researchers from many different countries, it has become evident to me that there is a need to consolidate our current understanding of how to measure the intelligibility of closely related languages and quantify relevant factors for interpreting the results.

The aim of this book is to provide an overview of existing knowledge regarding mutual intelligibility between closely related languages. This field of research has a long tradition and has served various practical and theoretical purposes. For instance, intelligibility research may help resolve issues that concern language planning and policies. It can inform the development of language learning materials, which can be tailored to make use of our knowledge of factors that determine intelligibility across specific language pairs – for instance, to create better materials for Norwegians learning German or Czech individuals learning Polish. It can also contribute more generally to improving cross-cultural communication: when we understand our differences, we are better equipped to overcome them. Intelligibility research can, therefore, help improve mutual intelligibility. Researchers have also conducted intelligibility research to address fundamental questions such as: How different can two languages be before they cease to be mutually intelligible? Is there a threshold for intelligibility breakdown? How different should two language varieties be before they can be considered separate languages rather than dialects of the same language? What are the most significant linguistic and extra-linguistic factors influencing intelligibility?

The book illustrates the diversity of current research in the field and identifies gaps in our knowledge that future investigators may fill. It presents methods for investigating intelligibility as well as case studies, combining theoretical and applied approaches to the field. The focus is on quantitative measurements. It discusses how to measure intelligibility in experimental settings and how to quantify factors that are relevant for explaining the results of intelligibility testing. Most of the examples in the book involve spoken language, but there are also references

2 Project number 360-70-430, 2011-2016, PI Charlotte Gooskens, co-applicant Vincent van Heuven, other project members Jelena Golubović, Wilbert Heeringa, Anja Schüppert, Femke Swarte, Renée van Bezooijen, and Stefanie Voigt. This project is referred to as the MICReLa project throughout the book.

to research on written language. Closely related languages are defined very broadly, encompassing language varieties from the same language family as well as dialects of a single language. However, foreign accents are outside the scope of the book, even though some of the methods discussed have been developed for research on foreign accents.

The reader is assumed to have a basic knowledge of linguistics, but hopefully, a diverse audience of students and researchers will find valuable information in the book. Individuals working in the area of language policy and language planning may also benefit from more knowledge of the intelligibility research and results presented.

This book would not have come into existence without my scientific "parents", Renée van Bezooijen and Vincent van Heuven. Their invaluable guidance and mentorship have shaped my intellectual journey, and I am profoundly grateful for the knowledge I have gained from them. Vincent supervised my Master's thesis and has been a continuous source of academic support since then. His critical feedback on this book played a pivotal role in refining its content, correcting numerous mistakes and inconsistencies before publication. Any remaining imperfections are, of course, solely my responsibility. Renée was my PhD supervisor, and after that, we worked together for many years. Her profound knowledge and infectious enthusiasm were extremely inspiring, and I think back at our time of cooperation as the best period in my academic career. A special thanks goes to John Nerbonne, who has shown unwavering interest in my research right from the beginning, providing invaluable support and advice.

I am also indebted to all PhD students, postdocs, student assistants, and colleagues I have been working with for the past two decades. In particular, I would like to thank Jelena Golubović, Wilbert Heeringa, Nanna Hilton, Sebastian Kürschner, Jens Moberg, Anja Schüppert, Femke Swarte, and Stefanie Voigt, integral to the success of the two large projects mentioned earlier. I continue to enjoy fruitful collaborations with many of them. Jan D. Ten Thije deserves recognition for establishing the LaRa group, providing a platform for discussing receptive multilingualism at the Dutch national level. I am also indebted to researchers worldwide who expressed interest in my work and made me aware of the need for a synthesis of research on mutual intelligibility. I collaborated with several of them, including Ahmet Kesmez, Bilgit Sağlam, and Fatih Özek on Kurdish and Turkic languages, Hanna-Ilona Härmävaara on Finnish and Estonian, Gerard Doetjes and Karin Beijering on Scandinavian languages and dialects, Remco Knooihuizen on Faeroese and Icelandic, Tekabe Legesse Feleke on Ethiopian languages, Stefan Bulatović on Croatian, and Cindy Schneider on languages in Vanuatu. Cindy also generously devoted time to reading drafts of the book and provided extremely valuable feedback.

The initial draft of this book took shape during my stay at the Polish Institute for Advanced Studies (PIAST) in Warsaw, where I held a fellowship from 2020 to 2021 amidst The COVID-19 pandemic. This opportunity was fundamental for the realization of this project.

I also express my gratitude to my parents, who gave me the freedom to pursue my academic interests, and to my children – Marijn, who proofread the first draft, and Jeppe and Ida, who never stopped encouraging me. Finally, my deepest gratitude goes to my husband, Mark, whose enduring interest in this project and critical, constructive feedback on each chapter's first draft have been indispensable. He also meticulously checked all references in the book. Most importantly, he made my life easy and enjoyable in so many ways.

Contents

Part II: **Further analyses**

Part III: **Applications and conclusions**

List of figures

https://doi.org/10.1515/9783111134697-204

List of tables

https://doi.org/10.1515/9783111134697-205

Chapter 1
Introduction

According to Ethnologue's list of living languages, there are currently 7,168 languages spoken worldwide (Eberhard, Simons, and Fennig 2023). In today's context of multilingualism and increasing mobility, speakers with different native language backgrounds are often faced with some of these languages in writing and speaking, and if they meet, they will need to find a way to communicate. One way to cross linguistic borders is for speakers to learn a second or foreign language. However, language acquisition is hard work. It requires mastering grammatical rules, memorizing word lists, and practicing pronunciation. It is only possible for an individual to learn a limited number of languages, and many people do not know foreign languages well enough to understand a message or to communicate with native speakers of these languages. To limit the number of languages to be learned, the solution is often to use a lingua franca, a language that makes it possible for speakers who do not share a first language to exchange information and communicate. Various lingua francas are used in different parts of the world, but English has become the global lingua franca of the 21st century. However, results of surveys (e.g., EF EPI 2023) show that many people have difficulties understanding and speaking English.

In many situations, speakers of different languages can take advantage of the fact that their languages are so closely related that they are mutually intelligible. In such situations, speakers can often communicate rather successfully using their own (individual) languages. This kind of communication is sometimes referred to as receptive multilingualism. To introduce this manner of communication more widely and improve mutual intelligibility, we need to be able to establish the degree of intelligibility between closely related languages and identify linguistic and extra-linguistic factors that play a role in explaining the level of intelligibility. This will enable us to decide in which situations it makes sense to introduce receptive multilingualism for communication or whether the use of a lingua franca or a foreign language should be promoted. It can also provide us with information about how to improve communication. If mutual intelligibility is insufficient for successful communication, understanding the factors that affect the level of intelligibility can assist researchers in identifying the cause of difficulties. This knowledge can be utilized to design educational programs aimed at resolving these issues.

There are many other reasons to undertake research on mutual intelligibility. Early investigations were oriented towards a formal paradigm, and researchers were mainly interested in the genetic relationships between languages and dialects. Examples are American structuralists (e.g., Voegelin and Harris 1951; Hick-

https://doi.org/10.1515/9783111134697-001

erson, Turner, and Hickerson 1952; Pierce 1952) who tried to establish mutual intelligibility among related indigenous American languages around the mid-20th century. Until the 1950s, the assumption prevailed that mutual intelligibility between two languages would be symmetrical (or reciprocal) and that structural differences (or distances) between two languages would also be symmetrical. Given symmetrical differences between two related languages, A and B, language A should be as intelligible to native listeners of language B as the other way around. Deviations from symmetry were attributed to either measurement error or to differences in extra-linguistic factors, such as prior exposure or attitudes to the other language. Casad (1974: 73) points out that reciprocity was part of the definition of mutual intelligibility in the work by the American structuralists in the early 1950s.

With the introduction of new linguistic disciplines, such as sociolinguistics, pragma-linguistics, and language acquisition, starting in the 1960s, the focus shifted to a more functional paradigm (ten Thije 2018). Since then, many investigations on mutual intelligibility have been conducted with diverse objectives, such as resolving issues concerning language planning and policies at the national and international levels. Intelligibility research may help policymakers deal with questions such as whether television programs in a neighboring language should be dubbed or subtitled, how to improve communication between speakers of different languages, or whether two language varieties should be considered separate languages or dialects of the same language. If smaller languages are to survive, it is crucial to understand the mechanisms involved in using one's own language for communication with speakers of other languages. Detailed knowledge about intelligibility can also play a role in sociolinguistic studies. It is, for example, valuable to understand how intelligibility may influence attitudes towards language varieties and their speakers. Intelligibility research has also been used to validate dialectometric methods for measuring distances between dialects. Finally, intelligibility research provides us with fundamental knowledge about the human language faculty and its limitations. The findings aid in enhancing our comprehension of the resilience of the human language processing mechanism. They enable us to determine the extent to which a language can differ from the native language before it becomes incomprehensible to the listener, and what linguistic and extra-linguistics factors are the best predictors of intelligibility. Chapter 8 presents a more extensive discussion of reasons for doing intelligibility research and discusses possible theoretical and practical applications of the results.

As mentioned, many researchers from around the world have been interested in the mutual intelligibility between closely related languages since the early 1950s. The methodologies developed in early projects are still being used for various purposes, such as in the standardization of language varieties and the development of unified writing systems that accommodate several closely related languages (Casad 1974; Nahhas 2007). This book builds on the ideas and findings of this early work (e.g., Casad 1974; Simons 1979). Many of the ideas remained rather theoretical for a long time due to the lack of tools for collecting and analyzing data. In traditional dialectology, before the introduction of personal computers, researchers were limited in the amount of data they could process and the kind of measurements they could carry out. Later, researchers extended their range of methods for measuring intelligibility, depending on their research interests, the circumstances in the research area, and the access to tools for quantifying the level of intelligibility and its determinants. Recently, intelligibility research has gained new momentum due to developments in different scientific areas. In particular, the introduction of computers and software has made collecting large amounts of data easier, and refined dialectometric algorithms for analyzing large data sets have been developed to measure linguistic distances between dialects. Intelligibility research has also benefited immensely from the refined measuring methods developed in psycholinguistics, second-language acquisition, and speech pathology. These algorithms and methods have resulted in significant advancements in intelligibility research.

Most of the examples in this book are concerned with the intelligibility of spoken language, but the methods can often be applied to written language as well. However, it is essential to be conscious of the fact that the experience of a listener differs from that of a reader (see Section 2.1.1.1). When listening to spoken language, the listener will often get only one attempt to process the input and the processing time is limited, whereas in the case of written language there is mostly no time limit and the reader can reread the text and search for additional cues in the context if necessary. On the other hand, in the case of spoken intelligibility, there may more often be interaction between speaker and listener. This allows them to negotiate meaning, i.e. to reach mutual understanding by means of various strategies, such as repeating, clarification, rephrasing, and checking whether the message has come across in the intended way. Written words are segmented by spacing while there are no consistent gaps between words in spoken language. Many languages have standardized spelling systems, which results in uniform representations of the letters in words, but the font of printed letters and the handwriting of individual persons can be very different. Likewise, the pronunciation of the sounds in spoken words can be highly variable.

As will become clear throughout the book, researchers have looked at intelligibility from many perspectives. Sometimes the main interest of the researcher is to establish the level of intelligibility, but it can also be a central aim to explain intelligibility results. Simons (1979) introduced a distinction between linguistic and social factors that may explain the level of intelligibility. Not all speakers of the same language A may understand a language B equally well. The level of understanding between two individuals with different native language backgrounds may largely depend on extra-linguistic determinants, i.e., factors outside of the language itself, such as speaker and listener competencies, characteristics and activities. In this book we refer to such factors as extra-linguistic determinants and we will discuss how to quantify them in Chapter 4. However, characteristics of the languages of the speaker and the listener also determine the level of intelligibility. Linguistic determinants can be found at various linguistic levels, most importantly at the levels of phonetics (the sounds in speech, e.g., vowels, consonants), morphology (the internal structure of words), lexicon (the vocabulary of a language), syntax (how words combine to form phrases and sentences), semantics (the meanings of words and phrases in language), and pragmatics (the study of how context contributes to meaning). In addition, differences between languages can be found at the paralinguistic level (phenomena such as pitch, volume, and speech rate). Measurements of linguistic and para-linguistic determinants are discussed in Chapter 5. However, there is no section dedicated to pragmatics since little experimental research has been conducted on the relationship between pragmatics and intelligibility of closely related languages. Pragmatics is touched upon at different places in the book when discussing discourse and strategies that listeners and speakers can use to reach mutual understanding. Figure 1.1 presents a schematic overview of the linguistic and extra-linguistic determinants of intelligibility that are discussed in the book.

Ideally, we would like to be able to predict and explain how well speakers of two closely related languages can understand each other on the basis of combinations of measurements of relevant linguistic and extra-linguistic determinants. Various determinants may overlap. For example, if two languages are lexically very similar, their pronunciations are often also similar. Heeringa, Gooskens, and van Heuven (2023) measured lexical, phonetic and syntactic distances between 35 pairs of closely related European languages. They found that the linguistic dimensions are generally significantly correlated. This means that if two languages grow apart from one another in their historical development, changes will not take place at only one linguistic level, but innovations will occur at all linguistic levels. However, the correlations are often weak and there is no a priori way of weighing the different determinants of intelligibility. The relative contribution of the various determinants and their interaction with each other make it complicated to develop a model of intelligibility. To establish statistical relations between the intelligibility results and their determinants, it is necessary to quantify both the level of intelligibility and the determinants that explain it. Most of the research discussed in this book therefore takes a quantitative approach to the study of intelligibility.

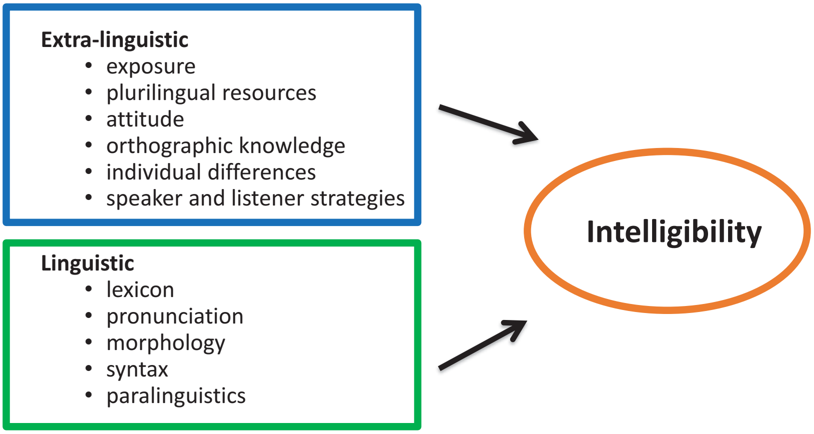

Figure 1.1: Linguistic and extra-linguistic determinants of intelligibility discussed in the book.

1.1 Concepts and definitions

There are a number of concepts that are central to this book. When referring to the various aspects of intelligibility research, scholars often define and use the same terms differently depending on their field of interest. Below, the terms and the definitions that are used in this book are presented.

Often, a distinction is made between intelligibility, comprehensibility, and interpretability. Smith and Nelson (1985) rely on different levels of understanding to differentiate the three. They conceptualize intelligibility as pertaining to the ability to identify the linguistic units in the stream of sounds (e.g., word and utterance recognition), comprehensibility as pertaining to the understanding of the overall meaning of words or utterances, and interpretability to the meaning behind the word or utterance, i.e. the speaker's intention. Intelligibility, comprehensibility and interpretability thus address three different stages in spoken communication, and each can be measured using various kinds of tests. Other researchers use the term intelligibility for the degree to which a speaker can be understood using functional tests that measure how well a listener *actually* understands the target language, and comprehensibility for the listener's opinion as to how well a speaker (or utterance) can be understood (Munro and Derwing 1995a; Munro, Derwing, and Morton 2006). Since the various terms are often used interchangeably and all ideas are concerned with the successful reception of a speaker's message by a listener, it is not necessary to distinguish between the different terms for our purpose, as suggested by Lindemann and Subtirelu (2013). One term, **intelligibility**, is therefore used in most of the book to refer to the property of a speaker or speech that ensures that it is “intelligible”. The listener understands the speaker (or the utterance). The process that the hearer goes through is called **understanding** (and includes speech recognition, comprehension, and interpretation).

In the literature, many examples of **asymmetric intelligibility** (or non-reciprocal intelligibility, see Kluge 2007) have been observed. Speakers of language A may have more difficulty understanding language B than the other way around. As mentioned above, until the 1950s it was commonly assumed that linguistic distances between pairs of languages are symmetric. Therefore mutual intelligibility between language pairs was also assumed to be symmetric, unless there are differences in extra-linguistic factors, such as attitude or previous exposure to the other language, or a deficiency in the measurements. More recently, however, many pairs of languages have been identified in which the intelligibility is asymmetrical, and where this asymmetry cannot be attributed to differences in exposure or other social factors. There is evidence that linguistic characteristics of the language varieties may also contribute to the asymmetry. The mechanisms behind asymmetric intelligibility are discussed in Chapter 6.

When talking about the separate directions AB or BA, the term **cross-language intelligibility** is used. Since intelligibility can be asymmetric, the cross-language intelligibility of language B to speakers of language A may differ from the cross-language intelligibility of language A to speakers of language B. **Mutual intelligibility** is used when referring to both directions AB and BA. Mutual intelligibility can be expressed as the mean of the two directions, AB and BA, in which case deviations from perfect symmetry in intelligibility are averaged out. If speakers of language A have a score of 40% when their comprehension of language B is tested and speakers of language B a score of 60% when their comprehension of language A is tested, the mutual intelligibility between language A and language B is 50%. It should be noted that in the case of strong asymmetry, such a mean may not be the best measure of how well the speakers of two languages understand each other. It seems logical to assume that there is a weak basis for communication if one of the groups of speakers has a low level of comprehension, even if the other group has a high level of comprehension. A possibility is to use the lowest of the two measurements to express the level of mutual intelligibility (see Simons 1979).

As mentioned in the previous section, a level of mutual intelligibility sufficient to exchange information can often be achieved if two languages are closely related. Speakers can then communicate using what is often referred to as **receptive multilingualism**.[3] In such a situation, the speakers are able to communicate

3 Other frequently used terminology that covers approximately the same concepts are plurilingual communication, semi-communication, intercompréhension, non-reciprocal bilingualism, rather successfully while both are speaking their own language. This mode of communication is discussed in detail in Section 8.5.

When investigating intelligibility, we are dealing with both listeners and speakers, each with a specific native language background. In intelligibility testing situations, listeners with language background A will listen to speech recorded by a speaker of language B, and when investigating mutual intelligibility listeners with language background B will also listen to sentences recorded by a speaker of language A. The language of the speaker is referred to as the **target language** and the language of the listener as the **listener language.**

A distinction can be made between inherent and acquired intelligibility (Simons 1979; Kluge 2007). The two kinds of intelligibility are distinguished by the existence of an acquisition process due to exposure or formal instruction in the target language and by the possibility to understand the target languages on the basis of the native language. **Inherent intelligibility** relies on similarities between the target language and the listener language, i.e. the linguistic determinants shown in the green square in Figure 1.1. These similarities follow from the close relationship between the two languages and they therefore often reflect traditional genealogic taxonomies among languages as presented in language trees. Examples of language pairs with mutual inherent intelligibility are Danish-Swedish, Czech-Slovak and Portuguese-Spanish. **Acquired intelligibility** presupposes some acquired knowledge and typically involves less closely related (or less similar) languages, such as Dutch-German, Polish-Russian, and French-Spanish. It could even include unrelated languages, such as Turkish-Dutch or Chinese-English, if the speakers have sufficient receptive knowledge of each other's languages. Other terms used for intelligibility as determined by extra-linguistic factors are social or contact-based intelligibility (Simons 1979).

Differences between inherent and acquired intelligibility are gradual rather than dichotomous (Sherkina-Lieber 2020). Swedish and Danish are often mentioned as languages that are inherently mutually intelligible. Generally, Swedes and Danes understand each other almost purely on the basis of the similarity of their languages. In a first encounter, Swedes may have problems understanding Danish. Still, intelligibility can be improved by some level of exposure to the target language that leads to an acquisition process, such as when the listener is exposed to the target language via television, books, newspapers, the internet or

---

non-accommodating bilingualism, mutual passive bilingualism, receptive bilingualism, polyglot dialogue, and lingua receptiva. The choice of terminology mainly depends on the focus, approach, populations, and research paradigm (formal or functional) being used. For discussions of terminology, see for example, Bahtina and ten Thije (2012), ten Thije (2018), and Sherkina-Lieber (2020).

personal communication. After some exposure to Danish, Swedes often get used to the sounds of the Danish language and the intelligibility problems decrease. Dutch and German, on the other hand, are normally regarded as examples of acquired mutual intelligibility. Even though their languages belong to the same branch of the Germanic language family (West Germanic), Dutch and German speakers would only reach a low level of mutual intelligibly in a first encounter. However, since the languages are fairly similar, they can benefit from the fact that German and Dutch share a large part of the lexicon due to their common ancestor. Therefore, some of the processes of understanding resemble those of inherent intelligibility. In this book, the term **general intelligibility** will be used to refer to intelligibility that may include various degrees of inherent and acquired intelligibility.

Furthermore, a subtype of inherent intelligibility referred to as **mediated intelligibility** can be distinguished. This refers to situations where listeners use the similarity to the target language of a **bridge language**, i.e. a third language that they know, to understand the language of the speakers. An example is Estonian speakers with knowledge of Russian, who can understand Ukrainian because it is closely related to Russian (see Branets, Bahtina, and Verschik 2020).

When we talk about **closely related languages**, we refer to language varieties that belong to the same language family and are so similar that they are mutually intelligible to some extent. The degree of mutual intelligibility can range widely from very low to very high. **Neighboring language** is another term that is sometimes used with the same meaning when we talk about languages that are spoken in close geographic proximity to each other. This term is often used in the Scandinavian context when referring to the three closely related mainland languages (Danish, Norwegian, and Swedish).

Linguists often use the words **variety** (or lect) to refer to a specific form of a language. These terms may include languages, dialects, registers, styles, sociolects, or other forms of language, as well as standard varieties and even personal varieties (ideolects). In the context of this book, it is not necessary to distinguish between these different kinds of language varieties. The mechanisms behind mutual intelligibility are generally the same in all these contexts. Therefore, when talking about **languages** throughout the book, the point to be made mostly concerns languages, as well as dialects and other language varieties. In Section 8.2, criteria for distinguishing between dialects and languages in the traditional sense are discussed in detail.

When listening to a closely related language, listeners are often able to recognize **cognates** because they are similar to words in their native language. Unless they have had some exposure to the language variety, they cannot understand **non-cognates** because they have no related equivalent in their language. Cognates can be defined as historically related words in the vocabularies of two related languages, i.e. word pairs that can be traced back to the same word in the ancestor of the two languages (**inherited cognates**). An example of a cognate pair is French *père* Spanish *padre* ‘father’, both with the same Latin origin *pater*. In an intelligibility context, the focus may be shifted to a more synchronic definition of cognates, including instances of the same borrowings from another language. **Cognate loanwords** are then also regarded as cognates. An example of cognate loanwords are English *café*, Swedish *kafé* (both borrowed from French *café*). In the case of intelligibility research, this broader definition of cognates makes sense, since loanwords are often easily recognizable to a listener and the aim is to investigate to what extent lexical overlap makes it easier to understand a target language. In principle, it only matters if words are related or not, regardless of their origin. In this book, this broader definition of cognates and non-cognates is therefore used.

1.2 Structure of this book

This book is organized into three parts. Part I (Chapters 2 to 5) explains how to measure the level of intelligibility and how to quantify linguistic and extra-linguistic determinants of intelligibility (see Figure 1.1). Part II (Chapters 6 and 7) deals in further depth with specific questions that can be addressed once we have the tools to quantify intelligibility and its determinants. Finally, Part III discusses practical and theoretical reasons for studying mutual intelligibility, desiderata for future research, and draws some overall conclusions.

Chapter 2 provides an overview of various **methods** for measuring intelligibility at different linguistic levels (word level, sentence level, text level, discourse level). It discusses a number of considerations that should be made when choosing a method for carrying out an intelligibility investigation and considers advantages and disadvantages. Each method is exemplified with one or more examples.

Chapter 3 presents the **MICReLa project** as an example of an intelligibility project. We present details of the setup and some of the material used in this project on the mutual intelligibility of 16 closely related languages in Europe. Also, some general results are presented. In Appendix A, the texts used for the intelligibility experiment are provided, and in Appendix B to Appendix D, full tables with data from the project are presented. Many examples throughout the book are taken from the MICReLa project.

Chapter 4 provides an overview of **extra-linguistic determinants** of intelligibility and discuss how they can be quantified. It also presents examples of studies that have found a link between intelligibility and each of the extra-linguistic determinants.

Chapter 5 presents methods for quantifying **linguistic determinants** of intelligibility at various linguistic levels (lexical, phonetic, morphological, syntactic, paralinguistic). It also provides examples of investigations that have shown significant correlations between intelligibility measurements and linguistic measurements.

Chapter 6 presents various investigations that have looked into the origin of **asymmetric intelligibility**. Such factors can be both linguistic and extra-linguistic. By understanding the reasons for asymmetric intelligibility, we can get further insight into the factors determining the level of intelligibility.

Chapter 7 focusses on research that has been carried out to make advancements towards a **model of mutual intelligibility**, including an analysis of data from the MICReLa project.

Chapter 8 discusses various **theoretical and practical applications** of intelligibility research.

Chapter 9 concludes with the identification of **gaps in our knowledge** about processes and phenomena in the area of mutual intelligibility and points to directions for **further investigations**.

Part I: **Measurements**

Chapter 2
Methods for measuring intelligibility

In investigations set up to measure the level of intelligibility between closely related languages, several different methods have been used to collect data. This chapter describes several of them and discusses their advantages and disadvantages. Section 2.1 discusses questions that should be considered when choosing a method for quantifying the level of intelligibility and carrying out an intelligibility investigation. At the end of Section 2.1, an overview of the considerations is presented. Section 2.2 provides an overview of methods for measuring the level of intelligibility. Each method is exemplified with one or more examples. In Section 2.3, the results of different methods of measuring intelligibility are compared. These comparative analyses provide insights into the significance of selecting a particular method.

2.1 Methodological considerations

Like any other scientific research project, an intelligibility project begins with formulating a research question (and hypotheses) to provide the project with a clear focus (see, e.g., Sunderland 2018). The research question should guide the further process. However, when setting up an intelligibility project, numerous methodological and practical considerations need to be made. These considerations may influence the whole research process, so clearly and accurately defining the research question can become an iterative process. For instance, a researcher may intend to measure the general intelligibility of two languages but discover that it is only feasible to test a limited group of listeners that is not entirely representative of all speakers of a language. This could mean that the research questions would have to be (slightly) reformulated.

Methodological considerations applicable to intelligibility testing share many similarities with those pertinent to experimental investigations in sociolinguistics, psycholinguistics, and language acquisition. The choice of a testing method depends on several factors. There may be practical limitations and hurdles to be faced during the development and administration of an experiment. A limited amount of time or funding may be available, so the time needed to develop the experiment or to process and analyze the data needs to be taken into account. Other factors that may guide the choice of testing method are more fundamental because they depend on the purpose of the investigation (see Chapter 8).

https://doi.org/10.1515/9783111134697-002

It is imperative to plan an investigation well in advance and in detail to get a realistic picture of the amount and nature of the work that needs to be done at every step. Unless the object of investigation is their own native language variety, researchers may be dependent on support from native speakers when preparing stimulus material, for translating words or texts or making recordings. These native speakers should be chosen with care to ensure that they represent the target language optimally. Other phases in the investigation may also need planning. For instance, locating a sufficient number of willing participants for the experiment might require considerable effort. Ethical approval should be given before the investigation starts. This aims to protect both the researcher and the participants in the research by ensuring they have their dignity, rights, safety, and welfare respected. In this chapter, the processing and analysis of the data are not discussed in detail. Still, this part of the investigation should also be considered in advance to make sure that the data will allow for a reliable analysis.

2.1.1 Test material

The choice of test material for an intelligibility experiment depends on the aim of the investigation. It can vary along various dimensions: mode (written or spoken), style (spontaneous or read-aloud, formal or informal, monologues or dialogues), level of analysis (isolated words, sentences, or texts), recordings (audio, video), and speaker characteristics and demography (social background, numbers, voice quality, gender, age, etc.). If the intelligibility of more languages is to be compared, these factors should be kept as consistent as possible across languages when collecting material for tests, unless they are one of the variables under investigation. We will discuss this in further detail below.

2.1.1.1 Mode

The mode is the medium of communication, and it is fundamentally divided into speech and writing. It depends on the purpose of the investigation, whether tests will be carried out in written or spoken mode. Most of this book deals with the intelligibility of spoken language, and this chapter focuses on test techniques that are developed for spoken language. However, it is important to be aware of the differences between the two modes and how they may influence each other. While the spoken mode is coded in sounds, the written mode is coded in graphic symbols. Unlike with the speech mode, there is the opportunity to revise and correct in the written form. Writing is long-lasting, while speech is ephemeral. Listeners can ask speakers to repeat, which may be simulated in a test situation by presenting spoken stimuli twice (see Section 2.1.3.2), but other than that, they cannot check back that a word or sentence has been correctly identified. In addition, listeners usually cannot control the speed at which listening takes place, whereas readers can modify reading speed according to their needs.

Each mode may range from spontaneous (a casual conversation or a scribbled written note) to planned (a prepared talk or a formal essay). Note also that the introduction of various means of communication via social media has blurred the traditional distinctions between oral and written discourse in some contexts (Sindoni 2013). Iwasaki and Oliver (2003: 63) point out that online chat is similar to a face-to-face interaction because all participants continually take turns to exchange messages and are not given much time to review their written messages.

It is often easier to understand a closely related language in the written form than in the spoken form, at least if the writing system of the target languages is the same as that of the participant. Little research has been conducted to compare the ease of understanding written and spoken language. To compare the ease of comprehension of the two modes, the same tests and textual materials should be used. Gooskens and van Heuven (2017) included written as well as spoken versions of three tests in an investigation, using the same tests and the same textual materials in both modes in 70 combinations of closely related Germanic, Romance, and Slavic languages (see Chapter 3). This allowed them to directly compare mutual intelligibility between the two modes. The results showed that the written modality was easier to understand than the spoken modality in 65 of the language combinations. The written form of a closely related language may be easier to associate with its counterpart in the native language. Especially in the case of well-established languages with a long history of writing, the written language often represents an earlier stage of the language, where the two languages had diverged less from their original common form than in the spoken form (Gooskens and Doetjes 2009; Schüppert 2011, see also Section 4.4). For example, it may not be too difficult for Dutch listeners to recognize the written form of Danish *brød* [b̥ʁœɣˀ] as a cognate of Dutch *brood* [bʁoːt] 'bread', but when listening to the Danish pronunciation of the word it may no longer be possible for them to recognize the word. Among elderly participants, the advantage of printed forms may be attributed to a reduction in processing capacity for spoken language (Vanhove 2014). As mentioned above, among both younger and older adults, it may be an advantage that written text can be reread several times. In a natural situation, spoken text is usually presented only once and is always short-lived. This effect may be stronger for elderly participants.

However, the opposite can also be the case. Some closely related languages use the same alphabet but are spelled with such divergent orthographic conventions that it is difficult to read texts that are intelligible in the spoken form. For instance, it is difficult for Dutch readers to understand written Frisian because the spelling conventions are quite different from those used in Dutch and other languages. Still, Dutch listeners often understand spoken Frisian fairly well (van Bezooijen and van den Berg 1999; van Bezooijen and van den Berg 2000; Gooskens 2007). Dissimilar orthographic and phonological systems may prevent learners from discovering cognates. Helms-Park and Dronjic (2016: 87) illustrated this with Polish-English cognates such as *egzystencja* [ɛgzɨ'st̪ɛnct͜sja] 'existence' and *menedżer* [mɛ'nɛd͜ʐɛr] 'manager' that are presumably easy to understand in the spoken form for a listener who knows English but may be challenging to recognize in the written form.

Sometimes, there is an interaction between the written and the spoken mode. When reading a text in a non-native language, readers often self-pronounce the written words. Vanhove and Berthele (2015) describe how the assumed phonetic similarity to words in their native language can help readers achieve cognate guessing success in the written mode. To illustrate, a Dutch native speaker may have this experience when reading German. The German spelling may be rather deviant from Dutch (e.g., Dutch *brood* vs. German *Brot*, 'bread'). Still, if Dutch readers read aloud the German word to themselves, they may more easily discover the correspondence between the Dutch and the German forms because the Dutch *d* at the end of words is pronounced as a /t/ and because long vowels are not always reflected in the orthographic form (the plural of Dutch *brood* is *broden* while both are pronounced with a long vowel. In a think-aloud task, Möller and Zeevaert (2015) found that native German speakers used intuitions about the pronunciation of cognates in Germanic languages (Dutch, Frisian, Danish, Norwegian, Swedish, Icelandic, Luxembourgish, Low German) when asked to recognize them in the written mode.

At the same time, knowledge about the written form may help the listener to understand spoken words in the target languages. Schüppert et al. (2022) showed that speakers of Danish can use their orthographic knowledge of Danish to decode spoken Swedish because spoken Swedish is close to written Danish, while the pronunciation of words in these two languages is sometimes quite different (see examples in Section 4.4).

When developing an experiment, the researcher must realize that the written and spoken modes may influence each other. Written instructions, questions, and answer alternatives may influence the results of a test set up to measure spoken intelligibility because the participants do not need to rely on spoken information only. For example, the list of words to be filled in the gaps during a spoken cloze test (see Section 2.2.3.6) is usually presented as a written list. This means that the written mode may have some influence on the results of the spoken intelligibility measurements. In other tests, such as the picture-pointing task (Section 2.2.1.2), the sentence verification task (Section 2.2.2.4), picture-to-story matching (Section 2.2.3.5), and eye tracking experiments (Section 2.2.5), the written mode is less likely to influence the spoken mode.

Many experiments are set up in such a way that participants must be able to read and write to read the instructions, fill in questionnaires, and give written responses to stimuli. However, in cases where the participants cannot read and write (young children and illiterate adults), it is necessary to think of alternative ways of collecting material by recording spoken responses or developing experiments where no reading and writing is involved.

2.1.1.2 Style

If the aim of the investigation is to compare general intelligibility of languages, the material from the different target languages must be matched as closely as possible. One way to control the material is to use translations of the same words, sentences or texts in all the target languages.

It may be necessary to use read speech to control the material maximally, for instance to ensure that differences in word choice or grammatical constructions do not affect intelligibility. Different speakers are unlikely to produce the exact same sentences and texts in spontaneous speech, and as such, it would be necessary to collect a large amount of speech material to gather the same words or concepts from several speakers. When using translations, a text in one of the target languages is often translated into the other target languages. However, there is a risk that the translators may adhere too strictly to the source text when selecting words and expressions for translations, resulting in translations that do not accurately reflect the target languages. To prevent one of the target languages from being given a special status, one solution is to employ translations from a source language that is not one of the target languages or, alternatively, to use source texts from each of the target languages and have them translated into each of the other target languages. By doing so, common words and structures are more likely to be equally represented in all target languages.

It may be preferable to use spontaneous speech since, compared to read speech, this simulates a more natural, real-life situation, making it easier to generalize the results to real-world settings (ecological validity). Pronunciations in spontaneous speech will likely be closer to pronunciations in everyday communication, where speakers often reduce and assimilate sounds. Read speech will usually be closer to orthography and standard variants because speakers tend to pronounce words as they are written. However, since it is challenging to control spontaneous speech, a good compromise may be to use recordings of semi-spontaneous speech that demands speech production in a controlled setting with a predetermined subject of conversation, such as a picture description task (see example in Figure 2.1).

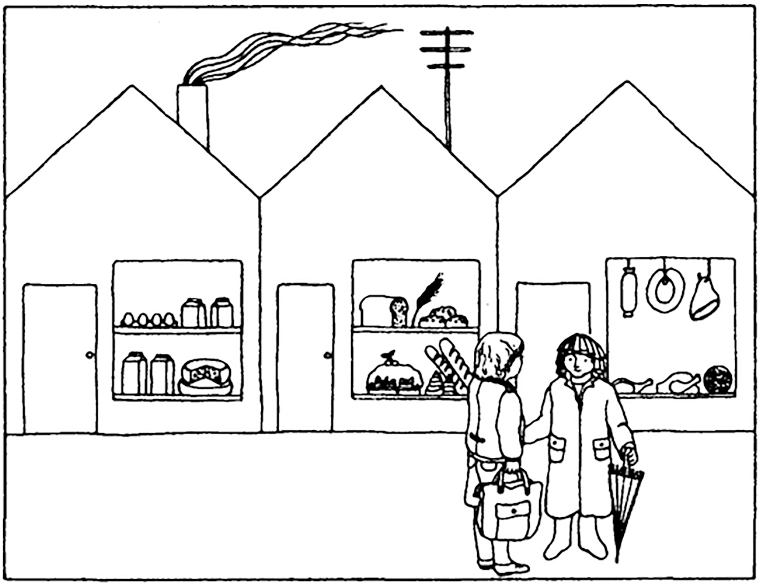

Figure 2.1: A drawing ("the street") used to collect semi-spontaneous speech. Speakers are asked to describe the picture. Source: van Bezooijen and van den Berg (1999: 4).

To elicit content-controlled discourse, Swerts and Collier (1992) had participants describe a network of geometrical figures differing in shape and color and connected by horizontal and vertical lines (see Figure 2.2).[4]

Bulatović, Schüppert, and Gooskens (2019) had speakers watch videos carefully and then describe the events from the beginning to the end. During the recording process, the speakers were presented with ten screenshots of the specific video to help them remember as many details as possible, and to ensure that they did not divert from the storyline too much while narrating the plot. This procedure resulted in recordings in different language varieties that were so similar in content that it was possible to formulate the same questions about the texts. Recordings of dialogues can be elicited with semi-structured games, such as the map tasks, where pairs of speakers are provided with maps, and one speaker gives directions to the other (see Section 2.2.4.2). It should be noted that even if the contents of the recordings of semi-spontaneous speech are similar, there may still be differences in how the information is structured. This can affect intelligibility, not

4 This network was originally developed by Levelt (1989) to illustrate his so-called lineralization model. It shows how speakers order information when they have to express a multi-dimensional structure.

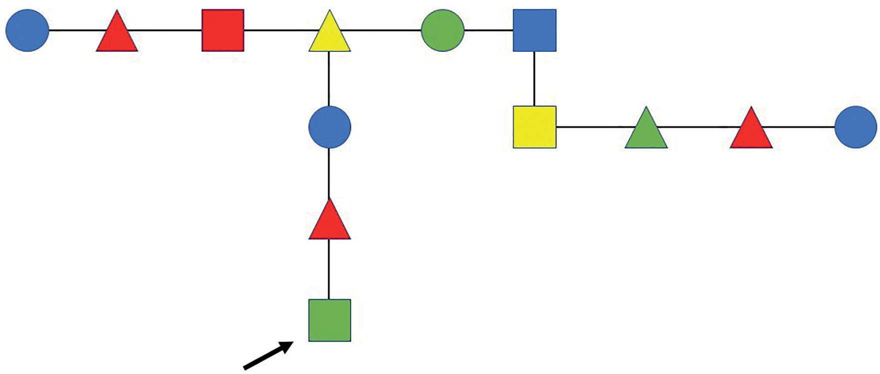

Figure 2.2: A spatial grid-like network consisting of geometrical figures differing in shape and color. Speakers are instructed to describe the network in such a way that a listener can reconstruct it based on the description. Adapted from Swerts and Collier (1992: 465).

because of word recognition or grammatical understanding but because of the variable processing costs of information.

2.1.1.3 Level of analysis

Methods for investigating intelligibility can be characterized by the level of analysis (word, sentence, or text). If the researcher is interested in an overall indication of the level of intelligibility of a whole language (higher-order intelligibility), methods at the level of entire texts are adequate. Investigations at the text level are often more realistic or natural (ecologically valid) than investigations at lower levels, such as sentences or words (lower order intelligibility), since we usually encounter more than only isolated words or sentences when we listen to a language.

The fragments for an intelligibility test are often selected to represent the language as a whole. If the sample is large enough, such as a longer text, it is often assumed to be a representative sample of the target language. However, the text must be selected with care. There are ways to control the complexity of a text, for instance, by counting or quantifying sentence length, number of syllables, word length, word frequency, etc. See Jensen (2009) for a discussion of indicators for text complexity. The topic of a text (daily life, science, society, technique, politics, etc.) should be selected in such a way that listeners will not be able to answer questions about the text based on their knowledge of the world but will need to understand the text itself. It is also important to be aware that one unintelligible word may result in lower intelligibility of the whole text if it is central to understanding the text. To illustrate, one of the texts used to test the mutual intelligibility of Scandinavian languages by Delsing and Lundin Åkesson (2005) was about frogs. The word for 'frog' is very different in Danish (*frø*), Norwegian (*frosk*) and Swedish (*groda*). Since this word was crucial for understanding the whole story, listeners who did not understand the word for 'frog' in one of the neighboring languages could not answer the open questions about the text.

However, if the investigation aims to pinpoint specific linguistic factors that play a role in explaining the level of intelligibility, methods that involve smaller entities, such as words rather than whole texts, are more suitable. To illustrate, the researcher may be interested in experimentally testing the separate contribution of vowels, consonants, prosody, cognates, or word order to the intelligibility of a language. By keeping all but one factor constant and systematically varying the characteristics of this one target factor, any difference in intelligibility must be the effect of the variations in the target factor (experimental setting). Suppose we want to test the hypothesis that Swedes have difficulties understanding Danish due to the weakening of consonants in Danish. In that case, we can substitute the weakened consonants in recordings of Danish with their Swedish unweakened equivalents. If Swedes understand the manipulated version better than the original version, the consonant weakening must be the cause of the reduced intelligibility of the Danish version.

An advantage of testing isolated words is that the influence of the context or the situational redundancy on the intelligibility of a word can be excluded. This makes it possible to draw conclusions about the role of individual word characteristics in intelligibility. It is more challenging to trace back poor intelligibility to specific sources at sentence or higher levels. If the words are presented in a sentence or whole text, the context or the situational redundancy may make up for the poor intelligibility of single words. The word predictability depends on the content of the context (see Section 2.2.2.2). For instance, the predictability of the word *car* is low in the sentence *I bought a car* but higher in *I brought the car to the garage*. However, this is complicated by the fact that the pronunciation of predictable words is often less clear than that of unpredictable words. Lieberman (1963) found that unpredictable words in sentences tend to have a longer duration, higher amplitude, and more precise articulation than highly predictable words. This may make predictable words recorded in a sentence more difficult to understand than unpredictable words if they are played in isolation to the listener.

Generally, a word test is ecologically less valid than a test involving whole sentences or texts. However, results from word-level tests tend to correlate highly with results from text-level tests (e.g., Gooskens and van Heuven 2017). This seems logical since a listener must understand individual words to understand a text and get help from the context (Field 2019: 290). Results of word intelligibility tests have also been shown to correlate highly with the general intelligibility of the target language as perceived by listeners in a judgment task (Gooskens and van Heuven 2017). Hence, tests measuring word intelligibility provide a good indication of the overall intelligibility of a language.

Smaller fragments (words, sentences) can be selected randomly under the assumption that a random selection will represent the language as a whole. However, several word characteristics may influence their intelligibility (see Section 5.1.1), and the selection may therefore require some level of control. For instance, one can make sure that the material is phonetically balanced, that is, in accordance with the statistical distribution of phonemes in the language (see Martin, Champlin, and Perez 2000; Field 2019). The concepts need to be known to the listener in order not to test world knowledge. Words can be selected based on a frequency list. The most frequent words in such a list generally cover a substantial portion of ordinary language use. It may be best to avoid homophones since they may confuse the listener. Other word characteristics, such as neighborhood density, false friends, and word length, which are known to influence intelligibility (see Section 5.1.1), should also be controlled for.

2.1.1.4 Recordings

Most of the methods discussed in this chapter are based on listening experiments and, therefore, involve recordings of the target languages. When making recordings, the recording conditions should be as consistent as possible. It is advisable to use the same recording equipment for all recordings and ensure that there is no background noise or bad acoustics. The best recording circumstances are found in a soundproofed recording studio. It may be necessary to normalize the sound level of the recordings afterward to make sure that the level is the same for all recordings. If read speech is collected, it is a good idea to ask the speaker to read the material more than once in case of unwanted effects, such as reading mistakes, background noise, etc. Readers can be asked to repeat if they read too fast, unnaturally, or with a non-optimal intonation. Speakers tend to pronounce word lists with so-called list intonation, using a rising tone on each word except the last one, which is typically pronounced with a falling intonation. It may help to ask the speaker to pause between words or to imagine that someone asks them for each word “How do you pronounce this word?”. The words can also be pronounced in a fixed carrier sentence, such as “I now say . . . . again”. However, it may not be easy to improve the way a speaker reads. Recording several readers and selecting the best one afterward may be a better solution.

When setting up intelligibility experiments, it may be worthwhile using video recordings rather than audio-only recordings as stimulus material. Such stimuli may be more natural and realistic since visual information like gestures and mouth and face movements may add information that makes it easier to understand the speakers (see Section 5.4.2). Innovative approaches with virtual reality (360° videos) offer the potential for an even higher degree of realism. However, opting for visual stimuli presents its own challenges, as visual information can influence the outcomes of intelligibility assessments. For instance, social information about a speaker can influence how well listeners understand the speaker. Kang and Rubin (2009), further discussed in Section 4.3, showed that visual primes in the form of photos of either a Caucasian or an East-Asian person could result in different levels of intelligibility even though the sentences were read aloud by the same American speaker.

2.1.1.5 Speakers

When selecting speakers for intelligibility experiments, it should be considered that some speakers are more intelligible than others because of differences in voice quality (e.g., hoarseness, nasality, breathy voice, creak), precision of articulation, and reading ability.

The gender of the speaker may play a role in intelligibility. In both British (Markham and Hazan 2004) and American data (Karl and Pisoni 1994; Bradlow, Torretta, and Pisoni 1996; Bent, Buchwald, and Alford 2007), there is a small but significant effect of gender. Women are slightly more intelligible than men, with a difference of 2% in the British word data and 3% in the American sentence data. On the other hand, Smorenburg and Chen (2020) mention that previous research on the role of speaker gender in verbal processing suggests that female voices require more, and thus slower, processing than male voices. However, they found no such gender effects in their own research, and neither did Tielen (1992).

It may also be desirable to control for other speaker characteristics, such as level of education, age, regional background, and socio-economic status, when selecting speakers since sociolinguistic research has shown that language use varies across such dimensions too (Krug 2013).

When comparing the intelligibility of multiple languages, it is crucial to consider the impact that a speaker's voice characteristics and background may have on their intelligibility. If the design of the experiment allows it, several representative speakers could be used per language variety. In that way, effects of variability between speakers will average out. Otherwise, a single speaker per language can be selected based on intelligibility scores as judged by native speakers of the same language. Drawing from a larger group of speakers, the speaker whose intelligibility scores are in the middle of the score range can be chosen to represent the speaker group. In a listening test where the intelligibility of only two languages is compared, one option is to use a balanced bilingual speaker to record the stimuli (matched-guise design). The same speaker fluent in two languages is used to record both languages, and therefore, the influence of voice characteristics on the intelligibility is matched or kept constant. To ensure that the speaker sounds native-like in both languages, a voice line-up can be arranged similar to the procedure used in forensic phonetics (Broeders, Cambier-Langeveld, and Vermeulen 2002). To select a bilingual speaker for a matched-guise experiment (see Section 4.3.1), Hilton et al. (2022) conducted two voice line-ups, a Danish and a Swedish one. They presented native listeners with a number of recordings of native speakers, including one recording by the bilingual speaker. They instructed them to pick out the one speaker that did not sound native. The results showed that the bilingual speaker was not picked out more often than the distractors in any of the two languages and could, therefore, be regarded as a representative speaker of both native Danish and Swedish.

2.1.2 Data collection

2.1.2.1 Equipment

In the past, the basic equipment needed for an intelligibility investigation was pen and paper and, in the case of a listening experiment, recordings of the target languages involved in the investigation. It was time-consuming to collect data because it was necessary to print forms to be filled in and to bring the tape recorders to the listeners. After the experiment, the researcher had to spend time on data-entering. The introduction of computer-based methods and software for conducting experiments has made life easier for researchers because data can be collected digitally. This makes it easier to collect and process the data. However, researchers should carefully consider whether they are willing to spend time acquiring the necessary skills to use the software (see Section 2.2.5) or whether somebody who already has these skills can support them. Even though software tends to become more and more user-friendly, some software packages still need programming skills. Furthermore, the software may not be suitable for the experiment that the researcher has in mind. Therefore, it is important to check the limitations of the software well before deciding to use it.

The software used for conducting the experiments may have to be installed on individual computers, which makes it necessary to have the listeners do the experiment on a computer that has the software installed. However, the internet has made it possible to carry out web-based experiments. The advantage of a web-based experiment is that the listeners can simply be provided with a link to the investigation. They do not need to meet the researcher or download any software, and the responses will be returned to the researcher in a format that is easily analyzed. The drawback is that researchers have less control of the experimental situation if they are not present during the sessions. They cannot be sure that listeners do not ask for help during the experiment or are distracted by noisy surroundings or bad equipment. Furthermore, it is not possible to immediately answer questions that the listeners may have, so the instructions must be very clear and provide all necessary information. At the same time, the instructions should not be too long since this may cause the listeners to lose interest even before starting the experiment.

A noisy or distracting test situation can influence the results, so the test location should be chosen carefully. It is a good idea to ask the listeners to listen to stimuli via good-quality headphones rather than over speakers.

2.1.2.2 Listeners

The task performance of listeners is always somewhat variable depending on many different listener characteristics, such as their level of education, intelligence, personality traits, age, gender, socio-economic status, geography, language background, and experience with the target language and other languages. Listeners may also be influenced by other factors, such as their motivation to carry out the test task and their attitude toward the target language. These factors are discussed in more detail in Chapter 4.

Depending on the situation and the nature of the test, listeners can be tested as a group or individually. Groups may be easier to convene than multiple individuals because only one appointment has to be made, such as with a class teacher or the organizer of an event. However, sometimes it is necessary to test listeners individually, which makes it more time-consuming to collect a sufficient number of tests for a reliable statistical analysis.

To ensure that differences in intelligibility between target languages cannot be attributed to unwanted factors, it is important to select a fairly large and well-defined group of listeners that should be comparable across the target groups. To illustrate, the researcher can choose to test a balanced number of male and female students between 18 and 25 years who are born and raised in specific locations and who have no prior exposure to the target language. However, even such precisely defined groups may not be completely comparable. For example, each country has its own school system, which makes it difficult to compare the educational levels in different countries.

To control for all the listener characteristics mentioned above, an intelligibility test is often accompanied by a background questionnaire that the listeners have to fill in. Here, questions are asked to provide demographic information (e.g., age, gender, places the listeners have lived, language background of the listeners and their parents, schooling, etc.). More detailed questions can, of course, be added if necessary. The questionnaire can also include questions about experience with the target language (see Section 4.1.1) and attitude toward the target language (see Section 4.3.1). If the questions are asked using an online form, multiple-choice menus can be used to make it easier for the listener to fill in and for easier administration and processing for the researcher. The answers to the questionnaires can be used to exclude certain listeners from further analysis because they do not meet all listener criteria or to investigate whether listeners with different characteristics perform differently in the intelligibility test. For example, the researcher may be interested in comparing the performances of listeners with different genders or age groups. Kaushanskaya, Blumenfeld, and Marian (2020) developed the Language Experience and Proficiency Questionnaire (LEAP-Q), a tool for collecting self-reported data about proficiency and experience from bilingual and multilingual participants ranging from 14 to 80 years old. It is available in over 20 languages and can be administered in a digital, paper-and-pencil, and oral interview format. In Figure 2.3, a simple example of a questionnaire is provided.

---

1. What is your gender?
   - ☐ Male
   - ☐ Female
   - ☐ Other/prefer not to tell
2. How old are you?_________ years
3. What is your highest completed level of education?
   - ☐ None
   - ☐ Elementary school
   - ☐ High school
   - ☐ Bachelor’s degree
   - ☐ Master’s degree
   - ☐ PhD
4. What is your profession?_______________________________________
5. What is your place of birth (city and country)? ________________________
6. Where did you grow up (city and country)? __________________________
7. What is your native language or dialect? ____________________________
8. Do you speak other languages at home besides your native language?
   - ☐ No
   - ☐ Yes, namely: ____
9. Which languages have you learned (e.g., in school)? ____________________

Figure 2.3: Example of background questionnaire.

A frequent stumbling block in investigations involving listeners is the availability and willingness to do the test among listeners that fulfill the criteria formulated by the researcher. Many researchers will recognize the situation where they have thought of an interesting setup of an experiment, but when they start to look for listeners, they discover that it is difficult to recruit enough listeners for their design. The number of listeners fulfilling the criteria may be small, or the potential listeners may not be motivated to participate. Researchers have to think of ways to get into contact with their target group and ways to make it attractive to participate. A useful point of departure is often to use one's own network and the networks of friends, family, and colleagues. If the experiment is web-based, it is possible to use social media to spread the link to the experiment and maybe benefit from a snowball effect by asking people to spread the link. It may help to offer something in return, such as a small gift or money, or to write an engaging text that makes it clear to the listeners why participating is worthwhile, for instance to help advance science or to test their language knowledge. Recently, it has become popular to present the test as a kind of game that is fun and entertaining and often contains some competitive element (gamification, see Leemann, Derungs, and Elspaß 2019).

When using web-based experiments, the researcher is in less control of who may get the link to the experiment. It may be necessary to remove listeners from the data set because the listeners did not fulfil the criteria for participation or did not take the experiment seriously. It may be challenging to decide whether the listeners made an effort to do the test as well as possible. However, this is sometimes clear, particularly if their score is very deviant from the rest of the group. Brysbaert et al. (2014) ensured that student participants took their task seriously and would not generate numbers at random by telling them in advance that they would only be paid if their results correlated positively with those of the other listeners.

Finally, it is a good idea to check whether listeners have any hearing problems in the case of a listening experiment or bad eyesight in the case of experiments that involve reading or looking at pictures. Also, note that illiteracy can be a problem. In developing countries, there are usually large segments of the population that are illiterate, but also in Western countries a large percentage of the adults are illiterate or have difficulties reading. As an illustration, around 15% of Dutch adults currently are functionally illiterate and may not be able to carry out a test involving reading and writing (Stichting Lezen en Schrijven 2021).

2.1.3 Designing the study

2.1.3.1 Qualitative or quantitative

When setting up an investigation, a major consideration is whether it is necessary to express the results in numbers and use a quantitative approach. This may be the case if the aim is to investigate how well one language is understood compared to another or to carry out statistical tests to show the relationship between intelligibility and various linguistic and extra-linguistic factors (see Chapters 4 and 5). An alternative to such a quantitative approach is a more qualitative approach where the main interest may be investigating people's strategies when trying to make sense of a closely related language (see Section 4.6).

Both approaches come with their own challenges. Test situations are often somewhat artificial, and the quantitative results from intelligibility tests may not reflect actual understanding optimally. In real life, mutual understanding depends on interactive cooperation, and in a natural context, the number of possible interpretations of an utterance is often reduced. Listeners are often good at achieving pragmatic communicative goals, even if they only understand a little of what is said. Some of the experimental methods for testing intelligibility resemble real-life situations more than others. Still, they may reveal less information to the researcher about the factors that may determine the level of intelligibility.

On the other hand, it can be challenging to determine if listeners genuinely comprehend the target language as they are often proficient at concealing their misunderstandings and adapting their speech to their conversation partner. This may be a potential disadvantage of the qualitative approach unless the investigation aims to study how negotiation of meaning and accommodation works towards successful communication. Moreover, qualitative approaches typically require significant effort and labor as they involve a meticulous examination of conversations, making them time-consuming.

In the rest of this chapter, the main focus is on experimental methods for testing intelligibility that will result in quantitative results. However, even in quantitative investigations, it may be informative to ask the listeners more qualitative questions about their experience with the languages and the particular characteristics of languages that make them difficult to understand in addition to the quantitative measurements. This information may be helpful when interpreting the quantitative results.

2.1.3.2 Ceiling/floor effects

When designing an experiment, it is crucial to consider the constraints of the assigned task. If the task is too simple or challenging for a particular audience, it can impede or prohibit the interpretation of the findings. Also, the listeners may get bored if the test is too easy or frustrated if it is too difficult and, therefore, decide not to finish the whole test. Thus, it is vital to avoid such circumstances.

Some investigations may involve target languages that are very similar to the listener languages. In such cases, the task may be too easy, and (almost) all listeners may get a score close to the maximum. This will result in a ceiling effect, i.e. a situation where a measurement cannot take on a value higher than some limit or "ceiling". It may then be necessary to use more sensitive test procedures. In addition to choosing a more difficult task or text, there are several other ways to avoid ceiling effects. The intelligibility of a spoken text can be reduced by artificially speeding up the spoken text (Janse, Nooteboom, and Quené 2003; Syrdal et al. 2012), adding noise (Gooskens, van Heuven, van Bezooijen, and Pacilly 2010; Lecumberri, Cooke, and Cutler 2010), applying filtering (Wang et al. 2011), or signal compression as used in GSM telephony (Nooteboom and Doodeman 1984). The task can also be made more difficult by putting the listeners under time pressure, either by requesting them to complete the task as quickly as possible or allowing them only a limited amount of time to respond. Another way to minimize a ceiling effect is to measure reaction time. The underlying assumption is that the quicker the listeners respond, the higher the level of intelligibility. Reaction times give a precise and sensitive measurement of processing costs, and even if the listeners answer all questions correctly, there may still be a difference in the time it took the listeners to understand the various stimuli correctly. However, if a task is really easy, listeners will react with a minimum reaction time. In such a situation, reaction time measurements will not solve the ceiling effect problem.

On the other hand, if the task is too difficult, the percentage of correct answers may be so low that it is also difficult to interpret the results (floor effect). To check that the task itself is not too difficult and to test the level of difficulty in advance, a reference condition can be built into the experiment, with a control group of native speakers listening to their own language. If they cannot perform the task in their own language, the task is obviously too difficult. If the memory limitations of listeners are not considered, the task may become excessively challenging. Hence, it is advisable to avoid complex tasks or lengthy sentences that could overburden the memory capacity of the listeners. However, it should be kept in mind that even in optimal situations, native listeners tend to make mistakes. To make the task easier for the listeners, it may help to have a recording played twice (double play). It may be argued that this does not reflect a real-world situation, where a listener is usually not able to hear spoken input more than once. On the other hand, the possibility of listening twice may reduce anxiety among listeners. In addition, it simulates situations where listeners ask speakers to repeat what they said. Field (2019: 306) discusses the pros and cons and the effects of double-play in detail.

2.1.3.3 Priming effects

When the same stimulus is presented more than once, this may result in a priming effect, an effect that occurs when listeners' exposure to a certain stimulus influences their response to a subsequent stimulus, which has something in common (form or meaning) with the previous stimulus. To illustrate, if English listeners first hear the Danish word *træ* ‘tree’ and later hear the Swedish equivalent *träd,* there may be a greater chance that they will translate the Swedish word correctly than if they had not heard the Danish word first. Therefore, in an experimental setting (see Section 2.1.1.3), the same (or similar) stimuli should not be presented to the same listener more than once. This contradicts the fact that it is desirable to use identical stimulus material when establishing the listener’s understanding of more than one language (see Section 2.1.1).

A solution is to use a Latin square design whereby each listener is exposed to an equal proportion of the stimuli in each target language and never hears the same stimuli (for example a word, a sentence or a text) in more than one version (Godfroid 2020). An example of a Latin square design is presented in Table 2.1. In this simple example, the test material consists of four different stimuli (1 to 4) in four languages (A, B, C, and D). This results in four different test versions (I to IV). Only one of the four versions is presented to each group of listeners. As an illustration, listeners who are tested with version II hear stimulus 2 in language A, stimulus 3 in language B, stimulus 4 in language C, and stimulus 1 in language D. It takes some effort to administer a Latin square design because different listener groups will be presented different stimuli. Many listeners are needed to fill all the cells in such a design with enough listeners for a sufficiently powerful statistical analysis. Four groups of listeners are required for the example in Table 2.1, one for each version.

From Table 2.1 it becomes clear that the different stimuli 1, 2, 3 and 4 are presented at different places in the experiment. Stimulus 2 follows after stimulus 1 in three of the versions. The place and order of the languages is always the same, starting with language A and ending with language D. This may have some influence on the results because of learning effects and fatigue among the listeners. A way to counterbalance this effect is to have eight different versions where versions 5 to 8 are presented in the mirrored order of version 1 to 4. The experiment can also be programmed so that the listeners are presented with a random selection of stimuli in a random order, but without the same words or texts presented twice.

Table 2.1: Latin square design with languages A–D, stimuli 1–4 and test versions I–IV.

Language	Test version			
	I	II	III	IV
A	stimulus 1	stimulus 2	stimulus 3	stimulus 4
B	stimulus 2	stimulus 3	stimulus 4	stimulus 1
C	stimulus 3	stimulus 4	stimulus 1	stimulus 2
D	stimulus 4	stimulus 1	stimulus 2	stimulus 3

2.1.4 Summary

To summarize the methodological considerations discussed in this section, a list of questions researchers may ask themselves before setting up an investigation is presented in Figure 2.4. It is advisable to keep track of all the decisions made in a log book since it is easy to forget why specific choices were made during the development of an experiment. It is also advisable to draft an initial version of the methods section for the paper intended for publication of the results. While writing this section, errors in method and design will often become apparent, and a peer reader may uncover blind spots. A final piece of advice is to carry out a pilot experiment before starting with the actual experiment.

- What is your **research question**? Formulate research question and hypotheses (expected answers to the question) as precisely as possible. (2.1)
- What is the **time frame** and is it realistic to carry out your project within this time frame? Make a timeline. (2.1)
- What are the **costs** for carrying out your project and do you have the necessary funding? Make a budget. (2.1)
- Do you need help and **assistance** from others for setting up the experiment, carrying out the experiment and analysing the data? Are these people available to help you when you need them? (2.1)
- Do you have access to **language expertise** to help you prepare your test material, e.g., making translations? (2.1)
- Do you have the necessary **equipment** (e.g., headsets and computers) and **software**? Do you have the necessary **skills** to use this software? (2.1.2.1, 2.2.5)
- What **method** will you use to test intelligibility? (2.2)
- Do you have **ethical approval?** (2.1)
- What **test material** will you use? (2.1.1)
  - Will you test with pen-and-paper experiments or via the computer? And if you develop a computer-based experiment, will it be stand-alone or web-based? (2.1.2.1)
  - Spoken or written? (2.1.1.1)
  - Spontaneous, semi-spontaneous or read? (2.1.1.2)
  - Words, sentences or texts? (2.1.1.3)
  - Random selection or based on certain characteristics? (2.1.1.3)
  - Topic? (2.1.1.3)
  - Audio, video, or both? (2.1.1.4)
  - Style and complexity? (2.1.1.2)
  - Monologues or dialogue? (2.1.1.2)
  - Is the method appropriate for your target group? (2.1.1.1, 2.1.2.2)
  - Who will be the speakers (gender, age, language background, voice characteristics?) (2.1.1.5)
  - What recording equipment will you use? (2.1.2.1)
  - Where will the recordings take place? (2.1.2.1)

Figure 2.4: List of questions that researchers may ask themselves before setting up an experiment.

- Did you take into account **methodological considerations**? (2.1)
  - Is there a ceiling or floor effect? (2.1.3.2)
  - Is there a priming effect? (2.1.3.3)
  - What does your design look like? (2.1.3.3)
- Who will be the **listeners** (2.1.2.2)?
  - Age, gender, social and geographic background, language background, etc. (2.1.2.2)?
  - How many (2.1.2.2)?
  - How will you approach them and motivate them to participate? (2.1.2.2)
- What **testing procedure** will be followed?
  - Will the listeners be tested individually or in groups? (2.1.2.2)
  - Is there a suitable testing location (quiet, reachable, no distractions)? (2.1.2.1)
  - Will listeners be paid? (2.1.2.2)
  - Are there clear instructions (2.1.1.1, 2.1.2.1)?
- How will you analyse the data statistically? How should you organise the data to do the **statistical testing** necessary to answer your research question (2.1.2.2, 2.1.3.1, 2.1.3.3)?

Figure 2.4 (continued)

2.2 Overview of methods

Listening to a closely related language is similar to other instances of nonoptimal speech input, such as listening to talking computers, foreign accents, individuals with a speech disorder, or native speech in noisy surroundings. Native listeners are generally successful in getting the speaker's intentions in such situations where the input speech is nonoptimal. It can be assumed that similar mechanisms are generally involved in decoding all these kinds of speech. As a result, methods for investigating mutual intelligibility between closely related languages can be adopted from other disciplines, such as speech technology, second-language acquisition, and speech pathology.

The rest of this chapter provides an overview of various common and less common methods for investigating intelligibility, discusses their advantages, disadvantages, and difficulties, and illustrates them with examples. Note that each method has different variants and that the overview of methods is not exhaustive. Each subsection deals with a particular level of analysis (word, sentence, text, and discourse), but some methods can be used for testing at various levels. The focus is on quantitative methods and functional tests, which test how well a listener *actually* understands the target language. However, opinion testing and a few qualitative approaches are also discussed. In addition, the chapter mainly concentrates on techniques developed for testing intelligibility of spoken language. Most techniques can easily be adapted to the testing of written language.

When choosing a method for testing intelligibility, it is essential to realize that each method may produce a different score. It is, therefore, possible to determine differences in intelligibility between different languages with one specific test method, but the absolute level of the scores will differ if the same languages are tested with different (more difficult or easier) methods or tasks (see Section 2.3).

2.2.1 Word level

First, methods that are mostly used to investigate intelligibility at the word level are discussed. As discussed in Section 2.1.1.3, word-level testing is generally less similar to naturalistic situations than testing at higher levels but will often more easily allow the researcher to pinpoint linguistic factors that can explain the intelligibility results. If listeners are tested with whole texts, it is difficult to trace the linguistic characteristics that make the text difficult to understand. Listeners can use the context to reach an overall understanding of the text even though there are individual words that they do not understand. If words (or sentences) are used, it is easier to trace difficulties back to particular characteristics of the test material.

2.2.1.1 Word translation task

The most widely used word intelligibility test is perhaps the word translation task. Listeners hear isolated words in the target language and write down or pronounce translations in their own language that capture the same meaning. The percentage of correctly translated words is a measure of overall word intelligibility.

If the test is administered digitally, the number of correct translations can be counted automatically through a pattern match with expected answers. However, it is generally necessary to check the incorrect translations manually. This may not be a straightforward task since words may have several meanings when they are presented out of context. For example, the Swedish word *brist* 'lack' can be translated into Danish *brist* or *mangel*, both meaning 'lack'. Both translations should be counted as correct since listeners have no way to know which of the two synonyms is the correct translation. In the case of homonyms, all possible translations should also be accepted as correct. An example is the Swedish homonym *här,* which can be translated correctly into Danish *hær* 'army' or *her* 'here'. Listeners often make spelling errors. These should be objectively defined, for example as instances where the listeners only misspelled one letter without this resulting in another existing word. In this definition, the mistake in Danish *ærende* (correct *ærinde* 'errand') is considered a spelling mistake. Therefore, *ærende* is counted as correct (only one wrong letter without resulting in another existing word), whereas *aske* (correct *æske* 'box') is not counted as correct because the mistake results in an existing word meaning 'ash'.

It is often not completely clear whether a word is correctly translated or not and it is left to the researcher to decide. Semantic overlap may pose a challenge for the researcher to objectively determine if translations should be classified as accurate or inaccurate. For instance, a Danish listener may translate Swedish *piga* 'maid' into the Danish cognate *pige* 'girl,' which has only a partly overlapping meaning (a young female person). There may also be a morphological overlap, such as if a singular noun is translated into the plural. In some cases, a solution may be to give partial points (for example, half a point) to translations that show overlap with the correct translations.

Because of difficulties with objectively correcting translations, an alternative solution could be to have the listeners select the correct translation from a closed list of alternatives (multiple-choice). Creating such a test can be challenging as the complexity of the test is heavily influenced by the range of alternatives presented. For instance, it will be more difficult to choose the correct translation if there are many words to choose from or if the list contains words that are very similar to the correct answer. These are so-called neighbors or false friends (i.e., word forms that closely resemble the stimulus word but have a different meaning, see Section 5.1.1). Furthermore, if the investigation aims to compare the intelligibility of more languages it is often difficult to select equivalent distractors (non-correct answers) for all languages involved. As a result the results may not be comparable.

Example 2.1: Word translation task Danish/Swedish

Kürschner, Gooskens, and van Bezooijen (2008) tested the intelligibility of 384 frequent Swedish words among Danish listeners via the internet. The translations were automatically categorized as right or wrong by the computer through a match with expected answers. The answers that were categorized as wrong were subsequently checked manually by a Danish mother-tongue speaker. Responses that deviated from the expected responses because of a mere spelling error were counted as correct identifications. The word intelligibility results were correlated with eleven linguistic variables that had been shown to contribute to intelligibility in earlier studies (see Chapter 5). The strongest correlation was found between word intelligibility and phonetic similarity.

The errors that listeners make when they translate words may often be mere spelling or typing errors. However, other kinds of errors may be a rich source of knowledge for the researcher who is interested in understanding how closely related languages are processed. The mistakes that listeners make tell something about how they try to match words in the target language with the corresponding cognates in their native language (see e.g., Gooskens and van Bezooijen 2013; Härmävaara and Gooskens 2019). To gain an even better understanding of this process, the listeners can be asked to reflect on their decisions by thinking aloud while translating (see also Section 2.2.3.4).

Example 2.2: Word translation task European languages

To test word intelligibility in 70 combinations of closely related languages involving 16 languages in Europe, a list of the 100 most frequently used nouns in the British National Corpus (*BNC Consortium 2007*) *was compiled (Gooskens and van Heuven 2017, see Chapter 3). The target words were translated into the other target languages and recorded by four native standard speakers per language. Each speaker contributed a different quarter (i.e., 25 words) of the stimulus words. To ensure that the test did not exceed a reasonable duration, each listener was presented with a randomized subset of 50 words from the larger set of 100 words. Listeners were instructed to translate each stimulus word into their native language using the computer keyboard. The results were correlated with those of a (sentence level) cloze test (*r *= .73), see Section 3.2, and with judged and perceived intelligibility (*r *= .72 and .78), see Section 2.2.3.1.*

2.2.1.2 Picture pointing task

Sometimes, it is not possible to have listeners write down their answers or choose from a list of words because they cannot read and write, as is the case with young children, oral cultures, and illiterate adults. In such cases, a special version of the multiple-choice test, the picture-pointing task, may be a good alternative (Gooskens and Schneider 2016).

Listeners hear recorded words in the target languages, and for each test word, they are shown a card with four (or more) pictures, one of which depicts the test word. They have to point to the picture that corresponds to the word that they have heard. The responses can easily be noted down on the spot, even by researchers who do not speak the target languages. This makes the test suitable for use in communities without a strong tradition of literacy and where the researcher is interested in the mutual intelligibility of lesser-known languages.

Alternatively, a touchscreen on a computer or tablet can be used to register the responses. This saves time because no data entry needs to be done afterwards and because it allows for automatic data analysis. It also makes it possible to measure reaction times. The time taken to make a decision can be viewed as a reflection of processing complexity. The level of intelligibility can then be expressed as the percentage of correctly identified words as well as the time it takes for the listener to make the decision. Gooskens and Schneider (2016) provided a checklist of steps to take when developing the material for a picture-pointing task. Figure 2.5 shows one of the picture-pointing cards.[5]

5 The full set of cards is available via http://www.let.rug.nl/gooskens/picture_cards_gooskens_and_schneider.pdf

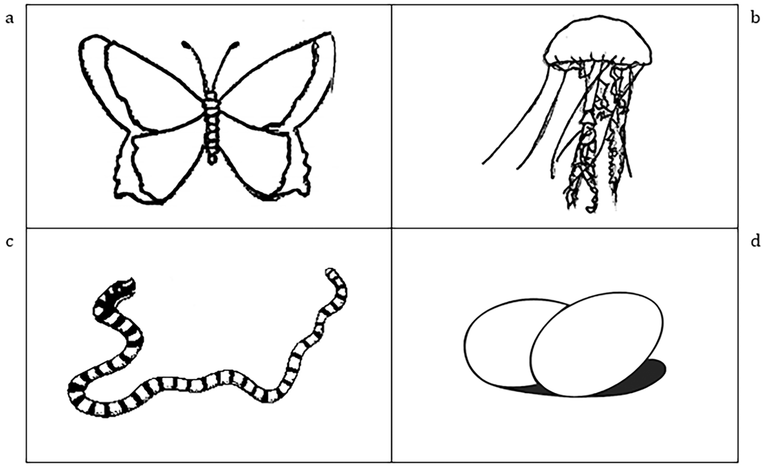

Figure 2.5: A card used for a picture-pointing task with nouns as target word and distractors ('egg', 'butterfly', 'jellyfish', and 'snake'). Source: Gooskens and Schneider (2016: 286).

The words that can be chosen for the test are limited because they need to be 'picturable'. This is difficult for abstract concepts, and nouns are generally easier to draw than other word classes. However, verbs have also been successfully pictured, such as in the Peabody Picture Vocabulary Test (Dunn and Dunn 2007). For her research on the mutual intelligibility between five languages spoken in the Shefa province of Vanuatu, Severin (in progress) created pictures using artificial intelligence. This made it possible to use a uniform style without irrelevant details; see example in Figure 2.6. It may also be an option to picture verbs with short video recordings showing an action.

When choosing pictures to represent the target words, it is necessary to test in advance that they are recognizable to the target group. In some oral societies, there may be a certain lack of picture literacy. It may be necessary to set up a pre-test to check labeling consistency with participants from the target group. They can be asked to label the depicted objects as spontaneously as possible and pictures that are not consistently labeled should be excluded from the experiment. In particular, some target groups might find it challenging to interpret line drawings. It may be better to use photos. On the other hand, photos make it more difficult to abstract from irrelevant details. Brodeur, Guérard, and Bouras (2014) provided a databank of standardized images to use for scientific purposes.[6]

6 Other resources can be found at the internet, see e.g., http://pics.stir.ac.uk/.

Figure 2.6: A card used for a picture pointing task with verbs as target words and distractors ('draw', 'inject', 'call', and 'paint'). Source: Severin (personal communication).

It is equally important to select the distractors with care. The distractors should be just as recognizable as the target pictures. Pictures showing words with high neighborhood density, ambiguous meanings, homonyms, or similar meanings as target words must be avoided. It might be necessary to check this with a native speaker. Both the target and distractor pictures should be presented in the same uniform style and size, preferably drawn by the same artist, so as not to draw more attention to some pictures than others.

Picture pointing tasks are enjoyable and easy to explain, and so far, tests have shown that listeners feel confident when taking part because it does not feel like a real test but rather like a game and because they do not have to produce any spoken or written answer. This makes the test suitable for use with oral communities and for testing children and illiterate adults. The results of the picture-pointing task reflect general intelligibility well (Gooskens and Schneider 2016). However, developing the test is rather time-consuming because of the necessity of preselecting test words, target pictures, and distractor pictures.

Example 2.3: Picture pointing task

Schüppert and Gooskens (2011, 2012) tested the mutual intelligibility of Danish and Swedish preschoolers and adults using a picture-pointing task. The stimulus material consisted of nouns chosen based on high frequency, cognateness, and "picturability". The pictures were pre-tested for labeling consistency among 4-year-old Swedish and Danish children. The scores of the children exhibited a near-normal distribution, with mean scores of 63% for the Danish and 65% for the Swedish listeners. However, the scores of the adults suffered from a ceiling effect (around 90% correct). The ceiling effect stresses the importance of pre-testing the material and recording reaction times in addition to accuracy.

Pinto and Zuckerman (2019) note that the explicit presence of alternatives in the picture-pointing task is ecologically invalid, since, under normal circumstances, listeners do not make a conscious choice between several alternatives when hearing a word in another language. They present an alternative, the Coloring Book. This is a web application where listeners are asked to color items on a coloring page by simply touching them on a touch-screen tablet. An example is provided in Figure 2.7. The subject colors the item in the coloring plate that they thinks corresponds to the word in the prompt. Each coloring plate contains several items (for vocabulary assessment up to 10–12 different items) in a natural context (a classroom, a farm, a birthday party, etc.). The application can also be used to test sentence comprehension.

2.2.1.3 Lexical decision task

Another method that can be used to circumvent the challenges associated with correcting translations and the ceiling effects is a lexical decision task where the listeners are played a word-like sequence of sounds in the target language. They are required to determine, as fast as possible, without making any errors, if the sequence is a word or a non-word in that language. The assumption is that the faster the listeners respond with the correct decision, the easier it is to process the word. Both the existing words and the non-words should be carefully selected. The non-words should be formed by sound sequences that adhere to the rules and restrictions concerning how syllables can be created in the language. In addition, they should comprise an identical number of morphemes as the existing words. Non-words can be generated by replacing certain sounds in existing words or by adding existing morphemes to phonotactically possible non-words. For instance, the long /ee/ in Dutch *komeet* 'comet' can be replaced with a long /oo/ to create the non-existing *komoot,* or the existing morpheme *-ig* can be added to the non-word

Figure 2.7: Example illustrating the Coloring Book method. The listeners hear a prompt containing coloring instructions, in this case *The balloons are red*. Source: Pinto and Zuckerman (2020).

fant to create *fantig*. The non-occurrence should be checked in dictionaries or with a search engine.

Example 2.4: Lexical decision task

Impe, Geeraerts, and Speelman (2008) presented 200 existing and as many non-existing words recorded in ten different varieties of Dutch to listeners from the Netherlands and the Dutch speaking part of Belgium. They were asked to decide as quickly as possible whether the items were existing Dutch words or not. The existing words were subdivided according to nationality-based typicality (typically Netherlandic or typically Belgian), frequency, and word class. In addition to the lexical decision task, the listeners were also asked to decide which of two possible alternatives best reflected the meaning of the stimulus words, one option being a synonym or semantically strongly related word and the other being a semantically unrelated word. In their analysis of the reaction times, they only took into account the stimuli that the listeners categorized correctly in the lexical decision task and identified correctly in the multiple-choice task. The results showed that Belgian listeners had significantly fewer problems understanding Netherlandic Dutch than vice versa. Furthermore, they found a clearly positive effect of the degree of standardness on the ease of word recognition.

2.2.1.4 Semantic classification task

Another method that can involve reaction time measurements is a semantic classification task where listeners hear words and, for each word, have to decide, as fast as possible while avoiding errors, to which of several pre-given categories the words belong. For instance, in an investigation on the mutual intelligibility of 15 Chinese dialects, Tang and van Heuven (2008) had listeners classify the dialect words that they heard as one of ten semantic categories by ticking one of ten boxes: body parts, plants (sweet fruits and nuts), plants (vegetables), animals (four-legged), animals (other), textiles/fabrics/articles of clothing, orientation in time/space, natural phenomena, perishables (food/drinks other than fruit and vegetables), and verbs of action/things people do. If, for example, the listeners heard the word for 'apple' they should categorize it as a member of the category 'fruit'.

The underlying assumption is that correct categorization can only be achieved if listeners correctly recognize the target word. The percentage of correctly categorized words is a measure of word intelligibility. Reaction time can serve as an indication of processing difficulty. The choice between categories can be binary (e.g., tangible or intangible, animate or inanimate) or multivalued. However, if the list of choices becomes too long, it may result in noise in reaction times because listeners spend time searching for the correct category in the list. On the other hand, the role of guessing is, of course, smaller the more categories the listeners have to choose from. The test affords fast and economical testing of the recognition of a large number of isolated words. Like in the picture-pointing task (see Section 2.2.1.2), responses can be collected and scored without any interpretation by the researcher. This can also be done automatically if a computer-based experiment is used.

Example 2.5: Semantic classification task

To test the mutual intelligibility of Maltese, Libyan Arabic, and Tunisian Arabic, Čéplö et al. (2016) used a classification task with 11 categories: animals, body parts, clothing/jewelry, colors/shapes/properties/, eating/drinking, emotions, family/other people, in the house, orientation in space, time, and world around us. The listeners select the semantic category by tapping one of 11 icons representing that category as both text and as a simple black-and-white image on a touchscreen (see Figure 2.8). The results showed that Tunisian Arabic has the highest level of mutual intelligibility with the other two varieties. Additionally, an asymmetric mutual intelligibility was found, with speakers of Tunisian and Libyan Arabic better able to understand Maltese (40% correct answers), than the other way around (about 30 % for either variety of Arabic).

Figure 2.8: Example of a classification task with 11 categories. From Čéplö et al. (2016).

2.2.2 Sentence level

The sentence level is the level between words and whole texts. The advantage of testing at the sentence level compared to the word level is that it is more natural and includes the syntactic and morphological levels. The advantage of testing at sentence level rather than at text level is that it has fewer memory limitations. The short-term memory has limited capacity, and a listener can only remember short sentences.

2.2.2.1 Full-sentence translation

Full-sentence translation was first used to assess the interlingual intelligibility of native American languages (e.g., Voegelin and Harris 1951; Hickerson, Turner, and Hickerton 1952; Pierce 1952; Biggs 1957). Listeners hear a (recorded) spoken sentence and write down (or say aloud) the translation into their native language.

If the sentences are longer than a few words, it may be necessary to repeat the sentence to reduce memory load (see Section 2.1.3.2). This can be compared to a natural situation where the listener asks a speaker to repeat a phrase or sentence. The intelligibility score is based on the percentage of words in an utterance translated correctly by the listeners (Derwing and Munro 1997).

The scoring of the responses poses problems similar to those of the word translation tasks discussed above (Section 2.2.1.1). It is labor intensive and often challenging for the researcher to decide whether a sentence is correctly translated. However, unlike the word translation task, it is difficult to have the translations corrected automatically. Syntactic, idiomatic, and morphological differences between target language and listener language make it more complicated to determine how large a proportion of the sentence is correctly translated. Content words (nouns, verbs, adjectives, and adverbs) are likely to be more fundamental for the understanding of an utterance than function words (e.g., auxiliary verbs, prepositions, articles, conjunctions, and pronouns) that add little meaning beyond defining the relationship between words (van Bezooijen and Gooskens 2007). As such, it may be reasonable to assign different weights to different word categories, but it is not clear how these differences should be weighted. Like in the case of word translations, the translations can be supplemented with a think-aloud task to gain more information about the translation process (see Section 2.2.3.4).

Example 2.6: Full sentence translation task

Gooskens et al. (2010, see Section 5.4.1) tested the mutual intelligibility between Swedish and Danish with spontaneously produced sentences. Intelligibility was expressed as the percentage of content words correctly translated, disregarding any errors in the recognition of function words. The asymmetry traditionally claimed between Swedish and Danish was indeed found, even when differences in familiarity with the non-native language were controlled for. In a separate task, the same listeners listened to the recordings in their own native language, presented in descending levels of noise. The results show that Danish is as intelligible to Danish listeners as Swedish is to Swedish listeners, thereby rejecting the hypothesis that Danish is an intrinsically more difficult language than Swedish.

2.2.2.2 Partial-sentence translation

In various types of translation tasks the listeners are asked to translate only part of a sentence. An example is the Speech In Noise (SPIN) test (Kalikow, Stevens, and Elliott 1977) that was developed for testing English speech intelligibility in noise but has also been used to measure the intelligibility between closely related languages (e.g., Tang and van Heuven 2008, 2009, 2015). Results by Wang (2007) showed that the SPIN test is highly sensitive to differences in intelligibility of foreign-accented English as a result of different language backgrounds of both speakers and listeners.

The task requires listeners to write down the translation of the final word they hear in a sentence presented in the target language. The predictability of the final word may vary, and may or may not be predictable from the earlier part of the sentence. An example of a sentence with highly predictability is *He wore his broken arm in a sling* (target underlined), and an example of a sentence where the target word is less predictable from the context is *We could have discussed the sling*. Wang (2007) showed that the highly predictable sentences in the SPIN test were more sensitive to differences between speaker and listener groups with varying levels of proficiency in English.

An advantage of this type of test is that it is easier for the researcher to analyze the data than full sentence translations.

Example 2.7: SPIN test

To investigate the mutual intelligibility of 15 Chinese dialects, Tang and van Heuven (2009) selected 70 sentences based on the high-predictability section in the SPIN test sentence lists and translated them into the 15 dialects. The 70 sentences were selected based on their applicability to the Chinese linguistic/cultural situation. In addition, only sentences that maintained the structure of the SPIN sentences were selected, such that each Mandarin sentence ended with a final noun as it does in English. The SPIN sentences were first translated into Standard Mandarin by the first author and next from Standard Mandarin into the 15 dialects by the designated speakers of each dialect. The results correlated significantly with opinion scores (Section 2.2.3.1) and objective lexical and phonetic distance scores (Sections 5.1 and 5.2).

A related task involves presenting words in a context with part of the message replaced with blanks for selected words only. It is easy for the researcher to process the data, but the disadvantage of this method is that it is uncertain what role the written context plays in the translation results (see Section 2.1.1.1).

Example 2.8: Partial translation task

Van Bezooijen and van den Berg (1999) played semi-spontaneous samples of various Dutch varieties to groups of listeners from the Netherlands and Belgium. The texts were printed in Standard Dutch but dotted lines of uniform length replaced the nouns. The participants were instructed to listen to the recordings and write the missing words on the lines. The results showed that differences in intelligibility between the varieties could be attributed to linguistic differences between the varieties and standard Dutch.

2.2.2.3 Translation of Semantically Unpredictable Sentences (SUS)

If researchers want to exclude the confounding influence of semantic or syntactic contextual cues on the intelligibility of words in a sentence, they may consider using Semantically Unpredictable Sentences (SUS) to determine word-level intelligibility. Such nonsense sentences are sequences of words with normal word order and prosody. They do not provide any clues for the listeners to predict the identity of content words based on sentence semantics or situational context. As an illustration, the syntactic structures are correct in semantically anomalous sentences such as *He drank the wall*, *The state sang by the long week*, or *Why does the range watch the fine rest?* Intelligibility can be expressed as the percentage of correctly translated words in the sentence or be limited to the percentage of correctly translated content words since they are more central to understanding a sentence than function words (see Section 2.2.2.1). However, the simplest and fastest way to score results is to only count the sentences that are entirely correctly translated as correct. This easy-to-obtain scoring approach is strongly related to measures of word intelligibility (Benoît 1990).

The method is commonly used in speech and language pathology settings to investigate the clarity of disordered speech and in testing text-to-speech synthesis intelligibility, but it has also been used to test mutual intelligibility between closely related languages. It is useful for testing intelligibility when the researcher is interested in testing the role of the lexicon and pronunciation but wants to exclude the influence of semantic contextual cues. Listeners receive cues about the syntactic category and prosody only, but they cannot use semantic contextual cues to make additional predictions about word identity. A disadvantage of the method is that the task is unnatural because of the semantically anomalous sentences.

Benoît, Grice, and Hazan (1996) developed a so-called SUS generator to generate Semantically Unpredictable Sentences for several European languages. These sentences consist of common syntactic structures and words randomly selected from lexicons containing commonly used "mini-syllabic" words (i.e., the smallest words within a given category). The syntactic structures are simple, and the sentence length does not exceed seven or eight words to avoid saturation of the listeners' short-term memory.[7]

Example 2.9: Translation task with SUS sentences

Gooskens et al. (2010) used SUS sentences to assess the (asymmetrical) mutual intelligibility of Danish and Swedish as well as the intrinsic intelligibility of the two languages. They used the SUS generator developed by Benoît, Grice, and Hazan (1996) to generate the Swedish SUS sentences. Since no Danish SUS generator was available, they developed one themselves. To counterbalance possible language-specific influences, such as differences in word frequency, half of the 12 SUS sentences originated from the Swedish SUS generator, and the other half from the Danish SUS generator. The Swedish sentences were then translated into Danish, and the Danish sentences into Swedish. The results showed the same overall pattern as translations of spontaneous sentences (see Example 2.6), confirming the Danish-Swedish asymmetric intelligibility.

7 Newer SUS generators are available online (e.g., https://github.com/ecooper7/SUSgen; Lilley, Stent, and Zeljkovic 2012).

2.2.2.4 Sentence verification task

The sentence verification task resembles the lexical decision task (see Section 2.2.1.3) and can be used to avoid a ceiling effect. Listeners hear a sentence that contains a logical proposition that is either true (e.g., *Cows eat grass*) or false (e.g., *Horses are known to climb up trees*). They have to decide as quickly as possible whether the statements presented are true or false (or plausible/implausible, see Example 2.10). This task requires a listener to process the overall information of an utterance and combine background knowledge and knowledge of word occurrences with their ability to perceive individual phonemes. The assumption is that more intelligible speech will allow listeners to understand the message correctly and that the decision will be faster for more intelligible language varieties than for varieties that are more difficult to understand (Munro and Derwing 1995b; Bohn and Askjær-Jørgensen 2017).

The sentences should not be too long and should require little cognitive processing. Their truth value must be easily determined by the listeners on the basis of everyday knowledge. This means that potentially ambiguous or misleading sentences must be avoided. To check that the sentences used for the experiment are all deemed unmistakably true or false to the listeners, a pilot experiment should be conducted with a group of participants with the same background as the prospective listeners in the real experiment. The participants in the pilot read the sentences in their native language. The real experiment should include only sentences that (almost) the whole group consistently finds either true or false. If higher error percentages are found in the real experiment this must be caused by the sentences being difficult to understand rather than by lack of background knowledge.

Since the response is binary, it is recommended to use a considerable number of test items to minimize the impact of guessing. In addition to error counts, reaction times can be measured. In the latter case, the decision times must be lined up with the earliest point in the acoustic stimulus where sufficient information is available to correctly decide on the truth value of the sentence. Responses given before this moment must necessarily be based on guessing. In the sentence *Cars cannot eat books* the listeners can decide on the truth value after they have heard the word *eat,* and therefore the reaction time should be measured from this point.

The appendix in Munro and Derwing (1995a) provides more examples of true/false test sentences.

Example 2.10: Sentence verification task

Hilton, Gooskens, and Schüppert (2013) used an alternative application of the sentence verification test, where Danish listeners presented with Norwegian test sentences were asked to judge the plausibility of the proposition. The aim was to test the influence of morphosyntax on intelligibility. This approach was chosen over a more traditional sentence verification test (where the content would be deemed as either "true" or "untrue") as it was considered impossible to design enough sentence types for the experiment that are universally true. An example of an implausible Norwegian sentence is Du lukter med hendene dine *'You smell with your hands'. Examples of plausible sentences are:*

1a	*En*	*bonde*	*arbeider*	*på*	*sin*	*traktor*
	A	farmer	works	on	his	tractor
	'A farmer works on his tractor'					
1b	*En*	*bonde*	*arbeider*	*på*	*traktoren*	*sin*
	A	farmer	works	on	tractor+THE	his
	'A farmer works on his tractor'					

While 1a is grammatical in Danish (and archaic in Norwegian), 1b is ungrammatical in Danish (and the unmarked variant in Norwegian). Longer reaction times were found for Danish participants for sentences like 1b, which have an extra marking of definiteness. Therefore it was concluded that this extra marking reduces the intelligibility for Danes.

2.2.2.5 Carry out instructions

A simple approach to access sentence intelligibility involves giving listeners a set of instructions to follow. The success rate in executing the instructions, the time it takes the listener to start carrying out the instructions, and the time it takes to complete the task successfully are measures of intelligibility. Typically, the instructions entail moving or arranging objects within a simulated environment displayed on a computer screen.

This method offers the benefit of assessing intelligibility within a realistic communicative context. However, the requirement for listeners to carry out the specified actions can restrict the range of syntactic structures and vocabulary that can be incorporated.

Example 2.11: Carry out instructions task

To test the relative contributions of pronunciation and morpho-syntax to the intelligibility of foreign-accented speech, van Heuven and de Vries (1981) collected short Dutch utterances in which Turkish speakers of Dutch as a second language spontaneously described simple acts that were performed by the experimenter, such as pouring beer into a glass. Next, for each act, three additional sentences were recorded so that, in total, the following four versions could be presented to native Dutch listeners in a Latin square design (see Section 2.1.3.3):

(i) *Original utterance by a Turkish speaker of Dutch, with non-standard pronunciation and morpho-syntax;*
(ii) *Same utterance repeated literally by a native Dutch speaker, with standard pronunciation, but imitating the non-standard morpho-syntax used by the Turkish speaker of Dutch;*
(iii) *Same utterance by a Turkish speaker, with non-standard pronunciation, but standard morpho-syntax;*
(iv) *Same utterance by a native Dutch speaker, with both standard pronunciation and standard morpho-syntax.*

The listeners were instructed to listen to the utterances and perform the act described in each, as promptly as they could, using the same array of objects that had been provided to the original speakers. Both the number of errors (failure to understand the description) and the listeners' reaction time in case of correct understanding were established. It was concluded that phonetic factors play a more important role than non-phonetic factors in the intelligibility of foreign-accented speech.

2.2.3 Text level

If the main interest of the researcher is to get a general impression of the overall intelligibility of a language, it is more natural and often simpler to test at the text level rather than at a lower level of analysis.

2.2.3.1 Opinion testing

To obtain a rapid assessment of the general level of intelligibility of a language, a simple and effective approach is to request participants to rate their comprehension on scale(s). This method was used by ethnographers in early anthropological fieldwork on intelligibility and referred to as 'ask the informant', as opposed to 'test the informant', where intelligibility scores are based on a functional test (see Voegelin and Harris 1951; Casad 1974; Simons 1979).

The most straightforward opinion testing (judged intelligibility) involves no speech fragments. Participants are asked to indicate the percentage of the words of a language they think that they can understand (for example, by moving a slider on the computer screen) or to judge on a Likert-like scale how well they think that they understand a speaker of the target language (for example using a scale from 1 "not at all" to 10 "everything"). Alternatively, they can be requested to assess the extent to which they believe that monolingual speakers of their own language would understand the speaker. This may provide a better estimate of inherent intelligibility. An advantage of using judged intelligibility is that no speech material has to be selected and recorded. Without recordings it is possible to abstract from individual speakers who may influence the results because of specific voice characteristics and speaking styles (see Section 2.1.1.5) as well as differences in recording conditions (see Section 2.1.1.4). On the other hand, it remains unclear whether respondents can consistently assess intelligibility in the absence of speech samples. They may never or rarely have heard the language, may not remember how well they understood the speaker or it may be unclear to them how a typical speaker of the target language sounds. Consequently, respondents may base their opinions on some extra-linguistic factor, such as their positive or negative attitudes toward the country and its speakers (see Section 4.3), political borders, desirable answers, or the geographic distance to the place where the language is spoken.

Often, a better alternative is to present recordings of the target language to listeners before they judge how well they can understand the language (perceived intelligibility). However, even with this approach, it is uncertain whether listeners are actually able to make the judgments on an objective linguistic basis without the influence of extra-linguistic factors.

Developing appropriate functional tests and administering them can be time-consuming, since special care must be taken to prevent priming effects, ceiling effects, memory overload, and other unwanted effects (see Section 2.1). Opinion tests are generally easier to set up and carry out than functional tests. In particular, the same material can be presented repeatedly without the risk of a priming effect (see Section 2.1.3.3). However, the results of opinion testing should be interpreted with some care. They provide us with information about individuals' subjective perception of the intelligibility of languages, but people's self-reported language behavior may not be in line with their actual language behavior. In research on the intelligibility of foreign-accented speech, correlations between opinion measures and functional measures are not always significant (see Lindemann and Subtirelu 2013: 579 for an overview). Lindemann and Subtirelu (2013) mention two potential reasons for low correlations found in previous research. First, they may be caused by a ceiling effect in the results of the functional test (see Section 2.1.3.2). Second, they may be caused by inconsistent listeners' biases where some listeners upgrade and/or downgrade their subjective ideas.

Other studies show that listeners are often good at estimating how well they can understand a closely related language. Significant correlations have been found between opinion scores and results from functional tasks. Gooskens and van Heuven (2017) found correlations between $r = .72$ and .97, and Tang and van Heuven (2009) between $r = .77$ and .81. Significant correlations were also found in studies carried out to evaluate the quality of speech synthesis, see for example van Bezooijen and van Heuven (1997), and references therein. Therefore, opinion testing may provide a useful shortcut to functional intelligibility. However, many researchers doubt the validity of opinion testing and prefer to test actual speech comprehension with functional testing methods. Tang and van Heuven (2009) note that since the correlations are not perfect, mutual intelligibility should be tested functionally whenever the resources are available.

Example 2.12: Perceived intelligibility

Tang and van Heuven (2009) played recordings of the same text, the fable The North Wind and the Sun, *in 15 Chinese dialects to 24 listeners from each place where the dialects were spoken. For each dialect, the listeners were asked to indicate how well they believed monolingual speakers of their own dialect would understand the speaker on a scale from 0 ("They will not understand a word of the speaker") to 10 ("They will understand the other speaker perfectly"). The results showed high correlations between perceived intelligibility and intelligibility tested functionally by a sentence and a word intelligibility test (*r *= .77 and .81) and with objective measures (lexical similarity and phonological correspondence,* r *= .83 and .71).*

2.2.3.2 Recorded Text Testing (RTT)

A common method to test overall comprehension of a text is to have listeners listen to a text and have them answer questions about the content of the text. This method was first used in the fifties of the previous century to establish the mutual intelligibility of American Indian languages by Voegelin and Harris (1951), Hickerson, Turner, and Hickerton (1952), and Pierce (1952). They referred to the test as the Recorded Text Testing (RTT) method. The intelligibility of a language is quantified by calculating the average percentage of correct responses provided by the listeners. The questions can be posed after the presentation of the text or interspersed in appropriate places throughout the text. However, it may be more realistic to provide the questions to the listeners in advance, enabling them to concentrate on the relevant aspects of the content since people mostly listen to a text with a specific purpose in a real-life situation. Casad (1974) and Nahhas (2007) provide a detailed overview of the steps that should be taken to carry out a test with RTT.

The questions about the texts should cover the content of the whole text as well as possible, and a correct answer should not depend on understanding a single specific word. To ensure that the evaluation solely measures comprehension of the text (and not that of the questions), it is recommended to present the questions (and multiple-choice alternatives, if applicable) in the listener's native language.

The test should not measure the listener's memory, general knowledge, or intelligence. Hence, it is crucial to pre-test the questions on a separate group of participants not provided with the text to ensure that correct responses are not based on general knowledge or logic.

Defining correct answers to the questions can be challenging, and researchers may need to distinguish between varying degrees of correctness, such as "completely correct", "partly correct", and "incorrect". A more objective solution is to use multiple-choice questions, where listeners have to choose between a limited number of possible answers. This approach allows for easy manual or computerized correction of answers. However, one drawback of multiple-choice questions is the difficulty in finding plausible distractors that the listener does not easily eliminate. Additionally, multiple-choice questions are rather unnatural since people are usually not given a choice between several possible answers in a natural situation. When comparing intelligibility among different groups of listeners, it may also be important to realize that some listeners are familiar with multiple-choice testing. They may do very well by virtue of understanding how these tests work rather than through actual understanding while others are less familiar with this kind of testing method and therefore do not have this advantage. Kluge (2006) discusses the disadvantages of the RTT method in more detail. She points out that questions may be culturally inappropriate in many societies and that listeners can often infer the correct answers from the questions.

A critical discussion of the reliability of past research with the RTT method is presented by Yoder (2017). The recordings used for the RTT tests have often been different narrative texts that are unique for each speaker and are therefore also followed by different questions. By analyzing 25 papers using the RTT method, Yoder demonstrates that the selection of a text for an RTT has significant implications on the results and could lead to incorrect conclusions.

Example 2.13: Recorded text testing

Delsing and Lundin Åkesson (2005) asked listeners to answer five open questions about short passages of continuous texts. Unlike many other investigations, they used read-aloud texts, which made it possible to use the same texts and questions for all languages in a Latin square design (see Section 2.1.3.3). This is important if results are to be compared (see Section 2.1.1). The texts were short news items about a kangaroo that has escaped from captivity and hops around the streets of Copenhagen and about counting frogs to indicate how well the natural environment is doing. They were read aloud by professional news reporters. Five open questions were asked about the content, such as "For how long was the kangaroo on the run?". A correct answer ("2 months") was given 2 points and a partly correct answer (e.g., "1 to 2 months") was given 1 point.

2.2.3.3 RTT retelling

Kluge (2006) discusses an alternative approach to the standard RTT question format, the RTT retelling method. This method was developed by Ring (1981) and has since been subsequently refined by various researchers. Listeners are required to listen to a narrative divided into natural segments of one or two sentences each and then retell the recorded text in their native language while keeping as much detail of the original as possible. In this way, the listeners do not have to answer specific comprehension questions. For each segment of the text, the number of correctly retold core elements are counted, and the segment scores are added up to obtain the overall score for a given RTT text.

Kluge (2006) provides the following example to illustrate the RTT retelling method. In this example, the listeners hear the following fragment in the target language:

> *As I pulled in the net, a crocodile was sleeping in the water under a Sago tree. I didn't see it. I pulled in the net with the crocodile behind me.*

The researcher defines three core elements (indicated by the three colors) in the fragment, each counting for 1 point:
- (he/the man) pulled/investigate net (1 point)
- doesn't know/see (crocodile) (1 point)
- crocodile in the water/under Sago tree/between Sago roots (1 point)

In the following examples of retellings by three different listeners, all core elements are correctly retold, and all listeners therefore get the maximum number of points (3) for this fragment:

Listener 1: He was pulling like this and didn't see the big crocodile between the Sago roots.

Listener 2: While the man was investigating the net, he didn't know that there was a crocodile there under a Sago tree.

Listener 3: While he was investigating, he didn't see that a crocodile was under the Sago (trees).

The RTT retelling method has several advantages over the standard RTT question method. Firstly, it tests comprehension of the entire text rather than selected sections only. Secondly, retelling a story may be more culturally appropriate and less intimidating than answering comprehension questions, especially in traditional societies. Additionally, this method eliminates the need to design and translate comprehension questions into different speech varieties. However, one major disadvantage is that it is time-consuming to develop the test and count the number of correctly retold segments.

As in the case of the original RTT method, it remains a concern that different narratives are mostly recorded in different languages (see Section 2.2.3.2).

Example 2.14: RTT retelling task

To classify Lalo, a Central Ngwi (Loloish) language cluster spoken in western Yunnan, China, Yang (2012) tested the mutual intelligibility of 18 Lalo villages using a RTT retelling task. To prepare the RTTs, native speakers related short narratives, typically one to three minutes long. Next, for each recording, a panel of eight native speakers listened to the whole text and then listened again section by section, with a pause *after each section. Each section was at most 10 seconds. During the pause, the listeners retold the section's content in Chinese, which most Lalo speakers master. Elements that all native listeners retold formed the baseline for scoring responses for listeners from other varieties.*

For the actual experiment, eight to ten participants with a minimum of previous exposure to the target language were selected in each village. Each participant followed the same testing procedure as the panel of native listeners. A listener's intelligibility score was the number of core elements mentioned divided by the total number of core elements identified by the panel of native listeners. Not all RTTs were tested at each village due to time constraints and the fatiguing nature of the RTT process for the participants.

2.2.3.4 Text translation

The text translation method is very similar to the word translation task (see Section 2.2.1.1) and the sentence translation task (see Section 2.2.2.1); it has the same advantages and drawbacks. The use of whole texts may provide the listener with more context than a sentence translation task. In the case of spoken language, the text is typically presented one sentence at a time while the listener responds by either writing out or pronouncing the translation as literally as possible. Compared to content questions, an advantage of this method is that the researcher does not have to formulate questions about the text.

Translating sentences and texts may be challenging for certain listeners since it is not a natural task for them. The skill of translation entails much more than intelligibility only, and it may heavily rely on the listener's memory. Therefore, the text must be presented in short segments with pauses between them, during which the listener can write down the translation. To facilitate the writing task and to decrease possible memory problems, Scharpff and van Heuven (1988) printed the function words in the target text on the answer sheets and left blanks for content words to be filled in (one blank of uniform length per content word).

To gain an even better understanding of the process of understanding a text in a related language, the listeners can be asked to reflect on their decisions by thinking aloud while translating (e.g., Kaivapalu 2015; Mieszkowska and Otwinowska 2015; Jagrova et al. 2019) but a more natural task may be what Mercer (2000: 98) refers to as exploratory talk, i.e. talk in which partners engage with each other's ideas about how to translate the text. This kind of reflection may provide even more information about the processes of trying to understand a closely related language.

Example 2.15: Text translation

Gooskens, Heeringa, and Beijering (2008) tested the intelligibility of the fable The North Wind and the Sun *in 18 Nordic varieties (Danish, Norwegian, Swedish, and Faroese) among Danish listeners from Copenhagen using a translation task. The listeners were presented with the six sentences of the fable one at a time. Each sentence was in a different variety following a Latin square design (see Section 2.1.3.3). While* *listening to the six sentences, the listeners had to translate each word they heard into standard Danish. Each sentence was presented twice. The sentence was presented as a whole before being repeated in smaller chunks of four to eight words (depending on the position of prosodic breaks) to make sure that saturation of the listeners' short-term memory would not influence the results and that the listeners had enough time to write down their translations. The percentage of correctly translated words constituted the intelligibility score of a given language variety. The results correlated significantly with lexical (*$r = -.64$*) and phonetic (*$r = -.86$*) distances.*

Example 2.16: Text translation with exploratory talk

Börestam and Hansen (2021) used exploratory talk to investigate the strategies of Danes when trying to make sense of a Swedish spoken text and vice versa. Danish and Swedish participants were asked to listen to a text read aloud in the neighboring language. They were instructed to either translate the text (or parts of it) immediately into their native language or to write down (in any form) what they supposed they had heard. Afterwards, the participants worked together in pairs with the same native language and a recording was made. Based on their notes, they made a translation that was satisfactory to both of them. They were told to motivate their choices to each other. The results give an impression of comprehension of a neighboring language as a dynamic and creative process of toggling between the macro and the micro perspective, i.e., from only looking at portions of words up to considering the full context.

2.2.3.5 Picture-to-story matching

The picture-to-story-matching test involves playing a complete text to the listener, who is then presented with several options of pictures on screen or paper and asked to select the picture that best matches the contents of the story. This is, in fact, a multiple-choice task without any written text involved. No questions are posed in writing, nor do the listeners have to write down answers. Like the picture-pointing task (see Section 2.2.1.2), it is therefore suitable for testing children and illiterate adults. If the test is administered on a computer, the answers can be registered automatically using a touchscreen or having the listeners click on one of the pictures with the mouse.

However, the test is a rather rough measurement that may only provide an impression of whether the listeners got the gist of the story. Therefore, it is important to pre-test the task with a separate group of listeners before administering the actual test. To make the test more sensitive, reaction times (see Section 2.1.3.2) can be measured in addition to the percentages of correctly chosen pictures. It may also be possible to improve the picture task by using a larger number and more complex pictures and by varying more message elements (see Example 2.17). This would require the listener to extract more key elements from the text before being able to choose the correct picture.

Example 2.17: Picture-to-story matching

Gooskens and van Heuven (2017) constructed four pictures to embody the correct or incorrect representation of two key elements in a text passage. An example is provided in Figure 2.9. For instance, if the passage was about driving a car in winter, one picture showed a car driving in summer (with a sunny landscape and trees and flowers in full bloom, picture a), another picture contained a plane in a summer setting (picture b), a third picture showed a car driving in a wintery landscape (picture c), and a last picture showed a plane flying over a wintery landscape (picture d). When both content features were correctly identified (picture c), the listeners got full points; when both aspects were wrongly identified (picture b), no points were given, and when one feature was correct (pictures a and d), the listeners were given half points. The correlations between the picture task and two functional tests at the word and text level showed correlations of = .73 and .71. The task seemed too easy for some participant groups, creating a ceiling effect.

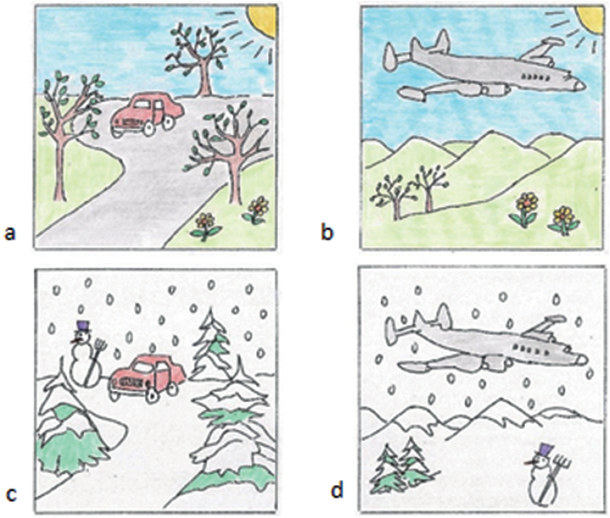

Figure 2.9: Picture used for a picture-to-story matching task. From Gooskens and van Heuven (2017: 28).

2.2.3.6 Cloze test

The cloze test was developed by William Taylor in the United States in 1953 and has since then been used extensively for measuring comprehension in the classroom (Abraham and Chapelle 1992; Keshavarz and Salimi 2007) but also for intelligibility research purposes (e.g., Smith and Rafiqzad 1979; van Bezooijen and Gooskens 2005a, 2005b, 2006). In a cloze test, selected words are removed from the text at regular intervals, for example, every tenth content word, and replaced by gaps, i.e., lines or empty spaces of uniform length (in the case of written language) or by beeps of uniform duration (in the case of spoken language). The percentage of correctly filled-in words is the measure of intelligibility. The cloze test requires understanding context and vocabulary in the text to identify the correct words or type of words that belong in the gaps. It is, therefore, an easy and useful way to test overall text intelligibility.

It is generally left to the listeners to think of suitable words. The scoring of this version of the test is difficult because the researcher has to decide whether the words offered by the listener are correct. Alternatively, the removed words are placed above the text, and the listeners are asked to put the words back in the correct position. Since the aim is to test the intelligibility of whole texts, a translation of these words should be provided. If some of the response alternatives were unfamiliar to the listeners, they would not be able to place them in the correct gaps, even if they did understand the texts in their entirety. This version of the cloze test can be scored automatically and is, therefore, an efficient and objective way of testing text comprehension, especially when a large number of listeners are involved. Usually, the test is employed to establish the intelligibility of written texts, but it can be adapted to test spoken language. In that case, the listeners will hear only one sentence at a time to avoid the influence of memory limitations. Chapter 3 provides a description of the spoken cloze test used to test the mutual intelligibility of spoken languages in Europe in the MICReLa project.

Example 2.18: Written cloze test

Van Bezooijen and Gooskens (2005a, 2005b) measured the intelligibility of written Afrikaans and Frisian texts among native Dutch participants using a variant of the cloze test. As a basis for assessing the intelligibility of written text, they used two Dutch newspaper articles (of 329 and 317 words) with an average level of difficulty. In both texts, five nouns, five adverbs, five adjectives, and five verbs were randomly selected. These were placed in alphabetic order above the text and replaced by blanks in the text. Next, the two texts were translated into Frisian and Afrikaans. The same words were removed and placed above the texts. Below, one of the Afrikaans cloze tests is shown (from van Bezooijen and Gooskens (2005b: 24). The participants were given ten minutes to put the 20 words back in the right place in the texts. The percentage of words placed back correctly was taken as a measure of the intelligibility of the texts. It appeared that it is easier for Dutch speakers to understand Afrikaans than Frisian.

aantreklike	al	blyk	doen	eksplosief
fantasties	internet	lank	misbruik	moeilike
moet	nooit	onbetroubare	paar	soms
stel	teleurgesteld	verhale	werk	woorde

Nie lank en blond nie maar ’n klein, kaalkop lamsak

Jaarliks soek tienduisende alleenlopers via ’n verhoudingsbemiddelingsagentskap of op die ——————————————— na ’n metgesel. ——————————————— met sukses, maar ook dikwels daarsonder. ——————————————— mans blyk in werklikheid oud en lelik, en ——————————————— agentskappe verdwyn skoonveld sodra die inskryfgeld betaal is.

Dit het ——————————————— geklink in die kletskamer op die net. Die ou met wie Annemarie op die internet in kontak gekom het, het aan al haar vereistes voldoen: hy was ———————————————, blond en atleties. Het hy geskryf. "Maar tydens die eerste afspraak in die kroeg het daar 'n lamsak op my gesit en wag," sê die 27-jarige vrou. "Hy was klein, kaalkop en het ook nog tien jaar ouer gelyk as die ouderdom wat hy op my webtuiste opgegee het." Annemarie het drie ——————————————— gewissel met die man en het kwaad en teleurgesteld vertrek. Die sprokie was verby.

Vir sulke kliënte is Joke Pronk aangestel. Sy ——————————————— vir die Algemene Vereniging Verhoudingsagentskappe (AVV) en bemiddel in geskille tussen kliënt en verhoudingsagentskap. 'n ——————————————— saak. "Die mense is ———————————————," aldus Pronk, 'n ——————————————— honderd euro terugbetaling bied dikwels min troos, hulle wil veral stoom afblaas." Pronk ——————————————— vas watter agentskappe nie hul afsprake nagekom het nie en deel "geel kaarte" uit.

Die aantal verhoudingsagentskappe het die afgelope jare – met name op die internet ——————————————— gegroei. Daar is ——————————————— vier – à vyfhonderd (op die net). Met die groei neem die ——————————————— ook toe. Baie internetagentskappe blyk te verdwyn sodra die inskryfgeld betaal is. "Skielik bestaan die webwerwe nie meer nie. Of hulle ——————————————— 'n kantoor in Kazachstan te hê." Pronk ken die ———————————————, maar dit kom volgens haar nie voor by die vyftien agentskappe met 'n AVV-keurmerk, waarvoor sy werk, nie. Sy raai kliënte aan om ——————————————— 'n eie e-posadres te gee nie, maar alleen e-posse te stuur via die webwerf. "As mense afsprake wil maak, adviseer ons nadruklik om dit te ——————————————— op 'n openbare plek. Dit is regtig nie net psigopate wat op die internet na 'n vrou soek soos sommiges beweer nie, maar jy ——————————————— wel versigtig wees.

2.2.4 Discourse level

The tests discussed so far test cross-language intelligibility in one direction, for instance, the intelligibility of target language A among speakers of language B. However, sometimes researchers' main interest is to establish how well speakers of different closely related languages can communicate each speaking their own native languages (receptive multilingualism). When two speakers communicate using receptive multilingualism, both the receptive and the productive proficiencies of the individual participants in the conversation determine the success of communication. On the one hand, communication can be compromised by a lack of knowledge on the part of the listener. Communication breakdowns may occur when the speaker uses an idiomatic expression or word unknown to the listener. On the other hand, problems can be found on the part of the speaker. Some speakers may be better at adapting their speech to the listeners than others (see Section 8.5).

So far, most research on receptive multilingualism has focused on testing the receptive proficiency of the listeners by presenting recorded speech stimuli of the target language to the listeners. The advantage of such a procedure is the experimental control that allows the researcher to vary variables systematically. However, since no communication occurs, the authenticity and similarity to real-life settings are low (ecological validity, see Sections 2.1.1.2 and 2.1.1.3). For the listener, there are no opportunities for negotiation of meaning or the use of communicative strategies, such as signaling a lack of understanding or correcting miscommunication (see Lindemann and Subtirelu 2013: 584). The speaker cannot check whether the listener has understood the message and reformulate it if this is not the case. It is currently largely unknown to what extent the productive and receptive proficiencies of the interactants each contribute to the common goal of communication. The success of receptive multilingualism depends on the participating interactants and is not simply the mean receptive proficiency of the two individuals.

Therefore, to model communication, it is not sufficient to test how well listeners understand the speakers' language. There is a need to create experimental situations where interactants communicate, for example, by solving a task together, each using their own language. The results can be analyzed quantitatively by counting the number of tasks carried out correctly in a certain amount of time. A more qualitative analysis can be carried out by looking at the strategies the participants use to communicate successfully. Below, a few of such communicative tasks are discussed.

2.2.4.1 Spot-the-differences task

The spot-the-differences task involves two interactants who each have a copy of a picture that displays a large number of objects arranged randomly. The two pictures differ in various aspects, including shapes, sizes, colors, and the presence or absence of particular objects. The participants cannot see each other's pictures, and their task is to work together to identify as many of these differences as possible as quickly as possible. Dependent variables are the number of differences correctly identified and the time taken to complete the task. The duration of the task is an indicator of the degree of communicative efficiency, with shorter durations denoting greater efficiency and longer durations denoting lower efficiency.[8]

8 For more information and examples of pictures that can be used for a spot-the-differences task see Baker and Hazan (2011).

Example 2.19: Spot-the-differences task

Van Mulken and Hendriks (2015) used spot-the-differences tasks to investigate the effectiveness of language use in computer-mediated communication between native speakers of German and Dutch. The communication took place using English as a lingua franca (ELF) or receptive multilingualism. For each task, they manipulated a picture containing a varied set of objects into two versions. To elicit a conversation, they chose objects with names likely to be unfamiliar to the participants, such as various toiletry items: tweezers, cotton buds, or panty liners (see Figure 2.10). For each task, the two participants were given two different versions of a picture with several objects. Each participant collaborated with a partner in an online chat to find ten differences between the pictures within a time limit of 15 minutes. The participants were instructed to complete the task in their own native language or ELF. The researchers also analyzed the communication qualitatively to gain insight into participants' communicative strategies to gain mutual understanding (see Section 4.6). The findings showed that the participants were more effective in realizing communicative goals in receptive multilingualism interactions than in ELF interactions and that participants made use of different communication strategies in the two kinds of interactions.

Figure 2.10: Pictures used for the spot-the-differences task by van Mulken and Hendriks (2015: 411).

2.2.4.2 Map task

A similar task is the map task, where one interactant is the instruction giver and the other the instruction follower (cf. Brown et al. 1984; Anderson et al. 1991; Grønnum 2009). Both interactants have a map with roads and landmarks. The giver's task is to instruct the follower on how to trace a route between two landmarks that are known to the giver but not to the follower. Introducing subtle differences between the two maps makes the task more challenging. After completing one task, the participants switch roles, and new maps are provided. The dependent variables are success rate and the time taken to complete the task. Example of maps developed for a map task is provided in the HCRC Map Task Corpus.[9]

9 https://groups.inf.ed.ac.uk/maptask/

Example 2.20: Map task

To investigate how well speakers of the two unrelated languages, Russian and Estonian, with receptive knowledge of each other's languages could communicate, Bahtina, ten Thije, and Wijnen (2013) used stylized maps that differed in several features (e.g., non-identical arrows to indicate movement in specific directions or gaps instead of connections, see Figure 2.11). These modifications forced participants to take the longest route to get from Point A to Point B. The participants were seated behind personal computers in separate rooms and connected via an online video link. One of the participants was assigned the role of instruction follower, and the other was assigned the role of instruction giver. The followers were instructed to explain their location (Point A), and the givers were to instruct the followers to get to their location (Point B). They were asked to use their mother tongues as much as possible. The results suggested that efficient communication is reached through the explicit negotiations aimed at establishing mutual understanding and that the level of proficiency in the non-native language did not strongly impact communicative success.

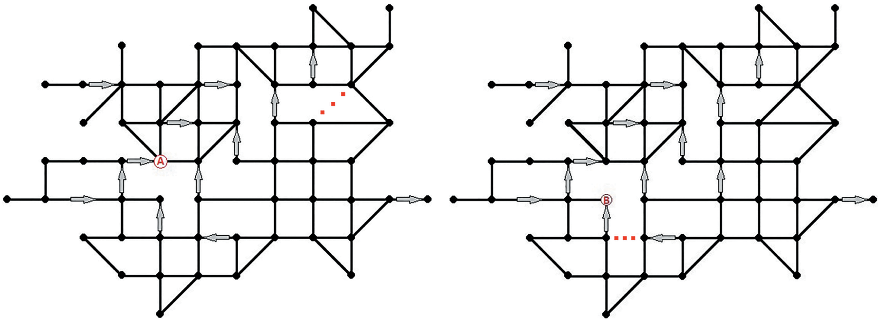

Figure 2.11: Stylized maps for a map task. From Bahtina, ten Thije, and Wijnen (2013: 166).

2.2.4.3 Observations

If we want to get more detailed qualitative insight into the strategies used by participants in a conversation to reach mutual understanding (see Section 4.6), it may be useful to observe real communicative situations. This is, first and foremost, a qualitative approach, but it is also possible to use the results as a basis for a quantitative analysis, for instance, by counting turn-takings, misunderstandings, reparations and length and frequency of pauses (Zeevaert 2004).

Examples 2.21: Observations

Börestam Uhlmann (1994) recorded around thirty inter-Scandinavian arranged conversations between Danes, Norwegians, and Swedes aged 18 to 25 who were unaccustomed to the neighboring languages. The material included conversations between speakers with different native languages as well as monolingual conversations. She was interested in which kind of strategies were used by the participants to enhance mutual comprehension, such as rephrasing, explaining, switching to English, repairing misunderstandings, and interrupting to either clarify something or make certain that the message had been correctly understood. While her data analysis was primarily qualitative, she also quantified the outcomes by counting the *number of repairs and misunderstandings that occurred. The results showed that the communicative flow was more often interrupted for repetition, clarification, confirmation, and paraphrase in, bilingual conversations than in conversations between speakers of the same language.*

In another observational investigation (Börestam 2015), the researcher investigated inter-Nordic communication in Iceland. In Iceland, Danish is taught as a foreign language. Since it is closely related to Swedish and Norwegian, Icelandic people are likely to understand these languages based on their knowledge of Danish (mediated intelligibility, see Section 1.1). Pretending to be tourists, fieldworkers who were native Danish, Norwegian, or Swedish speakers approached Icelandic adolescents, asking for directions to the Nordic House in Reykjavík in their mother tongues. The study, conducted across three different periods (1983, 1999/2004, and 2006), revealed a decline in the percentage of young individuals who comprehended the inquiry, dropping from approximately two-thirds in 1983 to just under 40% in 2006. Simultaneously, there was an increase in respondents opting to respond in English, with a significant majority (80%) doing so in 2006 compared to only a third approximately 25 years earlier, in 1983.

2.2.5 Recent methodologies

In most of the intelligibility tests discussed in this chapter, listeners are allowed time to consider their response, and only the result of the response, rather than the time taken to respond, is considered. Such tests have been employed more frequently than techniques that aim to gain insights into the listener's cognitive processes during comprehension, primarily because they do not require any specialized or costly equipment.

In the previous sections, reaction time measurements were mentioned a few times. Such measurements are claimed to offer an indirect measure of the degree of difficulty experienced in processing input. The assumption underlying reaction time measurements is that faster listener reactions indicate better intelligibility of the input being processed. Reaction time can be measured using software applications that register when a listener performs a certain action, such as a vocal response, pressing a button on the computer keyboard, or touching the computer screen. Tasks measuring reaction times, such as lexical decision tasks (see Section 2.2.1.3) or sentence verification tasks (see Section 2.2.2.4), can be employed to assess overall intelligibility. The measurements are temporally accurate to within a few milliseconds. The measuring point should be considered carefully. Generally, it makes most sense to measure from the disambiguation point, in particular, the point in a word where it can be distinguished from all other words or the point in a sentence where it can be decided whether the proposition is true or false (see Section 2.2.2.4). Stimuli are usually responded to faster with the dominant hand, the right hand for right-handed people and the left hand for left-handed people (Rastatter and Gallaher 1982; Shen and Franz 2005). It should be noted that reaction time measurements are very delicate. Many trials fail to record what they are intended to because of accidental keystrokes, lapses in the listener's attention, and distractions in the laboratory environment. Listeners should, therefore, be allowed ample practice on the task and be given breaks every once in a while. Another procedure for avoiding unwanted influence on reaction time scores is to normalize data by eliminating outliers, for instance, scores exceeding 2.5 standard deviations above a listener's mean reaction time.

Recently, web-based behavioral research is becoming more and more popular. This makes it possible for the researcher to collect large amounts of research data quickly, and programs for preparing and carrying out such research have become increasingly user-friendly. A unique problem with web-based testing is its reliance on listeners' hardware and software. When an experiment is carried out in the lab, the researcher can make sure that the same computer, stimulus software, and hardware are used for the whole response collection. In remote testing (over the internet), it is not possible to control all these conditions. Listeners may use their own computer (desktop, laptop, tablet, or even phone) with their own operating system and access experiments through a variety of web browsers. This may influence the measurements of reaction time. However, modern internet platforms now allow recording the response time with acceptable accuracy and precision (Anwyl-Irvine et al. 2021).

Several other new techniques for measuring cognitive processing during listening have become available in recent years (e.g., van Engen and McLaughlin 2018). Such measurements can serve as a valuable addition to functional intelligibility tests and subjective intelligibility estimates of listeners, as they allow for the detection of variations in listening effort that may not be apparent in the performance level of individuals in traditional functional tests. Eye-tracking measurements can detect the presence, attention, and focus of the users by determining what they are looking at and for how long their gaze is in a particular spot while listening to speech. It thereby provides information about processing difficulties and could be used to determine the moment of word recognition (fast/easy word recognition is the hallmark of intelligibility) by measuring eye fixation on pictures on the screen that correspond to the stimulus. The technique has been used to investigate processing among hearing-impaired patients (Wendt, Kollmeier, and Brand 2015). Since the difficulties encountered when listening to a closely related language can be compared to the challenge for hearing impaired when listening to their native language, it can be assumed that the technique could also be used for intelligibility measurements of closely related languages. As it does not require verbal responses, it is useful for testing various groups of listeners, including children and illiterate adults.

Example 2.22: Eye tracking

Kudera (2022) showed that eye-tracking can be used to measure how well speakers of four closely related Slavic languages (Bulgarian, Czech, Polish, and Russian) understand sentences in the other three non-native languages. He recorded eye fixations from participants who listened to sentences. The participants were instructed to listen to the sentences and look at a visual scene. Then, a pseudo-task involving answering a question in their native language regarding their understanding of the foreign sentence was given. Subjects were informed that pictures provided clues for sentence comprehension and were advised to pay attention to the objects on the screen. The moment of first fixation on a target picture of the direct object was the estimate of the difficulty of sentence processing. The measurements correlated significantly with surprisal scores, a measure of (un)expectedness of the sounds of words in one language given equivalents in another language (see Jagrova et al. 2019).

Pupillometry is a research method used in psychology studies to determine a participant's cognitive effort when listening to a stimulus by measuring the diameter of the pupil (the little black hole in the center of the eye). A dilated pupil and a longer interval between the stimulus onset and the time of maximal pupil dilation have been shown to correlate with lower intelligibility (Zekveld, Kramer, and Festen 2010). Even in situations where non-native listeners perform just as well as native listeners, they may still exert greater listening effort, which may result in increased fatigue and a reduced ability to perform multiple tasks simultaneously. The pupillometry technique can, therefore, be used to demonstrate subtle differences in cognitive effort in cross-linguistic intelligibility testing.

Example 2.23: Pupillometry

Borghini and Hazan (2018) compared the pupil response of 23 native English speakers and 27 Italian speakers of English as a second language. The researchers tested speech intelligibility of English sentences presented in quiet and in background noise at two performance levels. The results indicated that pupillometry is sensitive to differences both across listener types and across the levels of speech intelligibility. Importantly, the study revealed that pupillometry could uncover differences in listening effort even when those did not emerge in the performance level of individuals.

More immediate access to information processing can also be obtained from neurological techniques, such as event-related potentials (ERPs). This technique measures changes in the brain's electrical activity related to cognitive events. They are measured through electroencephalography (EEG) as changes in electrical charges in the scalp. The measurements can tell the researcher exactly when and approximately where in the brain decisions are being made by the listener. The measurements generally support those of behavioral studies. For instance, recognition of cognates is generally shown through a less pronounced peak in the negative wave in the 400 ms window (the N400 component, a negative-going change in the electrical charges related to semantic processing) than non-cognates. This can be explained by non-cognates requiring deeper semantic processing than cognates (Helms-Park and Dronjic 2016).

Example 2.24: ERP measurements

Goslin, Duffy, and Floccia (2012) used event-related potentials (ERPs) to investigate whether listeners process words spoken with a regional accent or native accent in the same way as word pronounced with a foreign accent. They recorded ERPs during the presentation of foreign and regional accented speech. They compared time epochs in the two most researched ERP components likely to be relevant to accent-related speech, the PMN (Phonological Mapping Negativity, 200–350 ms) and the N400 (350–600 ms) components. In the first epoch, significant differences in average amplitudes indicated that the PMN was greater for regional than native accents. Conversely, foreign accents led to a significant reduction in the PMN when compared to native or regional accented speech. In the later epoch, foreign accents elicited a significantly reduced N400 amplitude when compared to both other accent conditions. However, there was no longer any significant difference in amplitude between regional and native accents. These findings suggest that different strategies were adopted to process regional and foreign accents at a pre-lexical/phonological level.

Various software packages can be used to run the experiments described in this section. E-prime is probably the most commonly used software, but there are also several freely available options, e.g, PsychoPy. There are also options for running experiments over the internet, such as PsyToolkit, jsPsysch, oTree, and PsychoPy. Special equipment is generally still needed for some new techniques, such as eye tracking, pupillometry, and ERP measurements. This makes it difficult to carry out an experiment outside the lab, even though portable equipment has recently become available.

2.3 Cross-validation of methods

The overview of various methods for testing the intelligibility of closely related languages discussed in this chapter has made clear that choosing a method for an intelligibility project is not a straightforward task. The choice of the method to be used in an investigation depends on various practical considerations, such as budget constraints, time limitations, and the listeners' backgrounds. However, even with sufficient time, resources, and listeners willing and able to undergo complicated and lengthy tests, the choice of method should first and foremost be guided by the research question.

But apart from these considerations, does it matter which method is used? To shed more light on this point, we need to determine whether the same people who perform well in one test also attain high scores in another test while controlling for all other factors. Some researchers compared the results of different methods of measuring intelligibility. Such comparisons are informative, as they offer insight into the importance of choosing a specific method. As noted in the introduction to Section 2.2, it is not possible to compare the absolute level of the scores from tests that use different methods. Cross-validation of methods can, therefore, only be meaningfully done on the basis of correlation coefficients, not by comparing absolute score levels.

Maurud (1976) conducted a study to assess the mutual intelligibility of Scandinavian languages through word and content tests. The study involved soldiers from Denmark, Norway, and Sweden, who were required to read and listen to a text and translate several key words into their native language. The same text was used for both the word and content tests. Afterward, the participants also had to answer questions about the contents of the text. Maurud found correlations between the test results ranging from r = .6 to .8 for various groups of participants. This shows that different tests measure different aspects of intelligibility.

Doetjes (2007) investigated the effect of six different test types on the measurement of the intelligibility of Swedish for Danes. The same text was tested in six test conditions: true/false questions, multiple-choice questions, open questions, word translation, long summary, and short summary. Each test condition was presented to a different but comparable group of high school pupils in two small province towns in western Denmark. The percentage of correct responses decreased from 93.0% for the true/false questions to 66.2% for the short summaries. This shows that test results are strongly affected by the method chosen. There is no clear answer to the question of how well Danes understand Swedish.

Tang and van Heuven (2009) examined the mutual intelligibility of 15 Chinese dialects using functional intelligibility tests at both the word and sentence levels (see Example 2.7 and Section 6.2.2). They compared the outcomes of the tests with each other and with opinion scores. The correlation between word intelligibility and sentence intelligibility was high, with a value of r = .9. They found correlations between the opinion scores and the functional scores ranging from r = .7 to .8. The authors' conclusion was that functional sentence intelligibility tests should be the preferred method for assessing mutual intelligibility.

Gooskens and van Heuven (2017) administered six functional intelligibility tests involving 70 combinations of closely related pairs of European target languages and listener languages, see Chapter 3. The tests were a word recognition test, a cloze test at the sentence level, and a picture-to-story matching task at the paragraph level. Also, tests were presented in spoken as well as written mode. The scores on these functional tests were compared with each other and with measures obtained through opinion testing, i.e., judged and perceived intelligibility. The various tests revealed similar patterns of intelligibility (correlations between r = .71 and r = .79 for tests in the same mode). The correlations with the picture-to-story matching scores (which suffered from ceiling effects) were generally lower than with the cloze tests and word translation scores. There was a strong correlation between both measures of intelligibility based on opinion testing (judged and perceived), with a value of $r = .99$. These measures also correlated highly with the functional test scores, particularly with the cloze test, with correlations ranging from $r = .94$ to $r = .99$.

Čéplö et al. (2016) tested the mutual intelligibility of spoken Maltese, Tunisian Arabic, and Benghazi Libyan Arabic using three listening tests, a word test where the listeners were asked to perform a semantic classification task with 11 semantic categories (see Figure 2.8), a sentence translation test, and a multiple-choice comprehension test at the text level. The correlations between the individual test results of the three tests were very low and sometimes even negative. This suggests that listeners' performance in either test is not a good predictor of their performance in the other and that the three tests constitute cognitively different tasks.

While comparisons of various tests generally show relatively high correlations, a large amount of unexplained variance is still left. Even though there is considerable overlap, different tests measure different aspects of intelligibility and some tests are less sensitive to interfering factors than others. Therefore, the best approach may be to have listeners carry out more than one task to confirm results or illuminate distinct aspects of the same phenomenon.

It can be concluded that it is not possible to say how much of a language a listener understands. To illustrate, a result of 70% correct answers in a test does not mean that the listener understands 70% of the language. The percentages of correct answers depend on the difficulty of the test material and the test used. Therefore, the same test method must be used to compare the intelligibility of different languages.

In addition, different levels of intelligibility are needed for different communication purposes. In some situations, only a very basic level of understanding is needed, while in others, it is essential to reach a high level of understanding of complex speech. Therefore, at present, it is not possible to give a definite answer to the question of how well speakers of language A understand target language B. Consequently, caution should be exercised when comparing the results of different investigations. So far, it has not been possible to develop one universal test that could be used to test and compare the intelligibility of languages pairs worldwide. Such a test would be instrumental but also difficult to realize because the results of intelligibility tests depend on many factors, as has become apparent from the overview in this chapter and as will be discussed in the rest of the book.

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.

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)

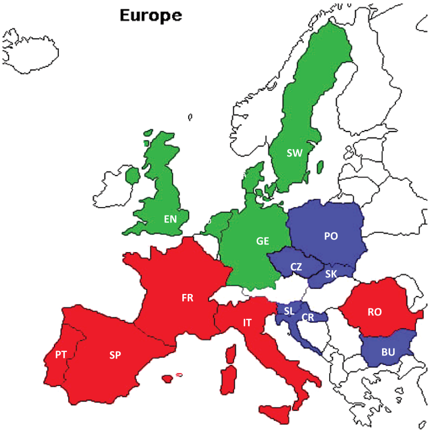

Figure 3.1: Languages included in the MICReLa investigation. Source: Gooskens and van Heuven (2017: 29).

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).

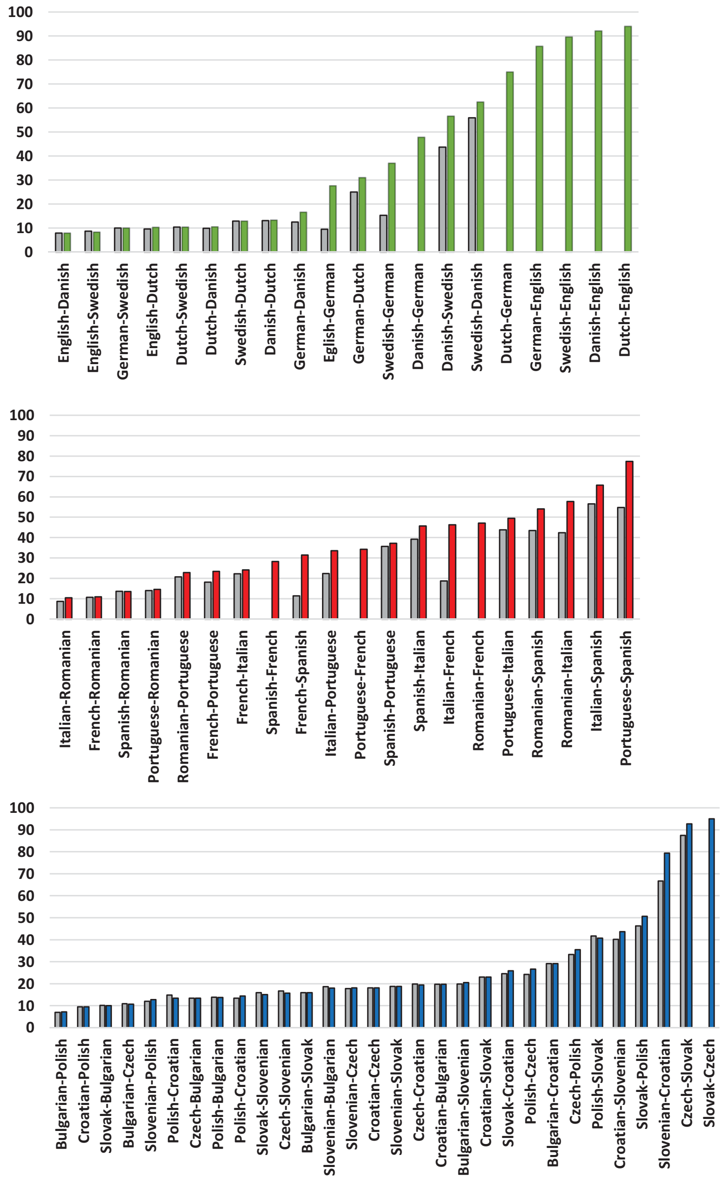

Figure 3.2: Results of spoken cloze tests in the Germanic (top), Romance (middle) and Slavic (bottom) language areas. For each language combination, the listener language is given first and the target language second (e.g., "English-Danish" means "English native speakers listening to Danish"). Colored 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.

Chapter 4
Extra-linguistic determinants of intelligibility

Chapter 2 discussed various methods for measuring how well listeners can understand closely related languages. Once the level of intelligibility has been established, the next step is to interpret the results and understand what factors determine how well a certain group of listeners can understand a particular language. Knowledge and awareness of such factors may help to improve mutual intelligibility. Of course, linguistic overlap between the listener language and the target language plays a significant role. Without a certain degree of overlap, listeners are unlikely to understand a non-native language, unless they have learned the language during formal language instruction, or have been exposed to the language in some other way. The role of linguistic determinants is discussed in detail in Chapter 5.

However, since intelligibility measurements are based on experiments involving human participants, various extra-linguistic factors also influence the comprehension level. Speakers and listeners have different backgrounds, properties, competencies, and personality traits, which may influence their level of mutual intelligibility. This chapter discusses extra-linguistic factors that have been shown to influence intelligibility. These factors are shown in the blue square of Figure 1.1 in Chapter 1.

Some extra-linguistic factors in this chapter show a general trend for specific groups of listeners and may explain differences found at the language level. For example, while the amount of exposure to a language may differ per individual listener, a general trend can often still be observed for a whole group. The average amount of exposure across an entire group of listeners can then predict how well listeners with a specific language background can understand a specific language. Other factors can be assumed to be more individual and, therefore, mainly explain differences in intelligibility at the listener level. Individual differences, such as intelligence, will likely explain why some listeners perform better than others. However, such individual traits are less likely to explain the level of intelligibility at the language level. For example, listeners with one specific language background are not expected to be more intelligent overall than listeners with another language background.

Sections 4.1 to 4.6 will show how some relevant extra-linguistic factors can be quantified so that they can be correlated with the results of intelligibility tests. This will be followed by a discussion of how to exclude the influence of extra-linguistic factors on intelligibility measurements (Section 4.7), which is important if the main interest of the researcher is to investigate the relationship between inherent intelligibility and linguistic factors (see Chapter 5). It is difficult to completely exclude all

https://doi.org/10.1515/9783111134697-004

extra-linguistic factors in a test situation, but there are ways to minimize their influence on intelligibility scores.

4.1 Exposure

The most obvious extra-linguistic factor that may determine the intelligibility of a non-native language is the nature and amount of previous exposure the listeners have had to it. The more exposure listeners have to the target language, the more likely they are to understand it because they will have learned some of the vocabulary of the language receptively. Depending on the kind of input listeners receive, they typically pick up words that they have encountered eight times or more (de Wilde, Brysbaert, and Eyckmans 2020: 352).

When exposed to a language, listeners will also become familiar with the sounds and develop contrastive awareness (Verschik 2012). They become aware of regular sound correspondences between their native language and the target language that they can use to recognize cognate words, i.e., historically related words (see Section 1.1). To illustrate, once German listeners discover that a Dutch /t/ often corresponds to a German /ʦ/ like in Dutch *tijd* /tɛit/ – German *Zeit* /ʦait/ 'time', they can generalize this knowledge to other words such as Dutch *tekenen* /tekənə/ – German *zeichnen* /ʦaixnən/ 'draw' and Dutch *twee* /tʋeː/ German *zwei* /ʦʋai/ 'two' (see Section 5.2). Vanhove (2016) had German participants translate Dutch written cognates containing either the digraph ‹oe› (corresponding to a German word with ‹u›) or the digraph ‹ij› (corresponding to German ‹ei›). The participants were provided with feedback in the form of the correct translation. In the second part of the task, they were presented with other Dutch cognates containing ‹oe› and ‹ij›. Participants who encountered Dutch cognates containing ‹oe› in the first part were more likely to translate German cognates containing ‹u› correctly than German cognates containing <ei>, and participants who encountered Dutch cognates containing ‹ij› in the first part were more likely to translate German cognates containing ‹ei› correctly than German cognates containing <u>. This showed that correspondence rule learning had taken place. The same processes can be assumed to play a role in spoken word recognition.

After some exposure, syntactic and morphological differences between the languages will also be less disruptive to understanding a closely related language than when such differences are encountered for the first time (Gass and Varonis 1984).

Milliken and Milliken (1996) note that the inherent learnability, i.e., the ease with which listeners can adjust to the other speech variety, may vary, depending on the listener language and the target language. Even if the inherent intelligibility of two languages is equal at first, it may be easier for a listener to discover phonological correspondences in one of the languages than in the other (see also the discussion of Cheng's "systemic mutual intelligibility" measure in Section 5.2). They may, therefore, need more exposure to a language with low inherent learnability than with high inherent learnability to reach a level of intelligibility that is sufficient for communication. Some researchers have suggested the inherent learnability of a language as a distance measure between languages (Chiswick and Miller 2005; van der Slik 2010).

In addition to the improved understanding due to the listener's increased knowledge about differences and correspondences between their own language and the target language, exposure to the language may also make listeners conscious of their ability to understand the language. Lack of exposure and familiarity with a language might make listeners apprehensive about their own skills, which could lead to their not even trying to understand because they are convinced that they will fail. Exposure and explicit training may change their attitude and make them open to understanding and communicating with speakers of closely related languages (see Section 4.3 on the role of attitudes). It is possible that this effect even works cross-linguistically. The boosted confidence listeners get when discovering that they can understand a language they do not speak may result in the willingness to try to understand even more languages they do not speak.

Exposure can occur in various formal and informal settings, including via personal contact, music, television, and social media, or in the written form via newspapers, books, etc. The listener will often discover correspondences and learn words by being exposed to the language (de Wilde, Brysbaert, and Eyckmans 2020). However, a listener can also be exposed to a language through formal schooling. Past studies have demonstrated that for closely related languages, a brief language course that raises awareness of the key differences and similarities between the speaker's native language and the target language can significantly enhance receptive proficiency. In an investigation by Golubović (2016), Czech native speakers received four and a half hours of instruction in understanding Croatian: they learned the 60 most frequent non-cognates, common phoneme correspondences, and syntactic similarities between Czech and Croatian. In addition, they practiced translating short stories from Croatian into Czech and inferring meaning based on what they had learned. Their level of understanding of Croatian was tested before and after the instruction. The results of this experimental group were then compared to the results of a control group that got no instruction. The results showed that the experimental group improved significantly more (a doubling of their score from 2.8 to 5.7 on a test where the maximum score was 10) than the control group (with a much smaller improvement of 27%, representing a learning effect from participating in the pre-test). Hedquist (1985) got similar results when teaching Swedish to native Dutch speakers.

4.1.1 How to measure exposure

Exposure can be assessed in several ways. A common method is using a questionnaire where listeners rate their level of exposure to the target language on various scales. This may include exposure through reading books and newspapers, watching television, interacting with speakers in person, and other similar activities (Delsing and Lundin Åkesson 2005; Gooskens and van Heuven 2020). It is also essential to know if the listeners have had formal instruction in the language. The amount of instruction can be expressed as the number of years of instruction, possibly supplemented by the number of weekly hours for greater precision. An online questionnaire can easily be developed with (freely available) survey software (see Section 2.1.2.2). An example of exposure questions in a background questionnaire is shown in Figure 4.1 (from the MICReLa project, see Chapter 3).

How often (on average) during the last 5 years have you used [language X] in the following situations?

Listening to people speaking in your presence (e.g., on vacation, at work, doing shopping):

Never	0	0	0	0	0	Every day

Watching television, movies, listening to the radio:

Never	0	0	0	0	0	Every day

Playing computer games:

Never	0	0	0	0	0	Every day

Chatting or surfing on the internet:

Never	0	0	0	0	0	Every day

Talking to people in your presence/on the telephone/on Skype

Never	0	0	0	0	0	Every day

Reading books, papers, magazines, and text on a computer screen:

Never	0	0	0	0	0	Every day

I have studied [language X] for ___________ years

Figure 4.1: Exposure questions from a background questionnaire used in the MICReLa project (see Chapter 3).

Figure 4.2 shows the relationship between the results of the spoken cloze test in the MICReLa project and the mean exposure scores across these six scales (see Appendix C). The correlations were significant ($p < .01$) and high, between $r = .87$ for Romance and .93 for Germanic languages. When all 70 language combinations were involved, the correlation was .90. This shows the importance of exposure for intelligibility. Correlation between intelligibility and number of years the lis-

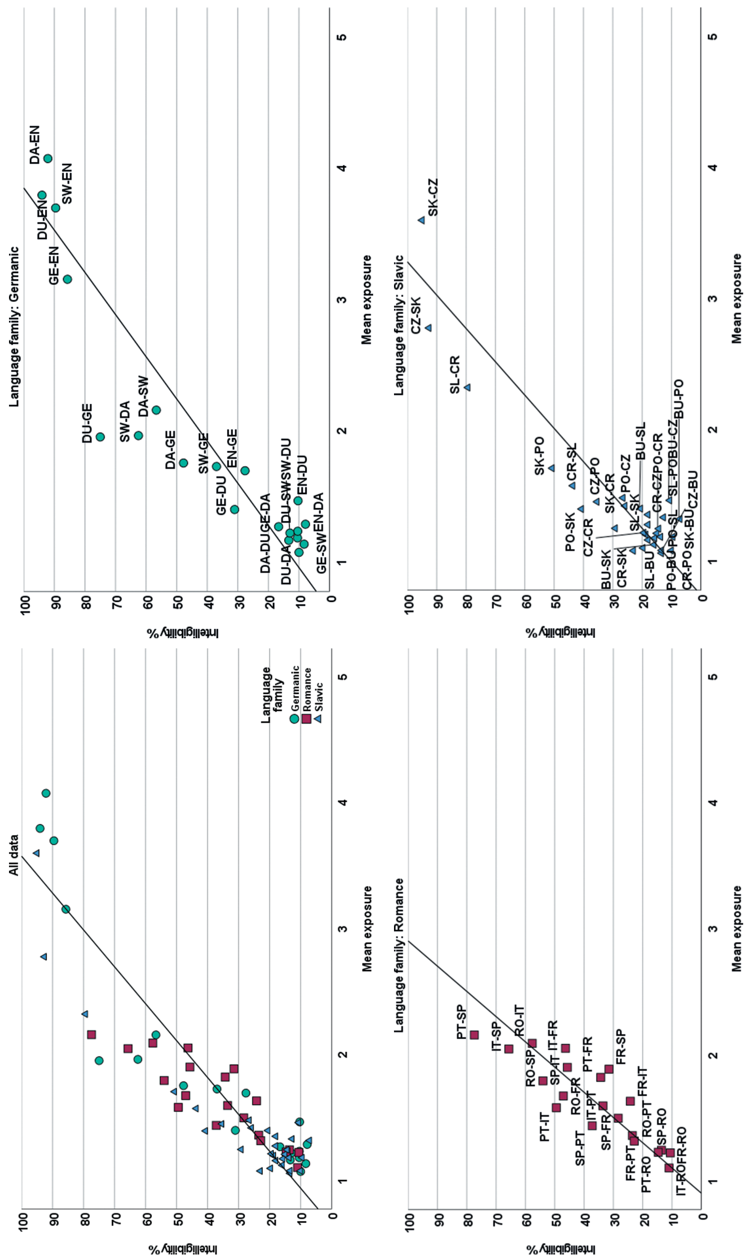

Figure 4.2: Correlation between intelligibility scores and mean exposure scores for all 70 European language pairs in the MICReLa project (r = .90) and separately for Germanic (r = .93), Romance (r = .87), and Slavic (r = .92). The abbreviations refer to the language varieties involved (see Figure 3.1). The first variety in a pair is the language of the listener, the second is the target language.

teners learned the target language was significant for the Germanic group only, presumably because the target languages in the Romance and Slavic language families are mostly not taught in school, while all Germanic listeners had learned English and many had learned German as a foreign language.

Another method of estimating exposure is by computing the percentage of non-cognates that a listener can comprehend. The underlying assumption is that listeners can only comprehend non-cognates if they have had prior exposure to the language. Thus, listeners with limited exposure to a language are likely to correctly translate fewer non-cognates than those with greater previous exposure. The non-cognates should be selected with care so that the listeners cannot translate them because of their knowledge of a third language. As an example, Danish listeners who know French may understand Swedish *fåtölj* [fɔ'tø̞lj] 'armchair', which is a loanword from French *fauteuil* [fo'tœj], even though the Danish equivalent is a non-cognate (*lænestol*). Moreover, non-cognates that are generally known even by people who rarely come into contact with the language should be avoided. To illustrate, Kürschner, Gooskens, and van Bezooijen (2008), see Example 2.1 in Section 2.2.1.1, found that Danes with very little previous exposure to Swedish hardly translated any of the non-cognates in the translation task correctly. An exception was the word *flicka* 'girl' (Danish *pige*), which was translated correctly by 68% of the listeners. This word is probably known to most Danes as a stereotypical Swedish word, and if listeners know this word, it does not necessarily mean that they have had more than very sporadic exposure to the language. In hindsight, it might, therefore, have been better not to include this particular non-cognate to measure exposure. Figure 4.3 shows another example of an investigation where the percentage of correctly translated non-cognates was used to quantify exposure. Gooskens and Schneider (2019) tested mutual intelligibility among speakers of closely related language varieties of Pentecost, one of the Pacific islands of Vanuatu. The results showed a high correlation (r = .92) between intelligibility tested by means of a word translation task (cognates only) and exposure expressed as the percentages of correctly translated non-cognates.

A third way to quantify exposure is by considering the geographic proximity between the listener's native language and the target language. Listeners tend to encounter languages spoken in geographically proximate regions more frequently than those spoken in more distant places. Thus, geographic distances can be used as a measure of exposure. It can be calculated as the straight-line distance in kilometers between the listener's native language and the target language ('as the crow flies'). Bender and Cooper (1971) tested mutual intelligibility within the Sidamo subgroup of East Cushitic (spoken in southern Ethiopia) by asking listeners questions about pre-recorded stories. The intelligibility results showed a strong relationship with various linguistic distance measurements, but also with the geographic distan-

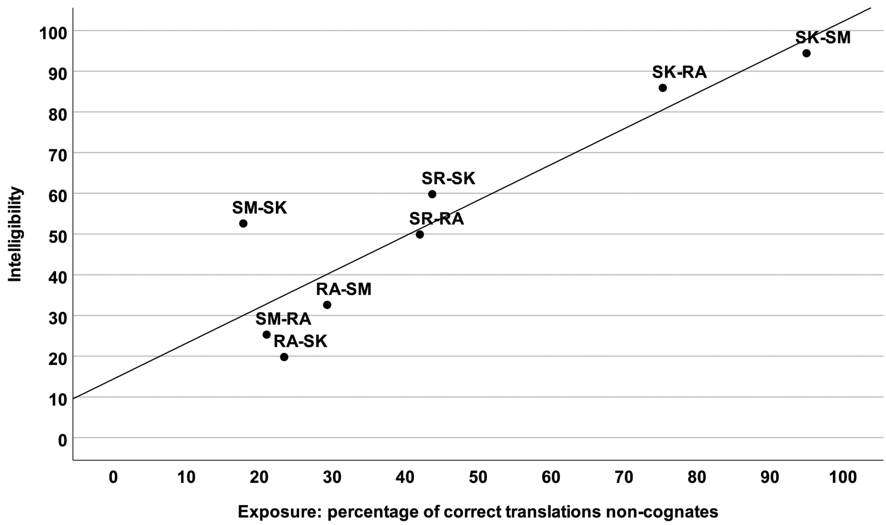

Figure 4.3: Exposure expressed as the percentages of correctly translated non-cognates and intelligibility measured in a word translation task (r = .92). The abbreviations refer to the language varieties involved (RA = Raga, SK = Suru Kavian, SM = Suru Mwerani, SR = Rabwanga), see also the map in Figure 6.2. The first variety in a pair is the listener language, and the second is the target language. Adapted from Gooskens and Schneider (2019: 78).

ces between main population centers of the languages (r = −.67). An even better reflection of the amount of contact with a variety may be travel distances expressed in the number of kilometers along the road or the time it takes to travel to a place by car or by public transport because it takes into account barriers to interaction, such as mountains, as well as boosters of interaction, such as roads and rivers. In an investigation involving 15 Norwegian dialects, travel distances from the year 1900 were found to correlate even better with phonetic and lexical distances than straight-line distances and modern travel distances (which are often an almost straight line), an improvement from r = .41 to r = .54 (Gooskens 2005). This shows that old traveling circumstances are reflected in linguistic distances between modern dialects. Travel distances are also likely to correlate significantly with intelligibility scores.

Fourth, demographic information, such as the population size, population density, or the number of community facilities (e.g., churches, shops, administrative headquarters, schools, factories, hospitals) and intermarriages, may predict how likely a listener is to be exposed to the target language (Simons 1979). The more speakers of a language, the denser the population, the more intermarriages, and the more facilities offered by the speakers of the target language, the greater the likelihood of interaction involving it.

Finally, people are likely to have been exposed to a neighboring language more often as they grow older simply because they have had more opportunities to hear and read it. Changes in curricula and media languages also play a role here. For example, the availability of more TV channels nowadays means that younger Dutch participants often hardly watch German TV channels, whereas these have constituted a considerable part of the popular TV channels (at least in the border provinces) up until the 1980s or 90s. In addition, while the majority of the older generations learned German at school, a large part of the younger generation have dropped German as a school language. Therefore, by asking listeners to indicate their age, the researcher might gain some knowledge about the amount of exposure they have had to a language. Indeed, various investigations have found that older listeners perform better in intelligibility tests than younger listeners (Delsing and Lundin Åkesson 2005; Härmävaara and Gooskens 2019). However, as we will see in Section 4.5, the relationship between intelligibility and age is not straightforward since several factors interact.

4.2 Plurilingual resources

Many listeners have knowledge of more languages or dialects than their own native language. The words that an individual knows in various languages are organized in a complex and dynamic way in the mental lexicon that contains information about the words' meaning, pronunciation, syntactic characteristics, etc. A number of factors determine which languages are activated and how. Mieszkowska and Otwinowska (2015) provide an overview of such factors. For instance, languages used more often and more recently are more easily activated than languages used less frequently and less recently. Additionally, languages that are perceived to be linguistically more similar to the native language are more easily activated, and languages in which the speaker has a higher level of proficiency are also more likely to be activated.

The mental lexicons of a multilingual speaker do not consist of separate compartments for each language. The lexicons are intertwined and can interact with each other and, as such, can be activated at the same time (Dijkstra 2003; de Bot 2004; Szubko-Sitarek 2011). This plurilingual knowledge can often be used to understand a closely related language. Listeners may understand some non-cognate words because they are loanwords from a so-called bridge language (another language they are familiar with). For instance, many Dutch people learn some German in school and can use this knowledge to understand some Danish words that are borrowed from German but are not Dutch cognates. Hence, Dutch speakers may be able to correctly translate the Danish word *bogstav* 'letter' (in the alphabet) into the Dutch non-cognate *letter,* through the German cognate *Buchstabe*.

Swarte, Schüppert, and Gooskens (2015) had native Dutch speakers (all university students) translate written and spoken Danish words into Dutch. The students had no previous knowledge of Danish and different proficiency levels in German. Examples of stimulus words are shown in Table 4.1. The words either only had cognates in German (such as Danish *bogstav,* German *Buchstabe,* Dutch *letter*) or only in Dutch (e.g., Danish *vante,* Dutch *want,* German *Handschuh*, 'mitten'). The results showed that participants with a high level of German performed better in this translation task (55% correct in the written task and 29% in the spoken task) than participants with a low level (30% and 17%). This shows that knowledge of German can help native Dutch speakers decode written and spoken Danish. This result can be generalized to many other language situations. Therefore, to interpret intelligibility results, it is helpful to gain knowledge about the linguistic background of the listeners.

Table 4.1: A target word *(bogstav)* that can be recognized by Dutch listeners by means of a cognate in the bridge language (German) and a target word *(vante)* that can be recognized because it is a cognate in the listeners' native language (Dutch). Cognates are bolded. Source: Swarte, Schüppert, and Gooskens (2015: Appendix A).

Danish (target language)	Dutch (listener language)	German (bridge language)	meaning
bogstav	*letter*	***Buchstabe***	'letter'
vante	***want***	*Handschuh*	'mitten'

The results of Swarte, Schüppert, and Gooskens (2015) also showed that the participants with a high level of German translated even more Danish words correctly if the words had a German cognate (see example 'letter' in Table 4.1) than if they had a Dutch cognate (example 'mitten'), while for the participants with a low level of German, the words with a Dutch cognate were translated correctly more often. This provides evidence for the existence of a foreign language mode, i.e., a mode where a foreign language is more highly activated than the native language with a resulting higher transfer from the foreign language than from the native language (Selinker and Baumgartner-Cohen 1995: Dewaele 1998; Fuller 1999). Among the highly proficient participants, German seems to be more highly activated than the native Dutch language.

Wenzel (2018) discusses possible explanations for the fact that the native language is a source of transfer in some cases and the foreign language in other cases. Some explanations are based on similarities between languages, while in others, the dominance of the native or the foreign language within the learner's mental network of linguistic knowledge is emphasized. Wenzel investigated German learners of Dutch with English as their second language and English learners of Dutch with German as their second language in a grammaticality judgment and correction task designed to get insights into intuitions about the third language (Dutch in this case). The native German subjects with English as a second language made more errors than the native English subjects with German as a second language, even though the Dutch structures in the test corresponded to that of the Germans' native language. This means that there was more transfer from the participants' second languages when the second and third languages were more closely related and more similar to each other than the first and third languages. Wenzel took this as evidence that the second language has a special status in the acquisition of a third language, possibly caused by the more explicit manner of learning the second language. This conclusion thereby confirmed predictions by Bardel and Falk (2007).

Smidfelt (2018), on the other hand, had multilingual native Swedish speakers read and decode text in Italian, an unknown language to them. The results showed that all the previously acquired languages of the participants (Swedish, English, German, French, and Spanish) were activated and used to infer the meaning of the words in the texts and were nearly equally helpful for comprehension. This supports a view of non-selective access to the mental lexicon of multilinguals and shows that comprehension can be aided by knowledge about languages within the same language family as the target language as well as by knowledge about languages that are less closely related.

The EuroCom project (e.g., Klein and Stegmann 2000; Hufeisen and Marx 2007) is based on the principle that learners of a new language can benefit from their prior knowledge of related languages. The project focuses on multilingual strategies and the observation that listeners who know other languages can apply their knowledge of those languages to help them understand an unknown language that is closely related to one of the languages they already know. To illustrate, German native speakers who have studied Italian can use their knowledge of Italian to understand other Romance languages, such as Spanish and Portuguese (mediated intelligibility, see Section 1.1). Also their knowledge of German and English may be helpful, particularly in the case of loanwords from these languages in the Romance languages.

Multilinguals tend to have a higher level of multilingual awareness and metalinguistic competence. Therefore, multilinguals are generally better able than monolinguals to use cross-linguistic similarity to understand a language. Berthele (2008, 2011) shows how people can use their linguistic knowledge to guess the meaning of written cognates in a related but unknown language (inferencing strategies). He investigated how well native German, French or Italian speakers could decode cognates from other Germanic or Romance languages with which they had no prior experience. His results showed that the more languages the participants knew, the better they were at cognate recognition. The competencies for good guessing capacities that he mentions are greater perceptual tolerance and a wider search space when dealing with linguistic differences from the native language. Multilinguals can make a flexible and selective comparison of features and patterns. They can concentrate on consonants while disregarding the vowels and utilize contextual cues to make informed decisions. Otwinowska-Kasztelanic (2011) showed that multilingualism correlated positively with the awareness of cognates among Polish learners of English. In a questionnaire set up to measure the respondents' awareness of cognate vocabulary, only multilingual learners proficient in several languages tended to notice and make conscious use of cognates, thereby showing a positive relationship between multilingualism and the awareness of cognates.

4.2.1 How to measure plurilingual resources

The results of research on plurilingual resources show that it is crucial to control for the language background of the listeners when carrying out an intelligibility experiment. This can be done by asking questions in a questionnaire (see Section 2.1.2.2) to collect information about the number and kind of languages the listeners have learned actively and passively in the written and spoken form and for how long. Listeners can also be asked to evaluate their own passive (reading, listening) and active (writing, speaking) language skills using a Likert-like scale (e.g., from 1 = very poor skills to 6 excellent skills). The language skill in one language can be expressed as the total of all skills in that language. A multilingualism skill can be calculated as the aggregated total of reported skills in all languages (Lambelet and Mauron 2017).

If knowledge about proficiency levels is required, it is also possible to have the listeners carry out functional language tests. However, it may not be feasible to prepare tests for all languages that the listeners know.

4.3 Attitude

Listeners' attitudes toward a language, its speakers, and the country where it is spoken are often mentioned as a determinant of intelligibility. Language attitude research shows that people have strong and consistent aesthetic associations with different languages (Giles and Niedzielski 1998). For example, many people feel that French and Italian are beautiful languages and that German and Dutch are unattractive. If listeners have positive feelings about a language and its speakers, they may be more motivated to make an effort to understand them. However, experimental support for the relationship between attitude and intelligibility has been rather weak, likely due to the difficulty of eliciting (subconscious) attitudes in experimental settings. People might be more likely to have negative attitudes toward a language or work less hard to understand it in real-life encounters than under laboratory conditions. Still, significant correlations between attitude and intelligibility have been found in some studies (Delsing and Lundin Åkesson 2005; Gooskens and van Bezooijen 2006; Impe 2010; Schüppert, Hilton, and Gooskens 2015; McDonough et al. 2022). Kumove (2020) showed that communication problems caused by language differences can explain low levels of trust between countries and that greater communication potential is associated with greater trust between countries. He established an "index of semi-communication" (i.e., the probability that any two randomly selected persons from each country are able to engage in receptive multilingualism) between 21 European countries (in total 359 pairs of countries). The results showed a significant correlation of $r = .138$ ($p < 0.001$) with measurements of the level of trust gained from a survey asking respondents in the countries how much they trust the other European nationalities ("How much trust do you have in Italians?", "How much trust do you have in Poles?" and so on). The relationship was not entirely linear. While having high communication potential appears to equate to high trust, low communication potential does not necessarily equate to low trust.

Attitudes toward the language of the speaker can have various sources. It is generally agreed that attitudes toward languages reflect stereotypes, imposed norms, previous experience, or social connotations, i.e, knowledge about the languages and their speakers. See Edwards (1999) for an overview. Giles and Niedzielski (1998) mention how positive or negative feelings concerning the speakers and the countries where the languages are spoken may cause speakers of English to perceive Romance languages like Italian, French, and Spanish as beautiful languages, while German or Arabic are seen as less attractive.

Inherent language qualities may also influence language attitudes. Certain languages or language varieties may possess more appealing sound characteristics than others, contributing to their intrinsic aesthetic appeal (phonesthetics, see Crystal 2009). Hilton et al. (2022) examined the language attitudes towards a Swedish and Danish balanced bilingual speaker (see Section 2.1.1.5) in a matched-guise test (see Section 4.3.1) held with Chinese listeners who had no previous exposure to Scandinavian languages and were unable to recognize the languages. The idea was that imposed norms and social connotations could not have influenced this group of listeners since they did not have the relevant sociolinguistic knowledge. Still, the Chinese listeners rated Swedish more positively than Danish on all seven scales, showing that they evaluated Danish as less aesthetically pleasing than Swedish. It was concluded that the more positive evaluations of Swedish that have often been found in the literature among listeners familiar with the languages can only partially be explained by asymmetries in cultural esteem or intelligibility. They also seem to be triggered by linguistic features. In a follow-up experiment, the intonation contours were removed from the speech signal by monotonizing the recordings. As a result, the difference in evaluative responses by the Chinese listeners to the Danish and Swedish speech samples disappeared. This suggests that the difference in Danish and Swedish intonation is the cause of the different evaluations among Chinese listeners. Chinese listeners may prefer Swedish to Danish because Swedish and Chinese are both tonal languages, while Danish is not. Studies on aesthetic evaluations in other scientific areas have found evidence for positive effects of familiarity, for example, in music (Zissman and Neimark 1990).

The findings by Hilton et al. (2022) provide evidence that languages have linguistic features that sound more beautiful to listeners and do not have a source in imposed norms, stereotypes, or social connotations. This confirms research by van Bezooijen (2002) who had Dutch listeners aesthetically evaluate various European languages. Van Bezooijen also asked phoneticians to rate the languages on global phonetic scales. The attributed degree of beauty proved largely predictable from a combination of judged “melodiousness” and “softness”. The aesthetic evaluations were positively correlated with faster speech rates, precise articulation, and fronted articulation. These outcomes suggest that aesthetic evaluations have a phonetic basis. Similarly, Reiterer et al. (2020) found a link between language attractiveness and phonetic factors. In a rating study, they had Central European listeners listen to 16 auditorily presented European languages and report their perceptions in terms of 22 binary characteristics. They observed that many factors, in a complex interplay, explain people’s aesthetic preferences. Language preferences seemed to be influenced mainly by societal and individual cognitive factors but also by universal phonetic factors, such as sonority, timing properties, and syllable patterning.

Establishing cause and effect in the relationship between attitudes and intelligibility is not a simple task. Several studies have reasoned that attitudes may affect the listener’s willingness and motivation to understand another language (e.g., Boets and De Schutter 1977; Börestam 1987; Lambelet and Mauron 2017). If listeners are positive towards a language, they may be more willing to make an effort to understand the speaker. The level of linguistic similarity between languages may not help facilitating communication if individuals lack the motivation to understand one another, as noted by Wolff (1959) and Giles and Niedzielski

(1998: 87). Negative attitudes or social stigmas attached to languages are, therefore, often seen as a potential obstacle to successful communication between speakers of different languages. Kang and Rubin (2009) report evidence that when listeners mistakenly believe they are listening to a nonnative accent, they perceive the speech as accented and have greater difficulties understanding the speaker. This effect of stereotyping can likely be generalized to the intelligibility of a closely related language.

Causality may also be reversed, such that people are more negative towards varieties that they find difficult to understand. Dragojevic and Giles (2016) found that a noisier surrounding makes listeners evaluate speech more negatively, possibly due to the increased difficulty in processing speech. Lev-Ari and Keysar (2010) found that low intelligibility caused non-native speakers to sound less credible. English listeners rated the veracity of trivia statements, such as “Ants don’t sleep”, on 14 cm lines, with one pole labeled “definitely false” and the other “definitely true”. Half of the statements were true. The listeners rated the sentences as less true when spoken by a non-native (mean 6.95 cm for mild accents, 6.84 for heavy accents) than by a native speaker (7.59). The listeners were told that the speakers were merely repeating a message from a person originally spoken by native speakers of English. Therefore, this effect was assumed not to be due to stereotypes or prejudice against foreigners, but rather, it must be a result of reduced intelligibility of non-native speakers.

4.3.1 How to measure attitudes

To measure language attitudes, listeners are often provided with rating questionnaires consisting of semantic-differential scales (Osgood, Suci, and Tannenbaum 1957). Such scales allow a respondent to express a judgment, using a scale with typically five to seven points on a continuum between two contrasting adjectives. They can, for instance, be asked to evaluate the beauty of languages on a scale from “ugly” to “beautiful”. To make the listeners less conscious that they are rating languages, they can instead be asked to state their opinion about the speaker on scales where two bipolar adjectives are the extreme values. An example of such scales is shown in Figure 4.4 (from Hilton et al. 2022: 38). The adjective pairs used here can be classified into the three categories that represent the three most important dimensions in the framework for language attitude testing in Zahn and Hopper (1985): dynamism (“old-fashioned – modern” and “strange – normal”), attractiveness (“unattractive – attractive” and “unfriendly – friendly”) and superiority (“stupid – smart” and “poor – rich”). For more discussion of rating tasks, see, e.g., Drager (2018: 59ff) and Kirchner (2022).

What impression does this speaker make?						
	1	**2**	**3**	**4**	**5**	
old-fashioned	0	0	0	0	0	modern
stupid	0	0	0	0	0	smart
unattractive	0	0	0	0	0	attractive
strange	0	0	0	0	0	normal
unfriendly	0	0	0	0	0	friendly
poor	0	0	0	0	0	rich

Figure 4.4: Example of bipolar attitude scales (from Hilton et al. 2022: 38).

Participants can be asked to rate the beauty of languages based on their knowledge about the languages without listening to them. Such ratings are often based on previous experience with or knowledge about the variety as well as stereotypical opinions about the languages. However, the listeners may not always be familiar with the languages they rate and, as explained above, they may base their evaluation not only on experience and stereotypes but also on the sound of the language itself. Therefore, the listeners are often played sound fragments to be sure that they are familiar with all languages and can base their ratings on their sound. It may be a good idea to ask the participants whether they recognized the language that they listened to or to tell them what languages they heard to get an indication of whether they based their ratings on the intended language.

As part of the MICReLa project (see Chapter 3), Swarte (2016) presented two attitude tests to listeners from five Germanic countries. First, they were asked to indicate how beautiful they found five Germanic languages on a scale from 1 ("ugly") to 5 ("beautiful"). Next, they listened to short fragments before judging the beauty of the languages. The two attitude measurements correlated significantly (r = .76) but not perfectly. This shows that it makes a difference whether listeners base their judgments on sound fragments or not.

Gooskens and van Heuven (2020) used attitudes based on sound fragments to predict and explain the mutual intelligibility between the 70 language combinations in the MICReLa project. Figure 4.5 shows the relationship between the results of the spoken cloze test (see Appendix B) and the attitude scores (Appendix C). The correlation was significant and relatively high with an overall correlation of .65 (.60 for Germanic language combinations, .70 for Romance, and .81 for Slavic, $p < .01$).

Individual speaker characteristics, such as voice quality, mean pitch level, and intonation, may affect how recordings of languages are evaluated (e.g., Zuckerman and Driver 1989). To neutralize the influence of voice characteristics on aesthetic judgments, the matched-guise technique can be used. This technique involves pre-

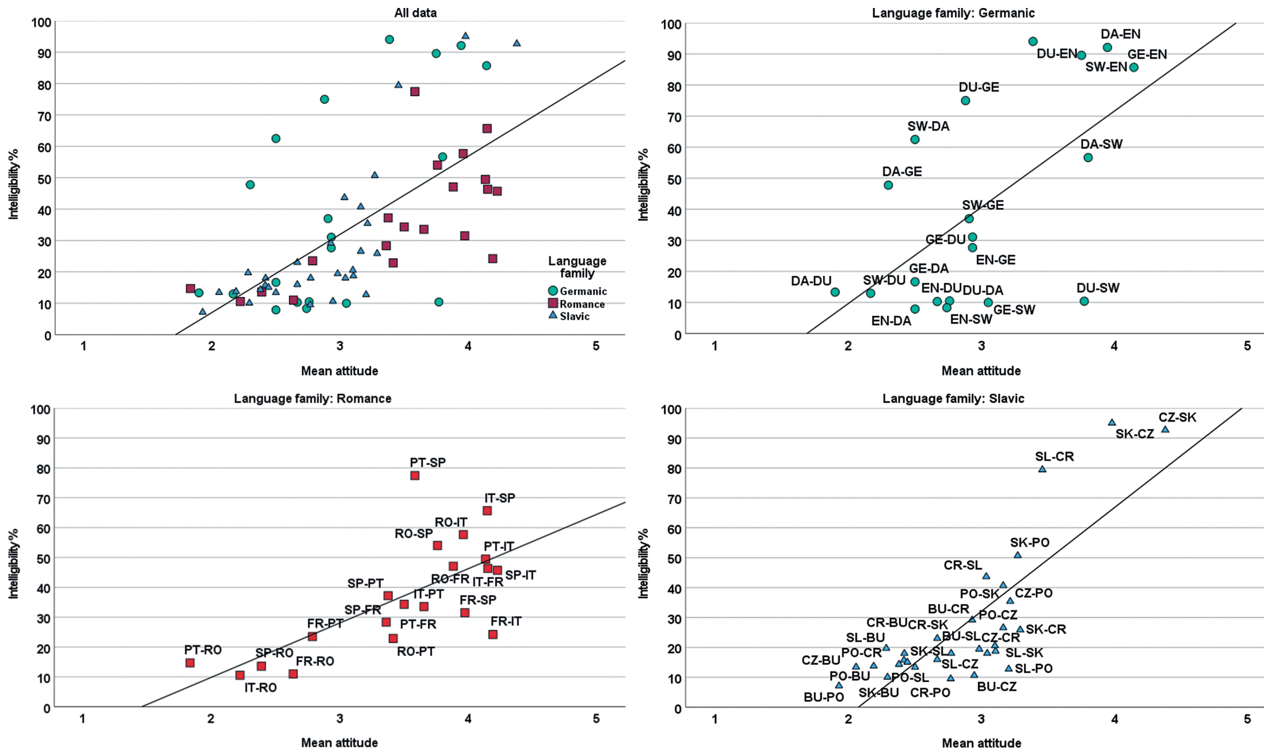

Figure 4.5: Correlation between intelligibility scores and mean attitude scores for all 70 European language pairs in the MICReLa project (r = .65), and separately for Germanic (r = .60), Romance (r = .70), and Slavic (r = .81). For explanations, see Figure 4.2.

senting lexically identical speech samples from a balanced bilingual speaker (i.e., a bilingual with equally high proficiency levels in both languages, see Section 2.1.1.5) in both of their languages, interspersed with other recordings to avoid listeners being aware of hearing the same speaker twice. Listeners then evaluate the speakers they hear for different personality traits, such as kindness and richness. Since the language is the only feature that differs between the two recordings, it can be assumed that speaker characteristics cannot explain differences in evaluations. This matched-guise technique was first used to investigate language attitudes in the French-English bilingual setting in Quebec, Canada (Lambert et al. 1960). The findings revealed a significant impact of the language spoken on listeners' assessments of the personality traits of bilingual speakers. Both English and French-speaking listeners consistently rated the speakers more favorably in terms of status and solidarity traits when they communicated in English than when they spoke French. This was believed to reflect the higher status of the English language in Quebec. For more information about the matched-guise technique, see Loureiro-Rodríguez and Fidan Acar (2022).

The statistical relationships between intelligibility scores and attitude scores gained from a matched-guise procedure have not often been tested. An exception is Schüppert, Hilton, and Gooskens (2015), who showed that the asymmetric intelligibility scores that have consistently been found for the Swedish-Danish mutual intelligibility in the literature (see Chapter 6) are also found for attitude scores from a matched-guise experiment. They found a weak but significant positive correlation ($r = .19$) between attitude scores and intelligibility as measured with a word recognition test. Listeners with a positive attitude towards the neighboring language performed better in the word recognition experiment than those with a negative attitude and vice versa.

The affective priming paradigm, which has frequently and successfully been used as a psychological measurement instrument, can also be used for measuring automatically activated language attitudes (de Houwer 2009; Walker Drager, and Hay 2022).[11] The core idea underlying the affective priming paradigm is that one can estimate the attitude towards a priming stimulus by examining how its presence influences the affective categorization of the target stimulus. For instance, listeners react quicker to a picture that evokes negative feelings if they have first

11 A related and more well-known method that has been adapted from social psychology to measure language attitudes is the implicit association task (IAT); see Campbell-Kibler (2012); Rosseel, Speelman, and Geeraerts (2019); Rosseel (2022). This is a digital sorting task where respondents have to categorize both target concepts (e.g., languages), and attributes (e.g., positive or negative images) by means of shared keys at the keyboard. This method has not yet been used to relate attitude to intelligibility and will therefore not be discussed here.

heard a priming stimulus (for example, a dialect word) that they find less attractive than if they have first listened to a stimulus that they find attractive. Similarly, they react faster to a picture that evokes positive feelings if it is preceded by a stimulus that they have positive attitudes towards than if it is preceded by a stimulus that they have negative attitudes towards. Since the priming effects occur automatically and spontaneously, this method has the advantage over other methods that it measures subconscious attitudes.

Impe (2010) measured attitudes among Dutch-speaking listeners from two regions in Belgium using an auditory affective priming experiment. Listeners were first presented with affectively positive or negative primes followed by positive or negative target stimuli. Auditory word stimuli, recorded in both Belgian standard and regiolectal varieties of Dutch, were used as primes. The target stimuli were positive pictures, such as a smiling child, and negative pictures, such as an exploding nuclear bomb. The listeners had to evaluate the picture as fast and accurately as possible as to whether it was positive or negative by pressing one of two computer keyboard buttons. The interaction between priming variety (Standard Belgian Dutch versus West-Flemish regiolect versus Antwerp regiolect) and target valence (positive versus negative pictures) was significant ($F(2, 32) = 5.27$, $MSE = 479.38$, $p < 0.01$). This means that target evaluation latencies were affected by the valence of the primes, which shows that automatic attitude activation did indeed occur. Importantly, there was a weak but significant correlation ($r = .36$) between the attitude scores resulting from the experiment and intelligibility scores as measured using a lexical decision task (see Example 2.4 in Section 2.2.1.3). This correlation was higher than when the intelligibility scores were correlated with explicit attitude scores measured by attitude ratings ($r = -.0\ 7$, $p < .001$). This shows that the subconscious attitudes of the Belgian listeners measured by means of the affective priming paradigm are better predictors of intelligibility of Netherlandic and Belgian Dutch varieties than their more conscious attitudes.

4.4 Orthographic knowledge

It is evident that the orthography of a closely related language can play a role in how well someone can read it. However, even in the spoken modality, orthographic knowledge of a listener may play a role in the intelligibility of a closely related target language.

Orthographies are often conservative and usually reflect the pronunciation earlier in history where two related languages had diverged less than in their current form. If certain sounds are pronounced the way they are spelled in the related target language while the pronunciation has changed in the native language of the listeners, listeners may be able to use their native spelling to recognize the spoken word in the related language because the written forms in the two languages have not diverged to the same degree as the spoken forms. This is shown with the examples in Table 4.2 which illustrates a situation where Danish listeners can use their own orthography to understand a Swedish spoken target word. In Danish, many sounds that have been lost in pronunciation are still preserved in the orthography, such as the spelling *lærere* for [lɛːɔɔ] 'teachers' where the two r-sounds are lost in the pronunciation, leaving only a long stretch of vowel sounds. In comparison, this word is written as *lärare* and pronounced as [lærarə] in Swedish. A Danish literate listener will have little difficulty understanding the Swedish spoken form even though it differs from the Danish pronunciation because it is very similar to the written Danish form. On the other hand, a Swedish listener will have less support from the Swedish orthography when trying to understand the spoken Danish word. The other two examples, the words for 'mild' and 'twelve', also show that the Swedish spelling is closer to the Swedish pronunciation than the Danish spelling is to the Danish pronunciation. In Section 6.1.3, the role of orthographic knowledge in explaining the Danish-Swedish asymmetric intelligibility will be discussed in more detail.

Table 4.2: Examples where Swedish has a stronger letter-to-sound correspondence than Danish. Note that Scandinavian languages have no final devoicing of obstruents. Source: Schüppert et al. (2022: 85).

	Danish	Swedish	English
Spelling	*lærere*	*lärare*	'teachers'
Pronunciation	[lɛːɔɔ]	[lærarə]	
Spelling	*mild*	*mild*	'mild'
Pronunciation	[milˀ]	[mɪld]	
Spelling	*tolv*	*tolv*	'twelve'
Pronunciation	[tɔlˀ]	[tɔlv]	

4.4.1 How to measure the contribution of orthographic knowledge

To test experimentally whether speakers of Danish do indeed use their orthographic knowledge of Danish to decode spoken Swedish, Schüppert et al. (2022) presented native Danish speakers with spoken Swedish words in a translation task. Native speakers of Danish listened to Swedish cognate words that differed in one phonetic segment and were asked to translate them. An example is provided in Table 4.3. Half of the words were pronounced in a way that is *consistent* with the spelling of the Danish word, e.g., Danish-Swedish *hat* ‘hat’ pronounced as [hæt] in Danish and [hatː] in Swedish. When listening to Swedish [hatː], Danish listeners can use their native orthography to understand the Swedish spoken word because the Swedish [a] can be recognized from the Danish orthography. The other half of the words were pronounced in a way that is *inconsistent* with the spelling of the Danish word, e.g., Danish-Swedish *gift* ‘married’ pronounced as [gift] in Danish and [jift] in Swedish. Here, the Danish listeners have no help from their native orthography and may even be confused by the [j] at the beginning of the Swedish spoken word because they do not recognize it from their native pronunciation or orthography. Schüppert et al. also obtained event-related potentials (ERPs, see Section 2.2.5) to consistent and inconsistent cognates to study the brain responses during decoding operations. Their results showed that ERPs to inconsistent cognates were significantly more negative-going than ERPs to consistent cognates between 750 and 900 ms after stimulus onset. This shows that native orthography is involved in non-native word recognition, at least when the two languages are closely related. Together with higher word recognition scores for consistent items, the data provides evidence that activation of native orthography enhances spoken Swedish word recognition in literate Danish speakers.

Table 4.3: The written and spoken forms in Swedish and Danish of the words for ‘hat’ (consistent Swedish pronunciation for a Danish listener) and ‘married’ (inconsistent Swedish pronunciation for a Danish listener). Source: Schüppert et al. (2022: 85).

‘hat’ – consistent			**‘married’ – inconsistent**		
	Danish	**Swedish**		**Danish**	**Swedish**
Spelling:	<hat>	<hat>	Spelling:	<gift>	<gift>
Pronunciation:	[hæt]	[hatː]	Pronunciation:	[gift]	[jift]

The evidence presented here shows that even when spoken intelligibility is tested, the orthography of the native language may influence the intelligibility results. Therefore, the role of orthographic knowledge should be considered when interpreting the results of a spoken intelligibility task.

4.5 Individual differences

Individual personality traits identified within psychology can have an impact on language learning (Dörnyei 2005). Similarly, personality traits can also be expected to play a role in understanding a closely related language. Examples of such traits are the ability to adapt to new situations, sociocultural and cognitive resources, and knowledge of the world. A taxonomy of personality traits has been identified within psychology, such as in the so-called OCEAN model (Goldberg 1993). It identifies five broad dimensions to describe the human personality, temperament, and psyche: openness, conscientiousness, extroversion, agreeableness, and neuroticism.

Various studies have used these five dimensions to predict success in foreign language learning but have mainly focused on language production. Limited experimental studies have been conducted to examine the impact of individual listener factors on intelligibility. Lambelet and Mauron (2017) quantified the five major personality factors through a questionnaire consisting of 60 five-point Likert scales, which was filled out by 181 French-speaking Swiss secondary school children aged 13 to 15 years. The children were also given four reading comprehension exercises to test their comprehension of Italian and were asked to answer questions about their appreciation of the task. Although there were significant correlations between task appreciation and intelligibility, no significant correlation was found between any personality traits and intelligibility. However, a relationship was observed between task appreciation and personality traits, such as "openness" and "extroversion". Consequently, the study suggested that personality traits should not be overlooked as a significant factor for reading comprehension.

Vanhove and Berthele (2015) focus on two individual characteristics that may influence the intelligibility of a closely related language, fluid intelligence (reasoning and problem-solving skills) and crystallized intelligence (pre-existing knowledge, including native vocabulary knowledge and foreign language skills), see Cattell (1963). Vanhove and Berthele (2015) examined the connection between age and intelligibility, which they measured using cognate guessing skills, i.e., the ability to recognize words that are historically related to their counterparts in one's native language. They found that, in the written modality, cognate guessing skills improve throughout adulthood. However, in the spoken modality, these skills remain relatively constant between ages 20–50 and then begin to decline, possibly due to a decrease in fluid intelligence. It may be cognitively more challenging to compare spoken phonemes across languages than letters and graphemes. Alternatively, the authors suggest that time pressure associated with auditory stimulus presentation may be a factor. Spoken items were presented only once, requiring quick application of cognitive flexibility, whereas written words remained on the screen until the participants had entered their translations.

4.5.1 How to measure individual differences

Various tools have been developed within psychology and second language acquisition to quantify individual differences and personality. An example is the NEO Personality Inventory questionnaire developed by Costa and McCrae (1992). Listeners are asked to assess their own personality by ratings on a large number of five-point Likert scales (from “totally disagree” to “totally agree”) that can be categorized according to the five principal personality dimensions (see Section 4.5). Other widely used personality self-report assessment tools are the Myers-Briggs Type Indicator (Myers et al. 2018) and the Eysenck Personality Questionnaire (Claridge 1977).

To measure fluid intelligence among their participants, Vanhove and Berthele (2015) used the advanced progressive matrices test developed by Raven (1965). This test contains 36 abstract puzzles, each with eight patterns in a 3-by-3 grid. The participants’ task is to select the logically fitting ninth pattern from a list of eight possible patterns presented below the grid. One point is awarded for each correctly selected pattern.

Vanhove and Berthele (2015) assumed that the native vocabulary size reflects the effects of learning and experience (crystallized intelligence). Therefore, they carried out a vocabulary test in the native language of the participants, an advanced German vocabulary test developed by Schmidt and Metzler (1992). The test consisted of 42 series of words and nonwords. The participants’ task was to select the existing German word presented alongside five nonwords, i.e., words without a meaning composed of a combination of phonemes that conform to the language’s phonotactic and orthographic rules. The target words ranged from the educated but common (e.g., *Ironie* ‘irony’) to the fairly unknown (e.g., *Heddur*, a type of aluminum alloy). The percentage of correctly selected target words was a measure of crystallized intelligence.

4.6 Speaker and listener strategies in interaction

Listeners can use various strategies to understand a related but unknown language. When it comes to interaction between speakers of closely related languages, both the listener and the speaker can employ various strategies to cope with and prevent misunderstandings and reach mutual understanding, depending on their proficiency levels. Discourse analysis experts have described interaction strategies for communication between native and non-native speakers, and various taxonomies have been proposed within second language acquisition studies.

Van Mulken and Hendriks (2017) based their taxonomy of communication strategies on some of these studies. They made a distinction between various strategies: appealing for assistance, signaling linguistic deficiency, signaling insecurity, offering assistance, compensatory strategies (describing, code-switching), paralinguistic strategies (such as using gestures, see Section 5.4.2), and meta-discourse (discussing task fulfillment). They found that different written communication modes (receptive multilingualism, English as a lingua franca, interaction between natives and non-natives of the same language) are characterized by a preference for particular strategies. In the case of the receptive multilingualism interactions that they set up for their investigation (see Example 2.19 in Section 2.2.4.1), the authors noted that speakers can devote more attention to the task at hand than in a non-native mode as they do not have to deal with lexical deficiencies while using their native language. Participants often resorted to paralinguistic strategies, and meta-discourse was also a common strategy used in receptive multilingualism interactions.

Blees and ten Thije (2017) note that in a receptive multilingualism interaction, it is more difficult for the interactants to adapt to each other's perceptual proficiency than in a lingua franca or native-nonnative dialogue. The participants may not detect intelligibility difficulties of the conversation partner because they both speak their own languages, making it difficult to adapt utterances to the listener's proficiency level. Therefore, the interactants should actively signal their problems with reception. Berthele (2011) discusses interlingual inferencing, i.e., how listeners use correspondences in their known languages to guess the meaning of cognates in closely related languages they do not know (see Section 4.2). However, he notes that in some situations a useful strategy is to know when to stop searching for correspondences between the native language and the target language so as not to waste time.

Braunmüller (2006) and Zeevaert (2004), summarized in Beerkens (2010), make a distinction between hearer strategies and speaker strategies. If speakers are monolingual, they can only adapt their language based on their knowledge of their own language and on communication with others who share their native language. They may, for example, speak slowly and reformulate sentences. They may also avoid using words they know to be difficult in their own language. Such words may be cognates in the listeners' language and could therefore have actually helped to improve mutual intelligibility. Speakers with knowledge of the language of the listeners can use additional strategies to reach mutual understanding, such as using particular words from the listeners' language that they know to be cognates in the two languages and avoiding non-cognates and false friends (see Section 5.1.1). The hearers, on the other hand, can make clear when they do not understand the speaker and provide feedback to show they have understood (e.g., through back-channeling, Heinz 2003). Grünbaum and Reuter (2013) formulate advice for interaction by means of receptive multilingualism among inhabitant of the Nordic countries, such as: “don’t speak too fast”, “articulate clearly”, “avoid certain words that may be difficult for the listener”, “repeat”, “explain”, and, “ask if something is not understood”.

4.6.1 How to investigate interaction strategies

Investigations concerning interaction strategies are most often of a qualitative nature, such as observations of interactions. However, experimental and more quantitative approaches have also been applied. Bahtina, ten Thije, and Wijnen (2013) used a map task (see Section 2.2.4.2) to measure interaction success and related the results to a quantitative analysis of strategies and factors that contribute to successful interaction.

Zeevaert (2007) recorded a discussion between three Swedish, one Danish, and two Norwegian university administrators and measured their speaking time in the discourse in minutes, number of turns, and the number of back-channel signals (signals of understanding and attention). He found that turn-taking took place mainly between the different languages (81%), a sign that interaction between speakers of the three languages was unproblematic and that the transitions between the turns were fluent with no conspicuous pauses and with frequent overlap between utterances of speakers. In addition, the high number of back-channel signals showed good understanding and attention. Based on his measurements and observations, Zeevaert concluded that the discourse between the six speakers was balanced and that the mutual understanding was good. A comparison with Danish speakers in other discourses shows that the Danish speaker made an extra effort to speak slowly with short utterances and some lexical adaptation, maybe because he was aware that Danish is a difficult language to understand for other Scandinavians (see Section 5.4.1 and Chapter 6).

4.7 Excluding extra-linguistic factors

Predicting and explaining the level of intelligibility between (closely related) languages is complicated. It involves a large number of factors that may influence intelligibility to varying degrees at the language level as well as at the individual level. By correlating quantifications of such factors with measurements of the level of intelligibility, the researcher may gain knowledge about the importance of the various factors in explaining intelligibility.

Sometimes, however, our main interest is to investigate inherent intelligibility (see Section 1.1) without any influence from extra-linguistic factors. In this case, we should exclude or minimize the influence of extra-linguistic factors on the results. It is difficult to exclude such factors completely, so it is essential to be aware of their influence throughout the analysis and interpretation of intelligibility results.

To exclude the influence of exposure, it is important to test a group of listeners who have had no previous exposure to the target language. This can be done by selecting listeners who have never been in contact with the target language, for instance, because they live far away from the country where the language is spoken or because the target language is not taught in the school system. Listeners can be excluded if their answers to questions in a background questionnaire or the number of correctly translated non-cognates show that they have had (substantial) exposure to the target language (see section 4.1.1).

If we want to measure the intelligibility of a language without influence from attitudes, we may exclude listeners with strong positive or negative attitudes towards the language and its speakers, as shown by their answers to attitude questions (judgements in the middle of an attitude scale, see Section 4.3.1).

The influence of orthographic knowledge can be excluded by testing illiterates. However, in many countries, it is not feasible to find a sufficient number of adult illiterate listeners who fulfill all the other criteria that the researcher set for listeners, such as social background and geographic origin. As an alternative, Schüppert and Gooskens (2012) tested young Danish and Swedish children. These children could not read or write, had no previous exposure to the neighboring language, and had positive nor negative attitudes towards the neighboring language. In this manner, they could exclude the influence of exposure, attitude, and orthographic knowledge. They found that the asymmetric intelligibility consistently found among adult Swedes and Danes disappeared and concluded that the asymmetry among adult Danes and Swedes must be caused by asymmetric influences of exposure, attitude, and orthographic knowledge (see Chapter 6).

Similarly, Gooskens, van Bezooijen, and van Heuven (2015) conducted a word translation task in which highly frequent Dutch and German cognate nouns, recorded by a perfectly bilingual Dutch and German speaker (matched guise design, see Section 2.1.1.5), were presented to Dutch and German children between 9 and 12 years old. The children were comparable in that they had no knowledge of the target language or a related dialect and expressed equally positive attitudes towards the other language, its speakers, and the country. Previous research has indicated that Dutch listeners understand German better than vice versa, possibly due to exposure and German being a school language in the Netherlands. However, the results of this study showed that even these young Dutch children were significantly better at understanding the German cognates (50% correct translations) than the German children were at understanding the Dutch cognates (42%). Since the relevant extra-linguistic factors had been excluded, the asymmetry must have a linguistic basis (see Section 6.2).

Since the plurilingual resources of listeners may influence intelligibility results in different ways depending on the number of languages and what languages they know as well as the level of knowledge, the researcher may wish to control these factors by testing monolinguals only. However, it may be difficult to find genuinely monolingual listeners in the present-day situation where English is dominant as a school language and a lingua franca in many countries. Second best may be to include only listeners who have minimal knowledge of foreign languages and do not know a language that is closely related to the target language.

By testing large numbers of listeners, the influence of individual differences on the intelligibility results is likely to be averaged out. Alternatively, speakers could be chosen with care using psychological tools to ensure that various idiosyncratic extra-linguistic factors are balanced across groups of listeners, so that they do not influence the results in unintended ways. In addition, outliers could be removed from the data set.

Another approach is based on the fact that mutual intelligibility can be asymmetric. It has often been assumed that extra-linguistic factors explain most of this asymmetry (see Section 6.1). Simons (1979: 101) assumes that extra-linguistic factors explain nearly half of the variance that remains after correlating intelligibility scores with linguistic distances. Simons (1979) and Grimes (1992), therefore, suggest that if intelligibility is assessed bidirectionally, i.e., measuring the intelligibility of language A among speakers of language B, as well as the intelligibility of language B among speakers of language A, the smaller measure is likely to be less influenced by extra-linguistic factors. This approach, they argue, allows for a more precise reflection of inherent intelligibility in both directions. Nevertheless, challenges arise when excluding extra-linguistic factors from intelligibility scores using this approach. First, strong and undeniable evidence exists that linguistic factors may also cause asymmetric intelligibility (see Section 6.2). Second, extra-linguistic factors may boost the intelligibility of both languages. Finally, some extra-linguistic factors, such as attitude, may have a positive as well as a negative influence on intelligibility. Still, by observing asymmetry, researchers can become conscious of factors that may have influenced their intelligibility results and feel challenged to look further into such instances of asymmetry (see Chapter 6).

4.8 Conclusions

This chapter has shown a relationship between various extra-linguistic factors and the degree of intelligibility of closely related languages as measured with functional tests. However, as discussed in the section about attitudes (Section 4.3), it is not possible to draw conclusions about cause and effect based on the correlations between attitude measurements and intelligibility measurements. This also applies to other extra-linguistic measurements and to any correlational study. It seems logical to conclude that the significant correlations between exposure and intelligibility measurements presented in Section 4.1.1 shows that exposure to a language will improve intelligibility. However, it may also be the case that people are more open to contact with speakers of languages that they (think they will) understand.

In addition, there is often interaction between various extra-linguistic and linguistic factors. In particular, it has been shown that a high degree of familiarity and exposure is an important factor in creating positive attitudes towards a language, resulting in significant correlations between exposure scores and attitude scores. This means that if both exposure and attitude scores correlate with intelligibility, a correlation between attitude and intelligibility could result from intercorrelation between attitude and exposure. Conclusions about causality should, therefore, be drawn with care, and are best avoided.

Chapter 5
Linguistic and paralinguistic determinants of intelligibility

Chapter 4 discussed extra-linguistic determinants that may at least partly determine how well speakers of closely related languages understand each other (the blue square in Figure 1.1). The present chapter focuses on linguistic and paralinguistic determinants of intelligibility (the green square in Figure 1.1).

To investigate the statistical relationship between intelligibility results and linguistic determinants, it is necessary to look for ways to quantify communicatively relevant linguistic differences. To measure the differences between dialects, computational and quantitative methods have recently been developed and refined within dialectometry. With such measures, the degree of linguistic similarity (or difference, which is the complement of similarity) at various linguistic levels (e.g., lexical, phonetic, morphological, and syntactic) can be quantified and used to characterize dialect areas by means of dialect maps and language trees. Linguistic difference measures show high correlations with distances as perceived by the dialect speakers themselves (Gooskens and Heeringa 2004a). Importantly, various investigations have demonstrated that such linguistic distance measures also correlate with intelligibility measures. Note that, in real life, mutual intelligibility is also influenced by extra-linguistic factors (see Chapter 4). Such factors may overrule linguistic factors. Therefore, if we want to draw conclusions about the role of linguistic factors in explaining intelligibility, it makes sense to focus on inherent intelligibility (see Section 1.1) and exclude the influence of extra-linguistic factors as much as possible. In Section 4.7, ways to do so were discussed.

This chapter presents and discusses a number of distance measures at different linguistic levels and shows examples of research on the relationship between intelligibility and measurements of linguistic distances. It also includes a section about para-linguistic factors that play a role in explaining and predicting intelligibility. The focus will be on linguistic measurements that have been used to explain functional intelligibility measurements.

Many insights from psycholinguistic research on speech perception in noisy circumstances, second language acquisition, and perception of foreign-accented language are relevant to this chapter. It is generally assumed that the spoken-word recognition system is, in principle, language-independent and that the task of listeners in recognizing speech in a second language is, therefore, very similar to recognizing speech in their native language (van Heuven 2008; Lecumberri, Cooke, and Cutler 2010). However, more uncertainty must be resolved at all lin-

https://doi.org/10.1515/9783111134697-005

guistic levels when listening to a related language rather than the native language. Listeners may encounter words that they do not understand because there are no cognates in their native language. They may be unable to distinguish between phonemes in the related language, resulting in incorrect word recognition, and non-native listeners may also be hindered by differences between their native language and the related language at the morpho-syntactic and pragmatic levels. All these uncertainties add up and make it complicated for listeners to process a related target language.

5.1 Lexicon

Lightbown and Spada (2006: 96) point out that even when words are not placed in the proper order, are pronounced imperfectly, or are not marked with the correct grammatical morphemes, it is still possible to communicate. However, misunderstanding one or more words often causes communication to break down. If two languages share no vocabulary, the languages are, in principle, not mutually intelligible; conversely, the larger the lexical overlap, the larger the mutual intelligibility will be. Grimes (1992) even sets a threshold for the percentage of vocabulary similarity needed for understanding a simple narrative (60% overlap) and for more complex communication (85% overlap).

A simple way to measure lexical distances between two languages is to calculate the percentage of non-cognates, i.e., the percentage of the vocabulary with no lexical overlap (see Section 1.1). In Table 5.1, an example of the calculation of the distance from Dutch to four other Germanic languages is provided. The six Dutch words in this small corpus have cognates in some languages but not in others. For example, there are cognates for Dutch *ding* ‘thing’ and *computer* in all four related languages, but Dutch *schaap* ‘sheep’ does not have a cognate in the two North Germanic languages, Danish and Swedish. The lexical distance from Dutch to each of the other four languages is expressed as the percentages of non-cognates. In this example, the lexical distance from Dutch to German is 0% (no non-cognates), 33% to Swedish (two non-cognates out of six words), and 50% to Danish and English (half of the words are non-cognates). Based on this small corpus, German listeners are expected to have less difficulty understanding Dutch than Swedish listeners. Among the four groups of listeners, the Danish and English listeners will have the greatest difficulties understanding Dutch.

Most scholars agree that since the degree of lexical overlap is fundamental for understanding a closely related language, a relationship between the degree of overlap and the level of intelligibility can be expected. Many investigations have found significant correlations between measurements of lexical distances

Table 5.1: Example of the compilation of a cognate list used to measure the lexical distances from Dutch (the target language) to four other Germanic languages (listener languages). The left column represents the Dutch source list. For each word, cognates in each of the listener languages are noted. An empty cell means there is no cognate. The numbers and percentages of non-cognates are shown in the bottom row. The percentages express the lexical distances between the listener and target languages.

Dutch target	**Listener language**			
	Danish	**English**	**German**	**Swedish**
ding	ting	thing	Ding	ting
computer	computer	computer	Computer	computer
schmink	sminke	–	Schminke	smink
schaap	–	sheep	Schaf	–
boom	–	–	Baum	–
vraag	–	–	Fragen	fråga
Non-cognates	3 (50%)	3 (50%)	0 (0%)	2 (33%)

(or similarity) and results of intelligibility tests. Simons (1979: 60) provided an overview of early work that has looked at the relationship between measurements of functional intelligibility scores and lexical distances.[12] He reproduced data from 10 intelligibility studies from various parts of the world (Africa, Oceania, and North America). Each study used different methodologies for measuring intelligibility, but they all had in common that they calculated lexical distance expressed as the percentage of non-cognates. In the case of strong asymmetry in intelligibility between AB and BA language combinations (more than 10%), Simons left out the language combination with the highest score, because he assumed that asymmetric social relations caused this asymmetry. The lowest score is more likely to reflect inherent intelligibility (see Section 4.7 and introduction to Chapter 6). He also left out two studies that he assumed to be outliers due to a lot of contact between speakers in the places included in the investigation (see Section 4.1). He correlated the intelligibility scores and the lexical distances from the remaining eight studies (175 language combinations), see Figure 5.1. The results show that lexical distance measures explain nearly 85% of the variation, and the

12 Simons calculated lexical similarity, but here, the similarity measures are changed into distance measures by subtracting the similarity measures from 100%. Lexical distance is the complement of lexical similarity. Lexical similarity is a measure of lexical overlap. Lexical distance is the percentage of words with no lexical overlap (the percentages of non-cognates in two languages), i.e., 100% minus the percentage of non-cognate pairs.

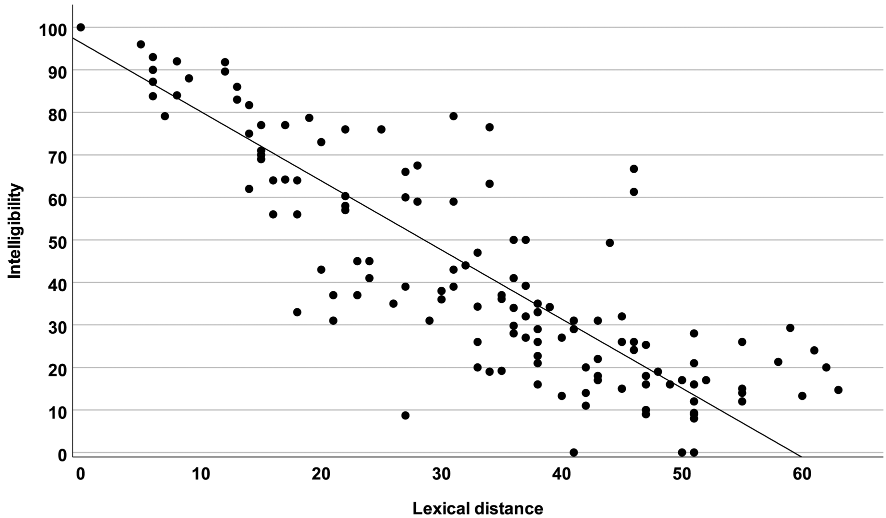

Figure 5.1: Scattergram showing the relationship between intelligibility scores and lexical distance scores in 8 studies. Reproduced from Simons (1979: 88).

resulting linear model predicts that when the lexical distance is above 60%, there will hardly be any intelligibility.

In Figure 5.2, scattergrams show the correlation between inherent intelligibility in the MICReLa investigation, including European language combinations (see Section 3.1.4 and Appendix B) and lexical distances between the same language combinations (see Appendix D). The overall correlation for all language combinations is high ($r = -.76$), and also the correlations for the individual language families show that lexical distances are a good predictor of intelligibility ($r = -.95$ for Germanic languages, −.69 for Romance languages and −.80 for Slavic languages, $p < .01$). Figure 5.2 seems to confirm the results by Simons (1979), which show that there is hardly any intelligibility if the lexical distance is higher than 60%, as discussed above.

However, correlations of lexical distances with results from an investigation on the mutual intelligibility between Danish, Norwegian, and Swedish among listeners from various parts of Scandinavia showed that lexical distances are not always good predictors of intelligibility (Gooskens 2007). The intelligibility scores in Figure 5.3, in general, confirm the results of previous findings (e.g., Maurud 1976, as summarized in Torp 1998). Norwegians are good at understanding the neighboring languages, Swedish-speaking listeners from Sweden and Finland have great difficulties understanding Danish, while Danes have less difficulty understanding Swedish. The correlation between lexical distances and intelligibility in Figure 5.3 is insignificant ($r = -.42$, $p = .11$). This can be explained by the fact that

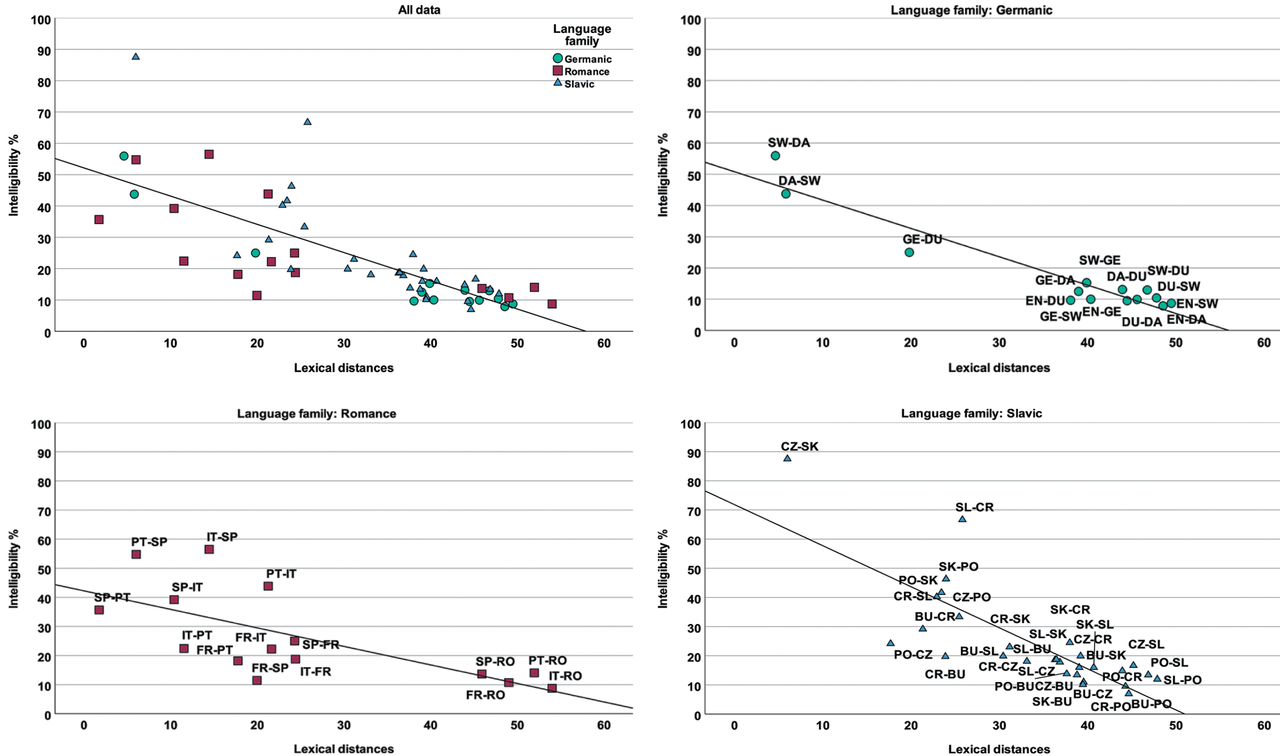

Figure 5.2: Correlations between inherent intelligibility scores and lexical distances for 57 European language pairs in the MICReLa project ($r = -.76$), and separately for Germanic ($r = -.95$), Romance ($r = -.69$) and Slavic ($r = -.80$). For explanations, see Figure 4.2.

there is very little lexical variation between the Scandinavian varieties, and none of the lexical distances are above 3.6%. Danish, Swedish, and Norwegian have almost the same vocabulary. Correlation analysis makes no sense when one of the variables has no variance. Both X and Y variables must be normally distributed for a parametric correlation analysis to be legitimate. Lexical distances can, therefore, not be used to explain the differences in mutual intelligibility between the Scandinavian languages. We discuss other explanations throughout the book (in particular in Sections 4.3, 4.4, 4.7, 5.2, 5.4.1 and Chapter 6).

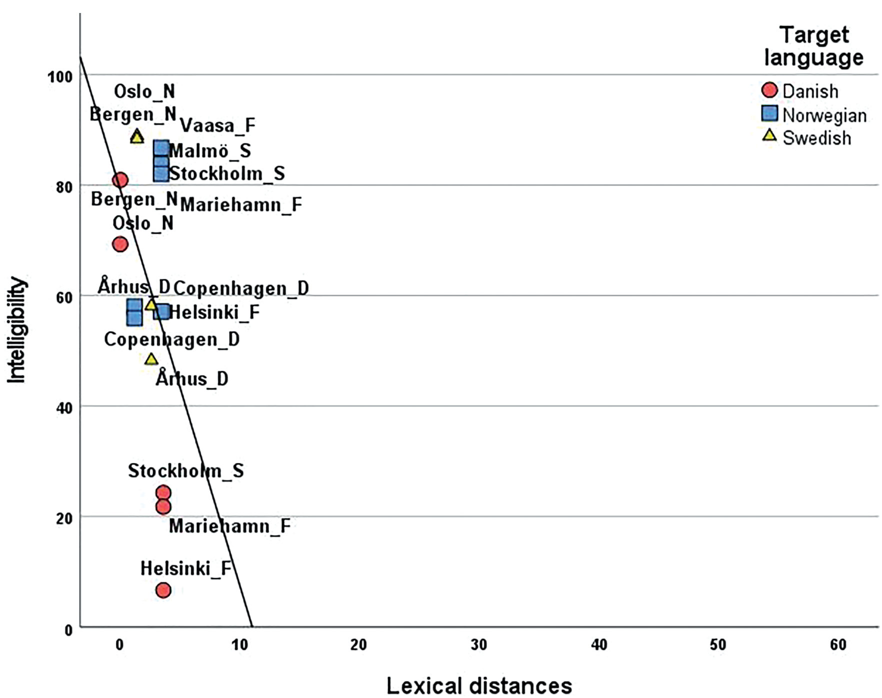

Figure 5.3: The correlation between intelligibility (percentages of correct answers in a cloze test) and lexical distances (percentage non-cognates), $r = -.42$, $p > .11$, from an investigation of Standard Danish, Standard Norwegian (represented by the Oslo variety) and Standard Swedish for listeners in nine different regions in either Denmark (D), Norway (N), Sweden (S) or Swedish-speaking Finland (F). Adapted from Gooskens (2007: 461).

Grimes (1992: 17) notes that high lexical similarity is a poor predictor of high intelligibility, while low lexical similarity is a better predictor of low intelligibility. If two languages have no vocabulary in common (i.e., a large lexical distance), listeners cannot understand the target language unless they have had some exposure to it before. On the other hand, a large overlap (a small lexical distance) between vocabularies does not guarantee a high level of intelligibility since the pronunciation of cognates in two languages may sometimes be so different that the listener may not recognize them as being cognates (see Section 5.2). In other words, lexical overlap is a necessary condition – without lexical overlap, none of the other measures make sense – but not a sufficient condition – the overlap needs to be recognizable for it to be useful to listeners. The European data in Figure 5.2 confirm the predictions by Grimes. The intelligibility is low for all language combinations if there is a large lexical distance. If the lexical distance is small the intelligibility scores tend to be spread over a large part of the scale from low to high intelligibility, just like in the Scandinavian investigation in Figure 5.3.

5.1.1 Considerations when measuring lexical distances

To calculate lexical distances between two languages, it is imperative to carefully consider what data set should be used for the calculations. To gauge the extent of chronological divergence among languages, historical linguists often use word lists, such as the Swadesh list (Swadesh 1971) and the Leipzig-Jakarta list (Tadmor 2009). The words in these lists are chosen for being resistant to borrowing and for being part of the vocabulary of most languages because they describe universal things. For example, the word for *head* has different unrelated forms in the languages of the world (e.g., French *tête,* Finnish *pää,* and Turkish *baş*), but it exists in all languages.

However, if the aim is to predict or explain intelligibility between modern languages, a better choice may be to base the measurements on a representative or random sample of the languages in question. In recent years, many corpora have been compiled for larger languages. Some researchers base their measurements on the most frequent words in such corpora, assuming that they are good representations of the languages. Nation and Waring (1997) showed that the 1000 most frequent word types in a large corpus generally cover more than 70% of the word tokens in the corpus. This means that if listeners recognize the 1000 most frequent word types in a language, they will be able to recognize around 70% of the words in a text. The top 2,000 words even make up around 80% of typical English texts. The exact numbers for spoken languages and languages without a strong literary tradition may differ and depend on the corpus. However, the general trend will be the same: the most frequent word types cover a large part of the word tokens that people will hear or read in their daily lives.

In some intelligibility investigations, the lexical distance measurements are based on the same material used for the intelligibility tests, i.e., a small but precisely targeted corpus. For example, the four texts of each approximately 200 words used for the cloze test in the MICReLa project (see Section 3.1.2) were used to calculate lexical distances. The percentages of non-cognates in the four texts for each language pair correlated with mean intelligibility scores in the cloze test using the same texts ($r = -.76$, see Figure 5.2). This shows that it is possible to predict intelligibility scores rather well on the basis of the test material itself. The same cloze test results were also correlated with the lexical distances measured on the basis of the 100 most frequent nouns in the British National Corpus (2007) translated into the 16 target languages involved (Gooskens and van Heuven 2020). The correlations with the intelligibility measurements were equally high or even better ($r = -.82$ for Romance languages, $r = -.86$ for Slavic languages, and $r = -.75$ for Germanic languages, $p < .01$). This shows that it is possible to predict intelligibility reasonably well on the basis of a short list of frequent nouns that are not part of the test material.

When coding the data, several considerations must be taken into account. The most straightforward measurement of lexical distances between word pairs is binary: word pairs are either cognates or non-cognates. It is not always easy to identify which words are cognates and which ones are non-cognates, since cognates may have changed so much over time that they are difficult to recognize as cognates. To illustrate, the cognate word pair English *fish* and Danish *fisk* can easily be recognized as having the same origin, Proto-Germanic **fiskaz*, but the word pair English *year* and Danish *år* may be difficult to recognize as cognates since the forms have diverged more from Proto-Germanic **jǣran*. It may be necessary to use etymological dictionaries to decide whether word pairs are cognates or non-cognates. However, such dictionaries are not available for all languages, and if they are, it is labor-intensive to look up a large number of words. In the case of a large data set, an alternative may be to use one of the algorithms developed to identify cognates automatically. For example, Ciobanu and Dinu (2020) proposed an algorithm for extracting cognates from electronic dictionaries with etymological information. They used the resulting data set of related words to develop machine-learning methods for identifying cognates. They did this by aligning orthographic sequences in pairs of words to extract rules for changes occurring when words enter new languages. They even refined the method to discriminate between historical cognates and borrowings.

Non-cognates can sometimes be so similar in form and meaning that they may be mistaken for cognates (false cognates). A well-known example of a false cognate is German *haben*, Latin *habere,* both meaning ‘to have’ but descending from two different Proto-Indo-European roots (*kap- and *ghabh-). Another example is Spanish *mucho,* English *much,* which have similar meanings but different roots (*ml̥tos and *meg-). Related to false cognates are false friends, i.e., words (cognates or non-cognates) that look or sound similar but differ in meaning. The degree of semantic overlap may vary. An example of false friends with the same etymological origin and some semantic overlap is English *embarrassed* 'ashamed', derived indirectly via French from Spanish *embarazado* 'pregnant'. Sometimes listeners may be so wary of the existence of false friends that they become suspicious of cognates and assume that they are false friends, a phenomenon that Kellerman (1978) referred to as "homoiophobia", see also the explanation in de Bot, Lowie, and Verspoor (2005).

While regular non-cognates will, in principle, hinder intelligibility, false friends may cause even larger problems because they may actually mislead the listener. The more similar the false friends are, the larger the chance that the listener will be misled. In addition, since listeners may not realize that these words have different meanings, they are less likely to use contextual cues to guess their meaning than in the case of clearly different word shapes. It could be argued that from an intelligibility viewpoint, false cognates may be less disturbing than false friends. Both false cognates and false friends are words that have similar forms in two languages. False friends may or may not have different etymological origins; false cognates do not share a common ancestry. However, false friends usually differ considerably in meaning, and the listener is, therefore, likely to misinterpret them. False cognates, on the other hand, usually share meaning so that listeners will not be completely misled when trying to match a false cognate with a word in their own language.

Huisman, van Hout, and Majid (2021) showed how semantic overlap and variation can be quantified. They collected naming data for 40 video clips depicting cutting and breaking events in Japonic and Germanic language varieties. The verb used for each event in each language was coded. For example, the Japanese verb used for both concepts 'saw' and 'cut' is *kiru,* and therefore, gets the same word code, while in English (and other Germanic languages), two different word codes are assigned. Based on the codings, semantic similarity between languages could be calculated. The results showed that, in general, related languages resembled each other more than unrelated languages and that Japonic languages resembled each other more than Germanic languages do.

In addition to semantic overlap there may also be instances of morphological overlap. For example, an English listener may hear the Spanish word *mentalidad* 'mentality' and recognize the stem but be confused by the non-cognate derivational (word-forming) affixes (*-ity* versus *-idad*). Kessler (1995) used two separate definitions of cognates. In the first definition, called etymon identity, words are defined as cognates if their stem has the same historical origin. In the second method, called word identity, words are defined as cognates only if each morpheme in the word is cognate in the pair of words. When calculating distances, a possibility could be to assign a smaller weight to a translation that contains the correct stem morpheme but a non-cognate affix. In Section 5.3.1, distances between inflectional (grammatical) morphemes are discussed.

Van Bezooijen and Gooskens (2007) introduced the percentage of cognates related via a suppletive paradigm as a measure expressing to what degree it is possible to deduce the meaning of a word paradigmatically. For instance, the Frisian translation of the Dutch third person plural present tense of the verb 'to be' *zijn* is *binne*. These two words are not inflections of the same stem. Nevertheless, a Dutch reader may understand the meaning of the Frisian word because it is related to the Dutch word *ben* (first person singular present tense of the verb 'to be'), which belongs to the same paradigm as *zijn*. Words in this category are often function words. Van Bezooijen and Gooskens found that the percentage of cognate Dutch-Frisian and Dutch-Afrikaans content words is almost identical (94.1% versus 94.6%). However, whereas in Frisian, nearly all function words (93.4%) are related directly to their Dutch counterparts, Afrikaans has relatively many function words (23.7%) related to Dutch via a suppletive paradigm.

When creating a parallel list of words to be used for lexical distance measurements, some of the words in the lists may be compounds, i.e., two or more words linked together to produce a word with a more specialized meaning, for example, *tooth* + *brush* = *toothbrush.* Compounds are transparent when their component parts (lexemes) are semantically related to the overall meaning of the compound, such as *snowball* (a ball made of snow). They are opaque when the lexemes do not directly contribute to the meaning of the compound, e.g., *waist* + *coat* = *waistcoat* (a sleeveless upper body garment). If the corresponding words in two languages are compounds consisting of either the same cognate components or two unrelated components, it is easy to award points to the lexical difference between the two words: they are either cognates or not. However, if one of the components in the compound is unrelated while the other is related, e.g., English *cheesecake* versus Danish *ostekage*, the lexical difference is often assigned half a point. Note that depending on the stage of compound evolution, English compounds may be written as two separate words, separated by a space or a hyphen. In contrast, some languages consistently write compounds as a single word, as seen in Dutch with *koffiemok* compared to English *coffee mug*. In the case of transparent compounds, there may be variation in the compositionality of non-cognate or half-cognate compounds, i.e., the extent to which the overall meaning of the whole word can be derived from the meaning of the components of the compound. For instance, German *Handtuch* 'towel' (literally 'hand cloth') or *Seehund* 'seal' (literally 'sea dog') more or less speak for themselves even if the literal translation does not exist in the native language of the listener (Weller et al. 2014).

Some word classes may be more central to understanding a text than others. Typically, since content words (such as nouns, adjectives, and main verbs) convey the semantic content of the message, they are considered more crucial for intelligibility than function words (such as articles, conjunctions, prepositions, pronouns, and auxiliary verbs), see van Bezooijen and Gooskens (2007). This is evident in telegrams and newspaper headlines, where function words are frequently omitted to convey the message succinctly, yet still allow for understanding. Additionally, even within the group of content words, certain words may be more central than others in particular contexts. Salehi and Neysani (2017) investigated the intelligibility of Turkish among listeners who were native speakers of the related Iranian-Azerbaijani language. They used a translation task and had the listeners indicate which words were problematic. They found that Turkish nouns and verbs were the most critical syntactic categories in facilitating or obstructing intelligibility among Iranian-Azerbaijani listeners, while adjectives and adverbs were less important. They explain this by the higher semantic load of the nouns and verbs compared to adjectives and adverbs, which modify the meaning of nouns and verbs. They illustrate this with the following Turkish example: *Ekip ikinci dalış yerine gitmek için harekete geçtı* ('The crew started to go to the second diving place'). In this sentence, the word *dalış* ('diving') is a noun that does not have a cognate in Iranian-Azerbaijani, and due to its high semantic load, the meaning of the whole sentence becomes unclear to the listeners. On the other hand, the non-cognate adjective *ufak* ('small') in the Turkish utterance *Ufak balıklar falan varmış orda* ('There are small fish over there') was less problematic for the overall understanding. This means that it may be possible to improve lexical distance measurements as predictors of intelligibility by weighing differences in some word classes more heavily than differences in other word classes.

In this book, loanwords are included in the definition of cognates when measuring lexical distances for intelligibility research (see Section 1.1). However, it makes sense to distinguish two kinds of cognates, namely inherited cognates and cognate loanwords. Recognizing cognate loanwords in a related language is often easier than identifying inherited cognates, which have evolved over a more extended period of time from a common ancestor language. Cognate loanwords are generally more similar, as they have had less time to diverge from their source language compared to inherited cognates, which have been part of the lexicon for a longer time. For example, the inherited cognates Swedish *lag* [lɑːg], Danish *lov,* [lɒʊ] ('law') have become so different, both in their spoken and their written forms, that they may no longer be recognizable for the speakers of the neighboring language. In contrast, a recent loan like *team*, which has not taken part in sound changes and retained its English pronunciation in Swedish and Danish, is undoubtedly easy to identify. In addition, many loanwords have specific segmental and prosodic properties that make them resistant to linguistic changes that have affected inherited words. For instance, most French words are polysyllabic and have stress on the final syllable, whereas Germanic languages are characterized by primary stress on the (monosyllabic) stem syllable. French loanwords keep the stress on the final syllable, like in the original French pronunciation in Swedish *miljö* [mɪl.ˈjøː] and Danish *miljø* [mil.ˈjø] 'environment'. Therefore, final syllables mostly maintain the full vowel in French loans, while in Germanic languages, vowels in unstressed final syllables of inherited words are often reduced (or limited to a small set like in Swedish), see Gooskens, Kürschner, and van Bezooijen (2012). In loanwords, full vowels are often retained in unstressed syllables, even if the non-accentuated syllable is final, cf. Danish *dato*, Swedish *datum* 'date' from Latin *datum*. Finally, loanwords are frequently known not only from the native language but also from foreign languages that the listeners are familiar with (see Section 4.2). For instance, the recognition by French listeners of Spanish *computadora,* French *ordiniteur* 'computer', may be facilitated if French listeners know the English equivalent *computer*.

All these characteristics of loanwords may have a facilitating effect on recognition compared to inherited words. Gooskens, Kürschner, and van Heuven (2022) investigated the role of inherited words compared to loanwords (from German, Latin, French, English, Greek, and other languages) for the intelligibility of written Danish for Swedes. They used a word translation task with inherited words and loanwords and two cloze tests, one with a large percentage of loanwords and one with few loanwords. Their results show that it is easier for Swedish listeners to recognize Danish cognate loanwords than inherited cognates and that the text with many loanwords was easier to read than the text with few loanwords. They explained these results by the fact that (recent) loans in Swedish have diverged less and are, therefore, more similar to the Danish counterparts than inherited words. In many Western countries, puristic movements have taken action against the large number of loanwords that have entered their languages. However, these results show that from a mutual intelligibility point of view, loanwords may constitute an advantage as long as languages borrow the same words (with the same meaning).

Several other word characteristics are likely to influence intelligibility. The frequency of words in the native language may affect the identification of cognates in a closely related language since frequent words are more likely to come to the listeners' minds immediately than infrequent words. In the mental lexicon, the activation of a word that is recognized remains high for a long time and never fully returns to its previous resting level. It has been observed that high-frequency words are known to more people and are processed faster than low-frequency words (see Brysbaert, Mandera, and Keuleers (2018) for a review of the word frequency effect). Therefore, highly frequent words have an advantage in the recognition process (Luce and Pisoni 1998). Van Heuven (2008) suggests that the frequency effect does not only apply to the native language but that cognates in related languages with high frequency in the native language may also be easier to decode correctly. Vanhove and Berthele (2015) confirmed the frequency effect for German-speaking Swiss participants who were asked to translate written stimuli from other Germanic languages (Danish, Dutch, Frisian, and Swedish). Their results showed that stimuli with high-frequency German and English cognates were translated correctly more often than stimuli with low-frequency cognates. A measure comprising both German and English frequencies was an even better predictor than the frequency of German and English separately, suggesting that multilinguals use their knowledge of other languages besides their native language when guessing the meaning of words in a related language (see Section 4.2).

Another lexical property that may influence intelligibility is neighborhood density. This concept pertains to the number of words in a language that share similar phonological characteristics and is typically quantified by identifying the number of words that are created by adding, deleting, or substituting a single phoneme in the same position in a given word (Luce and Pisoni 1998).[13] To illustrate, the target word *cat* has neighbors such as *scat* (one phoneme addition), *at* (one phoneme deletion), and *fat*, *cot*, and *cab* (one phoneme substitution). Some words have few or no neighbors, such as *elephant.* A large number of neighbors (high neighborhood density) enlarges the number of recognition candidates, causing delay or even failure of successful word recognition. Kürschner, Gooskens, and van Bezooijen (2008, see Example 2.1 in Section 2.2.1.1) found the neighborhood density of Danish words to be a significant predictor of the intelligibility of Swedish target words for Danish listeners. A measure of lexical distance might be refined by taking neighborhood density into consideration, high-density words being assigned a larger weight than low-density words. It should be noted that short words generally have a denser neighborhood than long words. There is also a relation between word length and frequency of word occurrence; short words tending to be more frequent than long words (Zipf 1972). From this it follows that the possible advantage of short words being more frequent than long words can be predicted to be neutralized (to some extent) by the neighborhood density effect: short words are generally frequent (advantage) but have a high neighborhood density (disadvantage). Long words are infrequent (disadvantage) but have a low neighborhood density (advantage).

As mentioned above, several investigations have found high correlations between intelligibility scores and percentages of cognates. However, as has become

13 Several online tools for automatically calculating neighborhood density are available, e.g., https://corpustools.readthedocs.io/en/latest/neighborhood_density.html, https://calculator.ku.edu/density.

clear, not all cognates are equally easy to recognize. A number of word characteristics have been discussed that may influence how well a cognate is recognized. Such characteristics should be taken into account when selecting data for measurements and interpreting intelligibility results. Some characteristics could even be built into refined measurements of lexical distances to improve the predictability of intelligibility. However, it is not a priori clear how each of the characteristics should be weighed. This can only be decided by experimental testing.

5.2 Phonetics

Section 5.1 showed that lexical distances tend to correlate highly with intelligibility scores. The more non-cognates there are in the target language, the more difficult it is to understand the target language, and the more cognates the easier it is. However, lexical distances can only predict intelligibility to a certain extent. For cognate relationships to be of actual help, it is, of course, essential that the listener is aware of words being cognates. The fact that a word is a cognate does not always mean that a listener will recognize it. Cognates in closely related languages may sometimes have changed so much over time that listeners will no longer be able to match them with the counterparts in their own language. They must, therefore, bridge minor or major differences to map the input onto the corresponding word in their native mental lexicon. It has been shown that the cognate facilitation effect depends on the degree of similarity between the cognates (Otwinowska-Kasztelanic 2011) and the degree of transparency of the lexical meaning.

Van Bezooijen and Gooskens (2007) developed a measure of lexical transparency to reflect that the deductibility of cognates varies. In their scoring, four grades of transparency were distinguished (see Figure 5.4). It could range between 0 and 3 points, with lower values denoting higher transparency and higher values denoting higher opacity. The total transparency was calculated by averaging the total number of points over the total number of word pairs. The authors tested the written intelligibility of Afrikaans and Frisian among native speakers of Dutch. Their study revealed that Dutch-speaking readers with no previous exposure to Frisian and Afrikaans could read Frisian and Afrikaans newspaper articles to some extent. However, they were considerably better at understanding Afrikaans than Frisian. In a cloze test (see Section 2.2.3.6), four of the five Afrikaans words were placed back correctly in their original sentence context, while no more than half of the Frisian words were placed back correctly. The percentage of cognates was very similar in the two languages (96.3% versus 97.9%). However, an analysis of transparency using the system shown in Figure 5.4 made clear that the transparency of the meaning of the cognates was higher for Afri-

- Meaning **completely transparent** (0 points). When two cognates have an identical form, there is no recognition problem. In this case, a score of 0 points is assigned. Example: Afrikaans *uitbuiting* vs. Dutch *uitbuiting* ‘exploitation’.
- Meaning **fairly transparent** (1 points). A score of 1 point is assigned whenever two cognates are so similar that the reader can be assumed to recognize the relationship fairly easily. In most cases, there is a difference in only one or two letters. Example: Afrikaans *sewentig* vs. Dutch *zeventig* ‘seventy’.
- Meaning **fairly opaque** (2 point). A score of 2 points is assigned whenever two cognates have so little in common that it will be quite difficult for a Dutch reader to recognize the relationship. Usually several (more than two) letters will be different. Example: Frisian *jierren* vs. Dutch *jaren* ‘years’.
- Meaning **completely opaque** (3 points). A score of 3 points is assigned if the two cognates bear so little resemblance that it must be (virtually) impossible for a Dutch reader to see the relationship. In most cases, the majority of letters will differ. Example: Afrikaans *hê* vs. Dutch *hebben* ‘have’.

Figure 5.4: Example of scores for different degrees of transparency between cognates, from high transparency (0 points) to complete opacity of meaning (3 points). Adapted from van Bezooijen and Gooskens (2007: 256).

kaans (mean score 0.50 on the scale from 0 “transparent” to 3 “opaque”) than for Frisian (mean score 1.33), reflecting the lower intelligibility scores for Frisian.

A disadvantage of the transparency measures is that they are, to some extent, subjective because the researcher has to decide how to score the degree of transparency for each word pair. This is also very time-consuming. In the 1960s, researchers already expressed the wish to develop algorithms for phonetic distance measurements that could be used to predict and explain intelligibility between dialects and closely related languages automatically and with a more refined and objective measure than lexical measurements. Ladefoged (1968, 1970) illustrated how such measurements of phonetic distances between languages could be developed by binary comparisons of segmental features, i.e., the smaller building blocks of speech sounds. Such as continuant, obstruent, nasal, strident, etc., for consonants, and high, low, front, back, round, for vowels. To illustrate, in the system that he developed, [b] and [d] have eight out of ten features in common, and [b] and [ʃ] have only four out of ten features in common. He suggested measuring the degree of segment similarity by counting the number of features each segment had in common. The features in each segment were compared with the corresponding features in the corresponding segment in all words in corpora of the languages to be compared. A dummy (unmarked) segment was added in the case of missing segments. The sums of all the comparisons involved indicated the degree of phonetic similarity of each pair of languages. Ladefoged applied this method to 20 Ugandan languages and found the results to be plausible and reflecting functional intelligibility. He also suggested improvements to the metrics, such as assigning different weights to various features.

Another early investigation was carried out by Cheng (1997) based on phonetic transcriptions of over 2,700 cognate words in seventeen Chinese dialects. He computed the complexity of the correspondence patterns required to convert the word strings from one dialect to their counterparts in the other dialect and referred to this as "systemic mutual intelligibility". Cheng's calculations involved assigning reward and penalty points to vowel and consonant correspondences in different parts of the word, with tone differences also factored into the metric. Positive values were assigned to frequent sound correspondences, while relatively rare correspondences were given negative weights. Cheng posited that the greater the complexity of the rule system necessary to convert cognate strings between dialects, the lower the cross-dialect intelligibility would be. However, he did not verify this hypothesis through experimental testing. Tang and van Heuven (2015) correlated Cheng's phonetic distance measure with functional sentence intelligibility scores for 210 Chinese dialect pairs (see also Section 6.2.2), and indeed found a relatively strong correlation ($r = -.77$), which can be interpreted as confirming Cheng's predictions to some extent.

Early methods for measuring phonetic distances are rather complex and difficult to apply to all language situations in a uniform way. Some of these methods also lack objectivity because the researcher has to make subjective decisions about points assigned to sound correspondences. This is probably one of the explanations for the fact that in early work, methods for measuring lexical distances to predict intelligibility were more widespread than methods for measuring phonetic distances. Recently, various dialectometric methods have been developed that allow for more objective phonetic measures on large data sets using the computer. Though initially developed to measure dialect distances and characterize dialect areas, these dialectometric measurements can also be used to predict and explain intelligibility. Below, the focus is on the Levenshtein algorithm since this measure has most often been used for intelligibility research.

The Levenshtein distance (Levenshtein 1966), is a string metric for measuring the difference between two sequences of symbols. Levenshtein developed the measure to automatically identify lookalikes and soundalikes for spell checkers (and later for use in intelligent search engines, which will answer questions even with spelling mistakes). Within dialectology, it was applied for the first time by Kessler (1995) to measure phonetic distances between Irish dialects and has since been applied successfully to characterize dialect areas in many countries (Nerbonne and Heeringa 2010). It was first used to predict mutual intelligibility in a study on the mutual intelligibility between Dutch, Afrikaans, and Frisian (van Bezooijen and Gooskens 2005a, 2005b; Gooskens and van Bezooijen 2006) and between Scandinavian languages (Gooskens 2007), and has since become the most widely used phonetic algorithm for predicting and explaining intelligibility. Phonetic distances can be calculated using the freely available, open-source web application Gabmap (Leinonen, Çöltekin, and Nerbonne 2016). An alternative, more user-friendly application is LED-A (Heeringa, van Heuven, and van de Velde 2023).

With the Levenshtein algorithm, the phonetic distance between cognates in two languages is computed as the smallest number of string edit operations needed to convert the string of phonetic symbols in a word in language A to the cognate string in language B (Kruskal 1999). Possible string operations are deletions, insertions, and substitutions of symbols. In its simplest form, each string edit operation costs one penalty point. The total number of penalty points can then be divided by the length of the alignment (number of alignment slots) to yield a length-normalized Levenshtein distance (Heeringa et al. 2006). This accounts for the fact that one segmental difference in, for instance, a word of two segments has a stronger impact on intelligibility than one segmental difference in a word of ten segments. Length-normalized Levenshtein distances have been shown to correlate slightly better with intelligibility scores than non-normalized distances (Beijering, Gooskens, and Heeringa 2008). To obtain alignments that respect the syllable structure of a word or the structure within a syllable, vowels are aligned with vowels and consonants with consonants, but /j, i, w, u/ can be aligned with both vowels and consonants, and schwa can be aligned with any sonorant (including vowels). In this way, unlikely matches like [o] with [t] or [s] with [e] are prevented.

In Table 5.2, the Levenshtein algorithm is illustrated by a simplified example that ignores diacritics, comparing the phonetic transcriptions of the English word *interest* with the Swedish cognate *intresse*. In the fifth slot, /ə/ is replaced by /ɛ/, in the seventh slot, /ə/ is inserted, and in the eighth slot, /t/ is deleted. The total number of operations is then divided by the length of the alignment (number of alignment slots) to yield a length-normalized Levenshtein distance. As there are three operations and the alignment has eight slots, the distance is calculated as (3/8) x 100

Table 5.2: Illustration of the Levenshtein algorithm, showing three operations: one insertion (ins), one substitution (sub), and one deletion (del). The total normalized Levenshtein distance between the English and the Swedish word is 3/8 x 100 = 37.5%.

Slot	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
English	ɪ	n	t	r	ə	s		t
Swedish	ɪ	n	t	r	ɛ	s	ə	
Operation					sub		ins	del

= 37.5%. The word *interest* can be mapped onto *intresse* in many different ways, but the Levenshtein distance always gives the cost of the cheapest mapping.

The overall phonetic distance between language A and language B is the mean of the normalized distances for all cognate word pairs in the research corpus. Some researchers base their Levenshtein measurements on cognates as well as non-cognates. Such measurements have been used as a way to combine phonetic and lexical distance measurements. For example, using the Levenshtein algorithm, Abunasser (2015) measured lexical variation that takes pronunciation variation into account by comparing all words in the Swadesh list. However, it should be noted that correlations between Levenshtein distances and intelligibility will be high if non-cognates are included in the measurements (because non-cognates have a Levenshtein distance of almost 1). You then get a point-biserial correlation coefficient that is always very high. Therefore, to avoid inflation, phonetic distance should only be correlated with intelligibility for the cognate part of the vocabulary.

In the investigation by van Bezooijen and Gooskens (2007) discussed at the beginning of this section, the lexical transparency scores correspond well with Levenshtein distance measurements. The Frisian cognates were the least transparent for the Dutch participants and had higher Levenshtein distances (34.2%) than Afrikaans (20.9%). The link between the Levenshtein distances and transparency measurements, on the one hand, and intelligibility measurements, on the other hand, was not tested statistically. However, the statistical link between intelligibility measurements and Levenshtein distances was established in many later investigations. These investigations found high correlations, typically between $r = -.7$ and $r = -.9$.

In Figure 5.5, scattergrams show the correlation between inherent intelligibility and phonetic distances in the MICReLa investigation (see Chapter 3). The overall correlation for all language combinations is significant ($r = -.66$, $p < .01$) and lower but still significant ($r = . -51$, $p < .01$) if the outlier Czech-Slovak is left out of the analysis. These two languages are very similar phonetically and lexically, and including this language combination results in an artificial inflation of the correlation coefficient. The correlations are lower than for the lexical distances (see Figure 5.2) but still significant at the .01 level, $r = -.68$ for Romance and $r =-.62$ for Slavic (without the Czech-Slovak outlier). In the case of the Germanic language family, the correlation ($r = -.53$) does not even reach significance.

The results presented in Figures 5.2 and 5.5 confirm the claim by Lightbown and Spada (2006) that lexical differences are, in general, more successful in predicting intelligibility than pronunciation (see the beginning of Section 5.1). However, Mckaughan (1964: 118) observed that measurements at the phonetic level are more useful for determining intelligibility between very similar languages. This

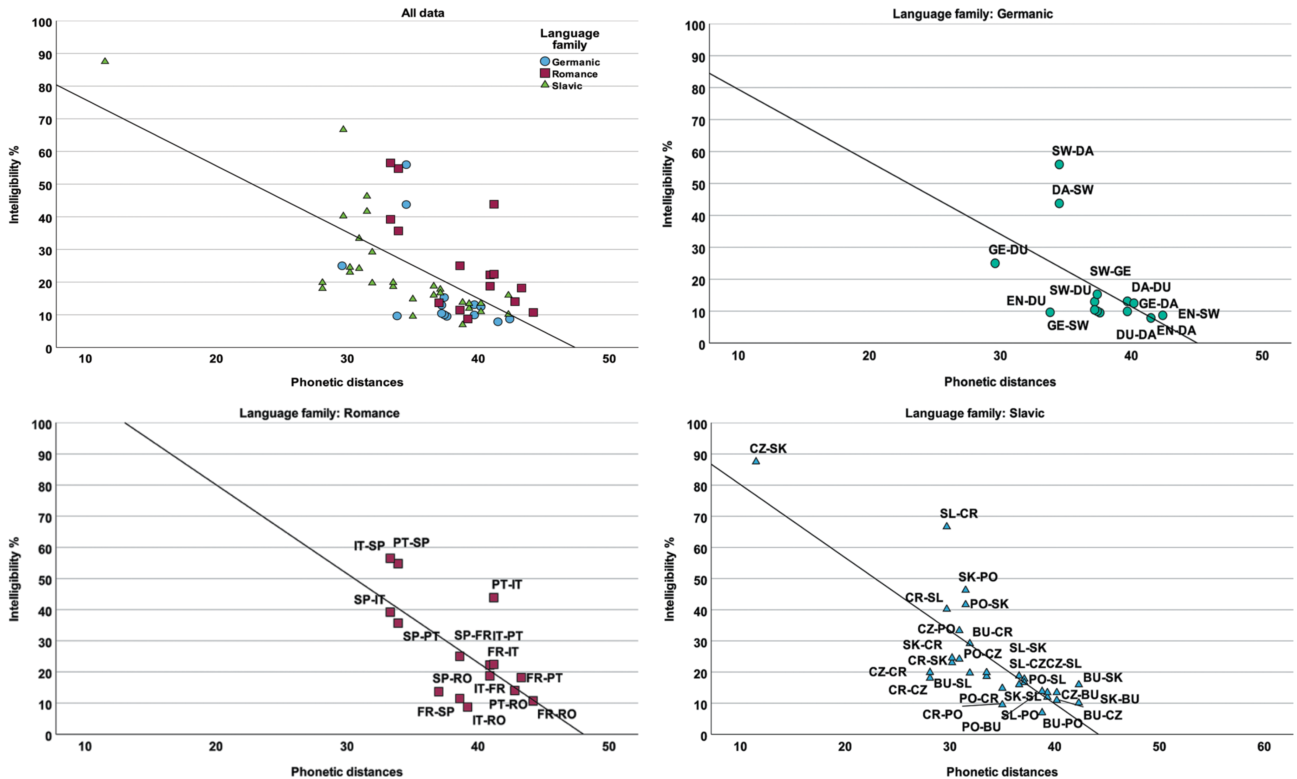

Figure 5.5: Correlations between inherent intelligibility scores and phonetic distances for 57 European language pairs in the MICReLa project ($r = -.66$), and separately for Germanic ($r = -.53$), Romance ($r = -.68$) and Slavic ($r = -.62$ without the Czech-Slovak outlier). For explanations, see Figure 4.2.

can be illustrated by a comparison of Figures 5.3 and 5.6 from an investigation by Gooskens (2007) on mutual intelligibility between Danish, Swedish, and Norwegian. Figure 5.6 shows that in the case of the Scandinavian language varieties, the correlation between intelligibility and Levenshtein distance is strong and significant ($r = -.80$, $p < .001$), much more so than the weak and insignificant correlation we found in the same study between intelligibility and lexical distance ($r = -.42$, $p = .110$, see Figure 5.3). This is a logical consequence of the fact that lexical distances cannot predict intelligibility if there is little variation (see discussion of Figure 5.3 in Section 5.1). Lexical distance and pronunciation are organized hierarchically. Pronunciation differences can only affect words that are lexically the same (cognates).The effect of pronunciation differences automatically increases when there are more cognates, and when there are no non-cognates, only pronunciation

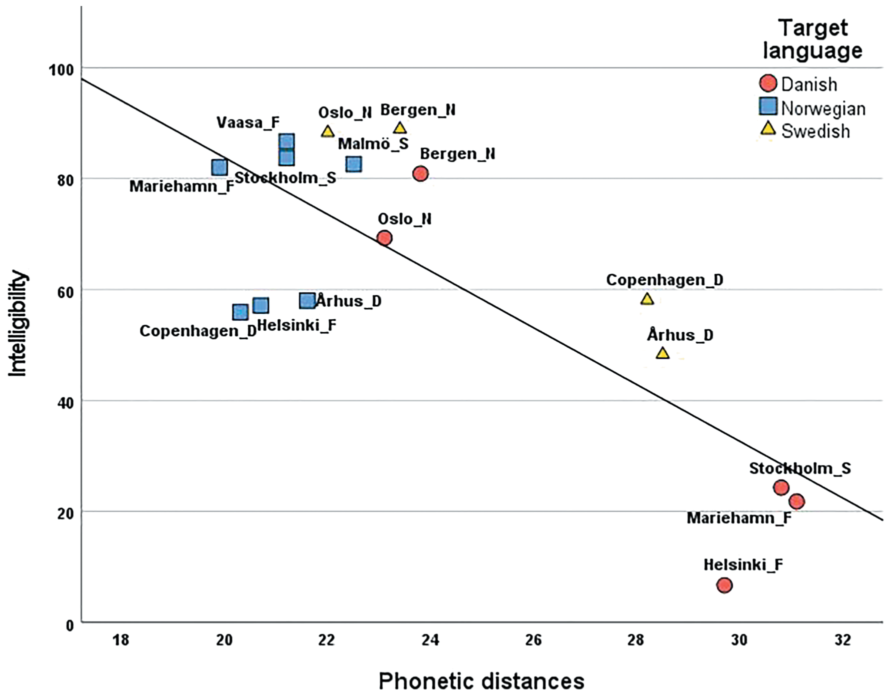

Figure 5.6: Correlation between intelligibility (percentage of correct responses in a cloze test) and phonetic distances (Levenshtein distance), $r = -.80$, $p < .01$, from an investigation of the intelligibility of Standard Danish, Standard Norwegian (represented by the Oslo variety) and Standard Swedish among listeners in nine different regions in either Denmark (D), Norway (N), Sweden (S) or Swedish speaking Finland (F). Language combinations involving Norwegian as a listener or target language have small Levenshtein distances and are understood best, while all other combinations have higher distances and score more poorly. Adapted from Gooskens (2007: 460).

difference can still have an impact. Therefore, it depends on the degree of lexical overlap between two languages how much the effect of sound difference between cognates can contribute as determinant of intelligibility.

5.2.1 Considerations when measuring phonetic distances

This section discusses some of the considerations to be made when measuring or interpreting phonetic distances to predict and explain mutual intelligibility between closely related languages. The example in Table 5.2 is a simplified version of the Levenshtein algorithm using binary differences between segments in the alignments (two segments are different or not). As suggested by Ladefoged (1968, 1970), it is possible to incorporate more linguistic details into the Levenshtein algorithm for measuring phonetic distance measurements by basing the costs of the basic operations on phonetic features. For example, the cost of replacing the sound [s] by [z] should be lower than the cost of replacing [s] by [k], since the first pair differs only in voicing, while the latter involves at least two phonetic differences (place and manner of articulation). The costs of differences between alignments can be assigned weights, for instance, between 0 and 1. More information is provided in Heeringa (2004), Nerbonne and Heeringa (2010), and Wieling, Margaretha, and Nerbonne (2012). These authors also discuss the processing of supra-segmentals and diacritics to reflect gradual segment distances. The refinements were mainly developed in dialectometry, i.e., to characterize dialect areas. It is possible that the algorithm needs other kinds of adjustments to optimally predict and explain intelligibility. Abunasser (2015) suggests optimization techniques to assign weight differences in phonetic features in order to explain mutual intelligibility between Arabic varieties. However, it is still unknown how differences should be weighted to optimally and universally reflect intelligibility between languages in different language families.

Möller and Zeevaert (2015) argued that to model intelligibility, it may not be optimal to assign weights to differences in the three phonetic dimensions (place of articulation, manner of articulation, and voicing) in the same manner as suggested by dialectometrists. Such weights are in line with experimental results concerning assessments of dialect distances (Gooskens and Heeringa 2004a). However, it may not conform to listeners' intuitions about cognate relationships and, therefore, be less optimal for predicting and explaining intelligibility. To draw conclusions about the relative importance of the three dimensions, Möller and Zeevaert proposed an approach where intuitions concerning possible cognate relations were elicited and used to weigh segment differences. They presented German readers with written words in several Germanic languages. The words generally differed in only one sound from German. Different types of written tests were carried out (free response, multiple-choice, judgments on the probability of two words being cognates) to assess the importance of different aspects of phonetic similarity for the transparency of written cognates. They incorporated the intuitions about similarity in the Levenshtein algorithm by assigning weights for distance measurement in line with the test results. For this particular group of listeners, these weighted Levenshtein distances showed better predictions of the transparency of the cognates. In particular, the same place of articulation made it easier for the listeners to recognize cognates correctly. Manner of articulation and voicing seemed to be less essential.

Even if an optimal and universal manner of assigning communicatively relevant weights to phonetic differences could be developed, certain sound correspondences may still have unexpected consequences for intelligibility. Gooskens, van Bezooijen, and van Heuven (2015) found that minor phonetic details that Levenshtein distances could hardly capture sometimes have a major impact on the intelligibility of isolated words in a closely related language. For example, Dutch *zoon* /zoːn/ 'son', which is transcribed phonemically with the same symbols as its German cognate *Sohn* /zoːn/, was poorly recognized by German listeners (21% correct identifications) in a word translation task with children as listeners (see Section 4.7). Their incorrect responses suggest that this must be due to subtle differences in the phonetic realization of Dutch and German /z/ and /oː/, which are normally not expressed in a phonemic transcription. In particular, the fact that Dutch alveolar fricatives /s, z/ have a less fronted articulation than their German counterparts (ten Cate and Jordens 1990: 57) seems to confuse German listeners who very often misperceived Dutch /s, z/ as German pre-palatal /ʃ, ʒ/. The German error responses (e.g., *schön, schon, John*) seem to be caused by this (pre-)palatal articulation in Dutch. In addition, the second half of Dutch /z/, as produced by the speaker in the experiment, was voiceless, which probably also influenced the responses given by the German participants. The solution may be to use more detailed phonetic transcriptions. However, this is not likely to improve the general predictive power of the algorithm since it depends on the specific language combinations what phonetic details are of importance for mutual intelligibility. An improvement of the algorithm to better capture intelligibility would require a complete perceptual assimilation study of how listeners match the sounds of the target language to their native sound system (Best 1995; Best, McRoberts, and Goodell 2001).

Some sound correspondences captured in the more detailed phonetic transcriptions may be more difficult to recognize than expressed by the algorithm because they are unfamiliar to the listener. For example, the retroflex consonants in Swedish are produced according to the phonological rule that /r/ and a following alveolar consonant merge into a retroflex sound in words such as *art* [ɑːʈ] 'sort', *bord* [buːɖ] 'table', *alternativ* [altəɳatiːv] 'alternative', *orsak* [uːʂɑːk] 'cause', and *parlament* [paɭamɛnt]. Danish does not have retroflexes, and such unfamiliar sounds may distract or confuse Danish listeners. This may have resulted in the incorrect translations of words with retroflex sounds found by Kürschner, Gooskens, and van Bezooijen (2008) in a word translation task, see Example 2.1 in Section 2.2.1.1. It may improve the predictive power of the algorithm to assign extra weight to correspondences with sounds that are unknown to the listener. However, this would result in an algorithm that is language-dependent.

Milliken and Milliken (1996) suggested that the degree of congruity, i.e., the difficulty that the listener has in discovering correspondences between sounds in their own language and the target language, should be reflected in the measurements. It is important to what extent correspondences between sounds in two languages can be generalized. They introduced the term "congruent correspondences" for exceptionless generalization, such as when /b/ in the language of the listener always corresponds to /p/ in the target language. "Incongruent correspondences" is used for situations where this is not the case, for instance, if /b/ in the target language sometimes corresponds to a /p/ and sometimes to /m/. Milliken (1988, reported in Milliken and Milliken 1996: 17) showed that the degree of congruity correlated well with tests of global intelligibility. The principles formulated by Milliken and Milliken (1996) are reminiscent of Cheng's (1997) systemic mutual intelligibility measure (see Section 5.2). Recently, conditional entropy measures have been used to model the regularity and frequency of sound correspondences. Such measures are discussed in Section 6.2.2.

Consonants have been found to contribute more to the transparency of cognate relationships and to carry more lexical information than vowels. Van Heuven (2008) explains this difference by the fact that syllables can contain several consonants but only one vowel, that consonants differ in more dimensions and are acoustically more contrastive than vowels, and that most languages have more different consonants than vowels. Cutler et al. (2000: 319) showed that, in a corpus of English words, there are about 2.2 times as many lexical neighbors (see Section 5.1.1) resulting from consonant replacement as from vowel replacement. This ratio is quite consistent across languages. Consonants serve as reference points within words, whereas vowels are more variable and prone to rapid changes over time compared to consonants (Ashby and Maidment 2005). Therefore, the occurrence of deviant segments in the consonantal structure of a word is likely to be more disruptive to the intelligibility than changes in the vowels. Ashby and Maidment (2005: 81) use the example of replacing all vowels in the sentence *Mary has a little lamb* by /ɛ/. They expect that most people will still understand the sentence. However, the sentence becomes unintelligible if all consonants are replaced with /d/ while maintaining the correct vowel qualities. Ashby and Maidment's assumption could be tested experimentally (using speech synthesis). Also interesting in this context is the existence of writing systems (e.g., Arabic) that only spell the consonants and omit the vowels or replace them with diacritics.

Gooskens, Heeringa, and Beijering (2008) examined the intelligibility of 18 Scandinavian language varieties among Danish listeners (see Example 2.15 in Section 2.2.3.4) and correlated the results with Levenshtein distances split up into consonant and vowel distances. When they calculated consonant distances, the distances were based on the alignment slots in which only consonants are involved, i.e., slots with either consonant insertion, deletion, or substitution. This distance was divided by the length of the full alignment. Similarly, vowel distances were calculated by considering only alignment slots involving vowel operations. In the example in Table 5.2, the length of the alignment of all segments is 8. Slots 5 and 7 concern operations involving vowels. Therefore, the vowel distance is 2/8 = 0.25 or 25%. For the consonant level, the distances may mutatis mutandis be calculated in the same way (1/8 = 0.11 or 11%).[14] The intelligibility results showed higher correlations with consonant distances than with vowel distances ($r = -.74$ vs. $r = -.29$), confirming that consonants convey more lexical information than vowels and therefore play a more central role in predicting intelligibility. Berthele (2011) and Möller (2011) found similar results for other Germanic language combinations. In Section 7.1.3, vowel and consonant insertion, deletions, and substitutions are included in a statistical model of inherent intelligibility based on data from the MICReLa project.

On the other hand Čéplö et al. (2016, see Section 2.3 and Figure 2.8) tested mutual intelligibility between three Arabic varieties and found that vowel differences generally affect mutual intelligibility more than consonants. This is an unexpected result given the Arabic root pattern morphology, where the root is a set of consonants arranged in a specific sequence that identifies the general meaning of the word and, therefore, plays a central role in lexical access. The stem's vocalic and syllabic features reflect additional information, such as part of speech and tense. For example, the Arabic words *kataba* 'write', *kaataba* 'correspond', *maktabun* 'office', *maktabatun* 'library', *kitaabun* 'book', *maktuubun* 'destiny', and *kuttaabun* 'Koran school' all have in common the root {ktb} with the semantic field of 'writing'. Any changes to the root would be expected to impede intelligibility. The authors explain the general finding that vowel differences are more important determinants of intelligibility than consonants in the case of Arabic, with the large

14 Alternatively, the vowel differences could have been normalized for the number of vowel slots and the consonants for the number of consonant slots. This would take into account that, in general, consonants are richer in number than vowels.

interdialectal and allomorphic variation in consonants that Arabic listeners seem to be well able to deal with.

The various investigations focusing on the relative contribution of consonants and vowels to intelligibility show that the contribution may differ across languages and language families. The size of consonant and vowel inventories can vary considerably, and so can the number of vowels and consonants used in running speech. Therefore, the weight that should be assigned to consonants and vowels may differ across languages.

The Levenshtein algorithm, as applied in dialectometry, assigns the same weight to different kinds of operations (insertions, deletions, and substitutions). However, in their investigation of the intuitions that participants have about the transparency of written cognates (see above), Möller and Zeevaert (2015) found that, in general, when participants look for word correspondences, they more often choose words that differ from the test item in a segment (= substitution) than correspondences that have a segment more or less in comparison with their native variety (= insertion/deletion). Segment insertions or deletions may change the structure of the words. This may alter the framework of the words and result in a different number of syllables in the language of the listener and the target language. For instance, cognates between Danish and Swedish sometimes differ in the number of syllables, cf. Danish bisyllabic *mængde* [mɛŋˈdə] vs. monosyllabic Swedish *mängd* [mɛŋd] 'quantity'. Kürschner, Gooskens, and van Bezooijen (2008, see Example 2.1 in Section 2.2.1.1) tested the intelligibility of 384 frequent Swedish words among Danes. They found that words with a different number of syllables in Danish than in Swedish were more difficult to understand than words with the same number of syllables.

Kaivapalu and Martin (2017) showed evidence that insertions have a larger impact than deletions on the perception of similarity between Finnish and Estonian. In a rating task, they found that Estonians who encounter insertions in Finnish words as compared to the Estonian cognates find it harder to see similarities than Finns who encounter deletions in Estonian words. In many test words, there is something added from the point of view of the Estonians but something deleted from the point of view of the Finns, either at the end of the word (e.g., in the elative case of Estonian *raamatu-st* 'from the book' – Finnish *raamatu-sta* 'from the bible') or both in the middle and at the end of the word (e.g., in the inessive case of Estonian *vangla-s* – Finnish *vankila-ssa* 'in the prison'). The authors conclude that the absence of a feature from the viewpoint of the native language is a smaller obstacle to perceiving similarity than the presence of an additional feature. Additional material in one of the words compared to the native language prompts participants unconsciously to search for a function for the extra material. However, the study was not designed to explore the value of deletions and insertions systematically. The relative contribution of insertions and deletions for explaining intelligibility still needs to be investigated experimentally.

The position of differences and similarities in a word may be important for intelligibility. Berthele (2011) and Möller (2011) note that listeners rely more on word beginnings than on later parts of words and that similarities of word onsets, therefore, have been found to be more essential for word recognition than similarities in the rest of the word. However, other research has found that auditory information in any part of the word contributes equally to the recognition process as long as the listener knows where the information fits in the temporal organization of the stimulus (Nooteboom and van der Vlugt 1988; Vanhove and Berthele 2015). Since experiments have mostly been conducted with native stimulus material or written language and the results are contradictive, the importance of the similarity of word beginnings for the intelligibility of spoken closely related languages remains unclear. In addition, it should be noted that even if segmental information contributes equally regardless of its position in the spoken word form, it is important that words are recognized as early as possible. In practice, longer words are recognized well before their end, see Marslen-Wilson William (1978). Therefore, the initial phase of the word gives important clues and helps to relieve working memory more than later parts of the word.

As already mentioned in Section 5.1.1, the position of word stress can play a role in word recognition (van Heuven 2022). Van Heuven and van Leyden (1996) showed that among Dutch listeners, correct recognition of low-quality synthesized words was severely reduced and delayed when stress was shifted to an incorrect position in Dutch words. In an investigation by Lepage and Grazia Busà (2013), word identification and reaction time results show that incorrect stress sometimes makes it hard to understand English with a French or Italian accent. These results predict that word recognition will also suffer from stress at another position in the word in the case of closely related languages. Kürschner, Gooskens, and van Bezooijen (2008, see Example 2.1 in Section 2.2.1.1) tested the intelligibility of 384 frequent Swedish words among Danish listeners. A logistic regression analysis showed that stress differences explain a small but significant part of the variance along with several other linguistic factors. Stress differences should, therefore, be included in the algorithm for calculating phonetic distances. Stress may be a central cue to lexical segmentation, i.e., the identification of word beginnings and endings, and this is important for recognizing the individual words in a sentence (Cutler 2012). Other cues to lexical segmentation can be prefixes, suffixes, and phonotactic constraints, such as phonotactically illegal consonant clusters at the end of one word and at the beginning of the next. Lexical segmentation cues are language-specific, and it may, therefore, be difficult for a listener to recognize words in connected speech of an unfamiliar target language (Field 2019: 291).

The use of pitch to differentiate lexical meaning, known as lexical tone, may be a critical aspect affecting the mutual intelligibility of tone languages. Yang and Castro (2008) computed tonal distance between dialects of tone languages spoken in the south of China in various ways. They found high correlations with functional intelligibility scores, with all measurements showing an approximate correlation of $r = -.7$. Tang and van Heuven (2015) correlated similar tonal distance measures with functional and judged intelligibility measures for all pairs of 15 Mandarin and non-Mandarin Chinese dialects (see Example 2.7 in Section 2.2.2.2), but found less strong correlations with intelligibility scores, probably because the distances were not measured on the same materials that were presented to the listeners. In an experimental setup, Wang et al. (2011) monotonized Chinese sentences to remove information about tones. They presented these (as well as the original sentences) to listeners in versions with high, medium, and low segmental quality. The results showed that tone information is fully redundant when segmental quality is high. However, when segmental quality is low, tone information greatly contributes to word recognition and sentence intelligibility. This is probably also the case when listening to a closely related tone language. It may improve the predictive power of the algorithm to incorporate tonal distances if the investigation includes languages with lexical tones.

In Section 4.4, evidence showed that orthographic knowledge may influence the intelligibility of spoken language. The fact that listeners can make use of their own orthography to understand the neighboring spoken target language could be included in the phonetic distance measurements. In that case, phonetic distances between word pairs could be a starting point. These distances could be reduced for sounds where the orthography can make it easier to find the corresponding word in the mother tongue. This can be illustrated by an example given by Gooskens and Doetjes (2009). They calculated orthographic and phonetic distances between 86 Danish and Swedish words that refer to everyday concepts. The distances were calculated using the Levenshtein algorithm. Next, the distances were corrected by assigning no weight to phoneme correspondences that could be recognized by means of the native orthography. The procedure for calculating these corrected distances is illustrated for a Danish listener in Table 5.3. The orthographic distance between the word for 'pan' from the perspective of Danish readers presented with the Swedish cognate is 40% (2 different letters, d/n and e/a, out of 5). The phonetic distance is higher (2 different sounds out of 4 = 50%). However, if Danish listeners hear this word spoken by a Swede, they will be able to recognize the /a/ (pronounced as [a]) as one of the two possible pronunciations of the written Danish <a> (the other possibility is [æ]). So even though [a] and [æ] are phonetically different, Danes may easily recognize both sounds as representatives of the written <a>. When calculat-

Table 5.3: Orthographic distance, phonetic distance, and phonetic distance corrected for orthography for a Dane presented with a Swedish cognate. Source: Gooskens and Doetjes (2009: 214).

	Orthographic distance	**Phonetic distance**	**Phonetic distance corrected for Danish orthography**
Danish reader/listener:	p a n d e	p æ n e	p a/æ n e
Swedish target language:	p a n n a	p a n a	p a n a
Differences:	0 0 0 1 1	0 1 0 1	0 0 0 1
Distance	2/5 = 40%	2/4 = 50%	1/4 = 25%

ing the phonetic distance corrected for orthography, this difference is, therefore, assigned no weight, resulting in a smaller distance (1 sound out of 4 = 25%).

The orthography-corrected distances showed higher correlations with the results of a word translation experiment with the 86 Swedish spoken words among a group of Danish listeners (r = −.63) than the uncorrected distances (r = −.54). These results provide evidence for the importance of considering orthography when interpreting intelligibility results.

The considerations in this section make it clear that it is difficult to make universal refinements to the Levenshtein algorithm that improve our ability to predict and explain the level of intelligibility of a closely related language. The approach of assigning extra weight to certain sound correspondences suggested by Möller and Zeevaert (2015) may result in a higher predictability of cognate recognition than binary measures or measures based on features. The Levenshtein algorithm may also be improved by incorporating other characteristics discussed above. The features may be assigned varying weights; for instance, consonants might receive a higher weight than vowels, while insertions and deletions could carry heavier weights than substitutions. Additionally, considerations, such as word stress and tone differences, could be taken into account. However, as Milliken and Milliken (1996: 18) note, it is difficult to make universal statements concerning the effects of specific phonetic differences on intelligibility since the significance of phonetic variation within which listeners will identify a sound with a given phoneme is often language-dependent. The range of tolerated variation for a given phoneme depends on what other phonemes there are in the language of the listener. This makes it difficult to apply an improved version of the Levenshtein algorithm automatically to several language pairs in a way that makes it possible to compare the communicatively relevant distances between different language pairs. To assign weights to the differences, experiments similar to those conducted by Möller and Zeevaert (2015) need to be carried out. They changed specific characteristics systematically to investigate their impact on intelligibility. Since the way we process speech is determined by our native vocabulary and repertoire of sounds, such experiments should take into account the human decoding processes as known from psycholinguistic research (see Lecumberri, Cooke, and Cutler 2010; Cutler 2012; Field 2019) and involve various target languages and groups of listeners with different language backgrounds. In this manner, it may be possible to get closer to an algorithm that assigns universal weights to phonetic differences that determine intelligibility.

5.3 Morpho-syntax

Most investigations of linguistic determinants of intelligibility have focused on lexical and phonetic distances. These two linguistic levels are better predictors of intelligibility than morphological and syntactic distances. Without word recognition, there is no intelligibility. Barcroft (2004: 201) illustrates the importance of vocabulary by comparing two types of errors in these sentence pairs: *It snow / *It nevs and *He want spoon / *He wants a fork. The first sentence in each pair is grammatically incorrect but still comprehensible, while the vocabulary errors in the second sentences distorts the intended meanings ('It snows' and 'He wants a spoon').

However, morpho-syntax still plays a role in predicting intelligibility and should, therefore, not be ignored. Hilton, Gooskens, and Schüppert (2013) conducted reaction time and plausibility evaluation experiments (see Example 2.10 in Section 2.2.2.4) to investigate whether certain Norwegian grammatical constructions not used in Danish may impede Danes' comprehension of Norwegian sentences. Their findings revealed that Danish listeners needed longer decision times and made more mistakes in a plausibility evaluation task when presented with sentences featuring Norwegian word order and morphology not used in Danish. It is easy to argue on logical grounds that a random scrambling of words leads to problems with speech understanding. However, the question is when morpho-syntactic differences become so great that misunderstandings and incomprehension arise. In this section, ways of measuring morphological and syntactic distances are discussed.

5.3.1 Morphology

An early attempt to develop morphological distance measures and correlate these with intelligibility measures was carried out by Bender and Cooper (1971) on six Sidamo languages, which belong to a subbranch of East Cushitic languages spoken in Ethiopia. Bender and Cooper used translations of the same six English texts into the six languages. They computed three similarity measures (the counterpart of distance measures) for each language pair. These measures were based on the proportion of morpheme types (stem morphemes or affixes), which were shared by the corresponding word pairs in the translations of the story (see Figure 5.7).

– The proportion of shared stem morphemes (morphemes that form the base of a word and carry its lexical meaning)
– The proportion of shared affixes (elements attached to stems to modify meanings)
– The proportion of all shared morphemes (roots and affixes combined).

Figure 5.7: Morpheme similarity measures proposed by Bender and Cooper (1971: 42).

Bender and Cooper played tape-recorded passages of the six languages to native listeners of each of the six languages in a Latin square design (see Section 2.1.3.3). Intelligibility was tested by means of multiple-choice questions about the texts. The three morpheme similarity measures for 25 language combinations were then correlated with the cross-lingual intelligibility measures. There was a substantial correlation between each of the similarity measures and intelligibility. Shared stems contributed more (r = .51) than shared affixes (r = .43). This means that both the stem morpheme and the grammatical affix contributed to intelligibility, but that the stem contributed more than the affix. The correlation with the combined stems and affixes was highest (r = .67).

It is also of interest to note that the similarity measures between grammatical morphemes and between root morphemes do not correlate with each other in Bender and Cooper's investigation (r = −.03). Similar results were found by Heeringa et al. (2014) who measured orthographic Levenshtein distances between five Germanic languages separately for stems and inflectional affixes. Their results showed that stem distances between these languages do not correlate with affix distances. The two investigations differ in the kind of affixes included in the measurements. Bender and Cooper (1971) included all affixes, and Heeringa et al. (2014) only included derivational affixes. Still, when modeling intelligibility, both investigations show that it is important to be aware of differences between stem distances and affix distances. Since the two measurements are independent, the total distances between whole morphemes (stems and affixes) explain all variance in the intelligibility scores that distances would also explain if calculated separately for stems and affixes. This is the case if a running text is used as the basis for measurements. Therefore it is safe to measure Levenshtein distances on the basis of whole morphemes in the case of whole texts. However, if the measurements are based on word lists that mostly have no variation in inflectional affixes, it is important to realize that the influence of affixes is reduced.

Van Bezooijen and Gooskens (2007, see also Section 5.2) suggested looking at the transparency of the grammatical meaning to understand its influence on intelligibility. They distinguished two types of transparency (see Figure 5.8). If the grammatical meaning was transparent, 0 points were assigned; if it was opaque, 1 point was assigned. The two types of grammatical transparency were scored separately and summed. The total transparency per word thus ranges between 0 and 2.

- Is it clear which word category (noun, verb, etc.) is involved? The Afrikaans *sing* 'sing' is likely to be recognized as a verb by Dutch listeners because it corresponds to Dutch *zing* and is therefore scored as transparent (0 points). An example of opacity is Afrikaans *die* (Dutch *de),* which Dutch listeners will interpret as a demonstrative pronoun rather than as a definite article because the feminine/masculine form of the Dutch demonstrative pronoun is *die* (1 point).
- Is it clear what tense, number, gender, person, etc. is involved? An example of transparency is Afrikaans *is*, which is the same as the Dutch singular form of the verb 'to be' and, therefore will be interpreted as such by Dutch listeners (0 points), whereas in Afrikaans the form is also used for the plural and is therefore opaque for Afrikaans listeners (1 point).

Figure 5.8: Measures of the transparency of the grammatical meaning. Adapted from van Bezooijen and Gooskens (2007: 256).

Even though the intelligibility results showed that it is easier for Dutch readers to understand Afrikaans than Frisian, there are more differences between Afrikaans and Dutch when it comes to the transparency of grammatical meaning (a mean score of 1.71 on a scale from 0 to 2) than between Frisian and Dutch (1.95). However, the differences with Afrikaans are mainly simplifications. For Dutch listeners (or readers), the simplified morphological system may be unusual, but in the end, it may have little effect on text comprehension. A text generally contains so much grammatical redundancy that the absence of explicit marking, for example, of the number in the verb system presents no problem. This can be illustrated by Afrikaans *werk,* which corresponds to Dutch *werk* (1st person singular or imperative)*, werkt* (2nd and 3rd person singular)*, werken* (infinitive, 1^{st,} 2nd and 3rd person plural). The morphological differences between Frisian and Dutch are of a different nature. Here, Dutch listeners encounter meaningful suffixes that they are unfamiliar with as verb endings, which may confuse them. To illustrate, some Frisian verbs have the ending *-je* in the infinitive, plural, and imperative, e.g., Frisian *wurkje*, whereas Dutch marks the infinitive/plural with the suffix *-en* as in *werken* and the imperative with either the zero suffix or *-t* as in *werk/werkt* 'work'. This may confuse a Dutch listener since the suffix *-je* is a diminutive marker that only attaches to nouns in their own language. The Dutch listener will experience difficulty reconciling the context that sets up the expectation of a verb form with the perception of a noun. This shows that a quantitative approach to measuring affix differences from the target language may be improved by assigning different weights to different kinds of differences. As in the case of distances at other linguistic levels, this makes it difficult to develop universal measures that take into account various refinements, and the weights can best be established experimentally (see Section 2.1.1.3).

5.3.2 Syntax

Research on the relationship between intelligibility and syntactic differences is limited because there has been little work on syntactic measurements in dialectometry. However, various dialectometric methods for measuring syntactic distances have recently been developed and applied in intelligibility research. Heeringa et al. (2018) developed two measures. The movement measure reflects the fact that listeners exposed to a sentence in the related target language may sometimes experience that certain words have "moved" compared to the corresponding sentences in another language. It measures the average number of word positions a word in a sentence in language A has moved compared to the corresponding sentence in language B. The indel measure captures the experience that listeners may have of words being added or removed when comparing sentences in a foreign language to their native language. This measure quantifies the average number of words inserted or deleted in sentences of one language compared to their equivalent sentences in another language.

The movement measure for the English sentence *It would be difficult* and the Dutch equivalent *Het zou moeilijk zijn* can be illustrated with the example in Table 5.4, adapted from Heeringa et al. (2018: 286). The verb *be* in the English sentence is moved 2 positions compared to the Dutch equivalent *zijn*. In the indel measure there is one deletion and one insertion. In this case, both measures result in a distance of 2.

Nerbonne and Wiersma (2006) introduced the trigram measure for measuring the impact of the native language on the acquisition of the syntax of a second language. This is a measure of aggregate syntactic distance where trigrams (sequences of three lexical category labels, such as "noun", "verb", and "pronoun") are inventoried and counted. Syntactic distance between two languages is then

Table 5.4: Illustration of the movement and indel measures in the sentences *It would be difficult* and *After a while it will become easier.*

English	*It*	*would*	*be*	*difficult*		
Dutch	*Het*	*zou*		*moeilijk*	*zijn*	
			>	>	2	Movement = 2
	0	0	1	0	1	Indel = 2

defined as 1 minus the Pearson correlation coefficient between the trigram frequencies in the two languages.

The example in Table 5.5 from Heeringa et al. (2018: 286) illustrates the trigram measure. Trigram frequencies are established in a small corpus of English consisting of two sentences: *It would be difficult* and *After a while it will become easier*. In the first sentence, there are four trigrams. In the second sentence, there are seven trigrams. Note that also the $, which marks the beginning of a sentence, and the #, which marks the ending of a sentence, can be parts of a trigram.

Table 5.5: Illustration of the establishment of trigrams in the sentences *It would be difficult* and *After a while it will become easier*.

	It	**would**	**be**	**difficult**	
$	pronoun	modal verb			
	pronoun	modal verb	verb		
		modal verb	verb	adjective	
			verb	adjective	#

	After	**a**	**while**	**it**	**will**	**become**	**easier**	
$	preposition	determinant						
	preposition	determinant	noun					
		determinant	noun	pronoun				
			noun	pronoun	modal verb			
				pronoun	modal verb	verb		
					modal verb	verb	adjective	
						verb	adjective	#

An inventory of the trigrams found in the two sentences shows eleven different types of trigrams, two of them occurring in both sentences (see Table 5.6). The last column shows the frequency of each trigram in the small English corpus. Most trigrams occur once in this small corpus, but the trigrams pronoun/modal verb/verb and modal verb/verb/adjective appear twice. Similar lists of frequencies are made for the languages to be compared to. How high each frequency is

Table 5.6: Inventory of trigrams in the two sentences *It would be difficult* and *After a while it will become easier*.

Trigram			Sentence 1	Sentence 2	Frequency
$	pronoun	modal verb	X		1
pronoun	modal verb	verb	X	X	2
modal verb	verb	adjective	X	X	2
verb	adjective	#	X		1
$	pronoun	determinant		X	1
preposition	determinant	noun		X	1
determinant	noun	pronoun		X	1
noun	pronoun	modal verb		X	1
pronoun	modal verb	verb		X	1
modal verb	verb	adjective		X	1
verb	adjective	#		X	1

depends on the relative frequency of each trigram in the corpus and the size of the corpus. Once a frequency list for each language has been made, the distance between each language combination can be calculated as 1 minus the Pearson's correlation of the two matched frequency lists.[15]

Gooskens and Swarte (2017) measured mutual inherent intelligibility between five Germanic languages (20 language combinations) using a spoken cloze test among groups of listeners with the same five native language backgrounds. They correlated the intelligibility scores with the three syntactic distance measures. Intuitively, all three measures may be adequate predictors of intelligibility. However, together with the indel measure, the trigram measure showed the highest correlation with intelligibility ($r = -.72$ for the trigram measure and $r = -.74$ for the indel measure, $p < .01$). The correlation for the movement measure was lower ($r = -.61$, $p < .05$).

Since the trigram measure had almost the same correlation with intelligibility as the indel measure and the procedure for calculating the distances is easier than for the other two measures (movement and indels), the trigram distances were used to predict intelligibility results in the MICReLa project. Both the movement and the indel measure require comparison of sentences using a procedure where words in stimulus language A are aligned to the corresponding words in language B. This alignment is difficult to automate and, therefore, may have to be

15 It is possible for trigram size to correlate negatively between two languages. Theoretically, the correlation could even be perfectly negative if language B orders all sentences of language A backward (mirror image order). To avoid a negative correlation, a solution could be to formulate an assumption: if $(1 - r) < 0$ then trigram distance = 0. The assumption is then that every negative correlation in trigram frequencies hinders intelligibility to the maximum.

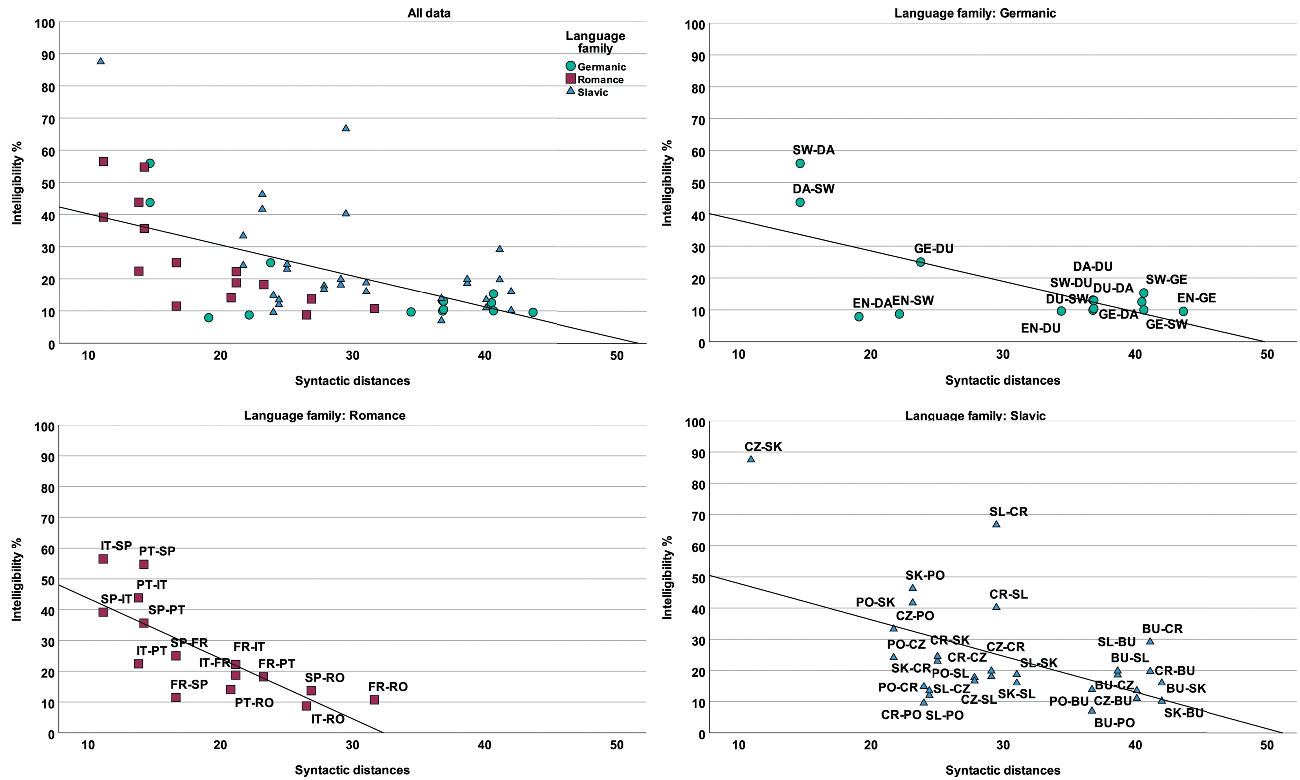

Figure 5.9: Correlation between inherent intelligibility scores and syntactic distances for 57 European language pairs in the MICReLa project ($r = -.56$), and separately for Germanic ($r = -.68$), Romance ($r = -.77$) and Slavic ($r = -.52$). For explanations, see Figure 4.2.

done manually. Alignment is not required for the trigram measure. Parallel corpora are not even required if the samples are sufficiently large. This makes it simpler and less labor-intensive to use the trigram measure.

In the MICReLa project, significant correlations were found between syntactic trigram distances (see Appendix D) and inherent intelligibility, both when including all language combinations ($r = -.56$, $p = .004$) and for the individual language families ($r = -.68$, $p = .008$ for Germanic language combinations, $r = -.77$, $p < .001$ for Romance language combinations, and $r = -.52$, $p = .004$ for Slavic language combinations), see Figure 5.9. This shows that intelligibility can to a large extent be predicted and explained by the syntactic trigram measure.

5.4 Paralinguistics

Paralinguistic factors include speech phenomena, such as pitch, volume, speech rate, and fluency. They may also include non-vocal phenomena, such as facial expressions, movements of eyes (gaze) and mouth, and hand (and body) gestures (Lyons 1977). Differences between these factors may be highly individual, but they may also be characteristic of whole languages. For example, some women speak with a higher pitch than other female speakers of the same language. However, van Bezooijen (1995) found that Japanese women, in general, speak with higher pitches than Dutch women due to sociocultural differences between the two countries. Mennen, Schaeffler, and Docherty (2012) found cross-language differences in pitch range between female speakers of English and German. Pitch differences have also been found between varieties that are even more closely related. Measurements by van Bezooijen (2000) show that the pitch of female speakers from West Flanders is 40 Hz higher than the pitch of standard Dutch speakers from the Netherlands and that the pitch of dialect speakers from Groningen in the north of the Netherlands is 25 Hz lower than that of speakers from Limburg in the south of the Netherlands. Many linguists stress the importance of paralinguistic factors for successful communication (e.g., Crystal 1975). However, there has been limited research conducted to experimentally evaluate their impact on the intelligibility of a closely related language.

5.4.1 Speech rate and segmentation

The relationship between speech rate and intelligibility is parabolic (a U-shaped curve). If the speech rate is fast, it will influence the intelligibility of a message negatively because the listeners cannot retain the previous words and speech fragments in their working memory. With somewhat slower speech, the intelligibility improves, only to decrease again at an even lower speed (Yenkimaleki 2017). Individual speakers can change their speech rate according to different circumstances. They can, for instance, choose to speak slowly to make themselves clear or to speak fast to bring across information as quickly as possible. However, there is also evidence that speech tempo can vary across related languages. For instance, Verhoeven, de Pauw, and Kloots (2004) found that Netherlandic Dutch varieties are articulated at a significantly faster rate than Belgian Dutch ones, Robb, Maclagan, and Chen (2004) found that New Zealand English is spoken at a faster rate than American English, and Santiago and Mairano (2022) found that Spaniards articulate faster than Mexicans when speaking Spanish.

Danish is often mentioned as an example of a language with many reduction and assimilation processes compared to other languages. Hilton, Schüppert, and Gooskens (2011) and Schüppert et al. (2012) presented evidence that Danes generally speak with a faster speech rate than speakers of the other Scandinavian languages and, therefore, assimilate and reduce their speech more. To measure the speech rate in the three Scandinavian languages, they collected material from the non-commercial public service radio stations Danmarks Radio (DR), Sveriges Radio (SR), and Norsk Rikskringkasting (NRK). They calculated the speech rate by counting the number of phonological syllables that the speakers produced per second (canonical speech rate). The results showed that the speech rate of the Danes was significantly faster (6.21 syllables per second) than that of the Norwegians and Swedes (5.37 and 5.35 syllables per second). Since some phonological syllables are often assimilated or even not pronounced, they also calculated the phonetic speech rate, i.e., the number of syllables that were actually pronounced per second. In contrast with the canonical speech rate, the researchers found no significant differences in phonetic speech rate between the three Scandinavian languages (the number of syllables per second is 4.38 for Danish, 4.41 for Norwegian, and 4.48 for Swedish).

Speaking quickly increases the demands on the articulatory apparatus; therefore, the speaker is likely to reduce and assimilate specific sound entities when speaking fast. Speech reduction and assimilation make it difficult to find lexical boundaries between words, resulting in difficulties in intelligibility. Danes transmit more phonological syllables per time unit than Swedes and Norwegians, but the same number of phonetic syllables. This confirms that Danish is pronounced with more assimilations and deletions than the neighboring languages. This makes it difficult to segment the speech signal – something which is often mentioned in a humoristic way among Scandinavians (see, for example, Figure 5.10 and the Norwegian parody of the Danish language in the Norwegian sketch comedy television program Uti vår hage[16]). Bleses et al. (2008) report a delay in vocabulary development in Danish infants and children compared to that of their peers from ten other European countries, the US, and Mexico, and they suggested that this delay could be attributed to these characteristics of Danish.

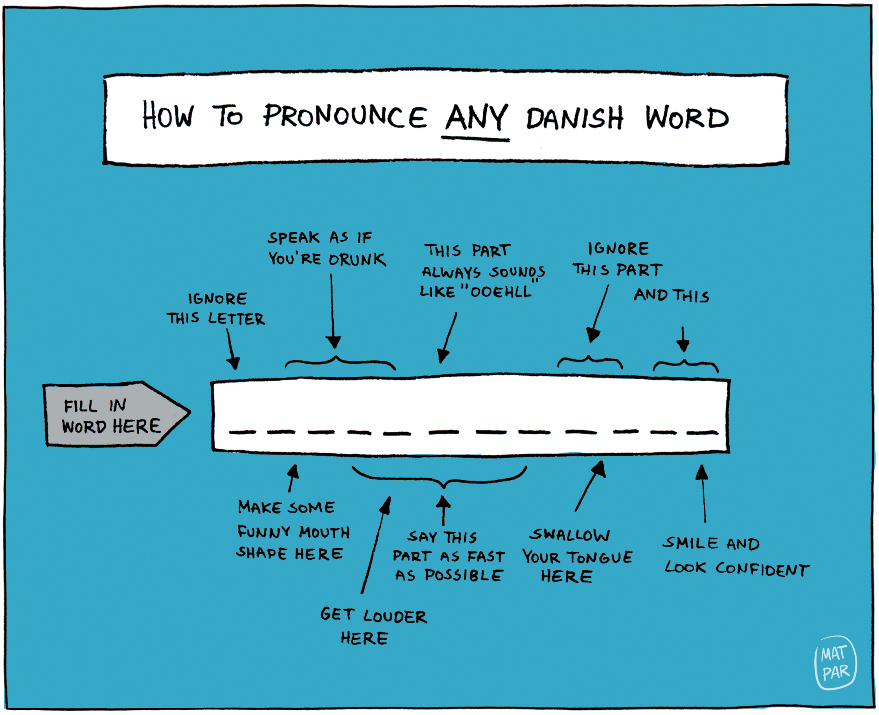

Figure 5.10: A caricature of the Danish language, showing the impression that many Scandinavians (including the Danes themselves) have of Danish as a language with a lot of reduction and unclear pronunciation. Made for the Copenhagen Language Center, copyright Matthias Parchettka.

A fast speech rate and unclear pronunciation due to assimilation and reduction phenomena are both likely to impair intelligibility and the two are also expected to intercorrelate. A fast speech rate results in less processing time and a higher demand on the working memory of the listener. In the case of a fast phonological

16 https://youtu.be/wGGX5gmwVbA

speech rate, the speaker often reduces the speech signal, and this may also make it harder for the listener to recognize words in the speech signal (Cutler 2012). In an experimental setup, Schüppert, Hilton, and Gooskens (2016) teased apart speech rate and reduction to test the role of the two factors in intelligibility. They recorded 50 semantically unpredictable sentences (see Section 2.2.2.3) that were read aloud by a native speaker of Danish in two different yet natural conditions, namely slowly and clearly (condition 1) and quickly and unclearly (condition 2). Conditions 1 and 2 were manipulated to form conditions 3 and 4. Each slowly and clearly produced sentence (condition 1) was time-compressed linearly by reducing the total duration to the duration of the same sentence produced quickly and unclearly (condition 2), resulting in condition 3 (quickly and clearly). Similarly, each quickly and unclearly produced sentence (condition 2) was time-expanded by increasing the total duration to the duration of the same sentence produced slowly and clearly (condition 1), resulting in condition 4 (slowly and unclearly).

Each sentence was presented to Norwegian and Swedish listeners in the four conditions in a Latin square design. Intelligibility was tested by means of a translation task. Not surprisingly, the results showed that spoken Danish is more intelligible when produced slowly and clearly than when produced quickly and unclearly. This suggests that either a fast articulation rate by Danish speakers or many reductions associated with a fast articulation rate, or both, are likely to impede the intelligibility for Norwegian and Swedish listeners (see Figure 5.11). Comparing the results of the four versions made clear that speech rate had a larger impact on the intelligibility of spoken Danish than reduction. The mean intelligibility is lowest in the two conditions (2 and 3) with fast a speech rate. That means that speaking slowly increases intelligibility to a greater extent than speaking clearly does, although the most efficient way of improving intelligibility is to speak slowly and clearly (condition 1).

The results by Schüppert, Hilton, and Gooskens (2016) show that in addition to poor segmentability, a short time frame makes it challenging for listeners to decode the message. They need to decompose and process the stream of speech sounds more quickly, which is more demanding on the working memory. In his hyper and hypo-articulation (H & H) theory, Lindblom (1990) argues that speakers of any language are constantly balancing between “hyperspeech”, i.e., clear articulation to maximize intelligibility in the listener, and “hypospeech”, i.e., unclear speech to minimize the articulatory effort for the speaker.[17] Generally, in communication be-

17 This insight was later also explicitly captured in the (very influential) theory of phonological optimality (Prince and Smolensky 2004). The optimal phonetic shapes of words are determined by effort constraints (designed to minimize the speaker’s articulatory effort) and faithfulness constraints (designed to keep the word shapes as recognizable as possible for the listener). The authors do not refer to Lindblom’s H&H theory.

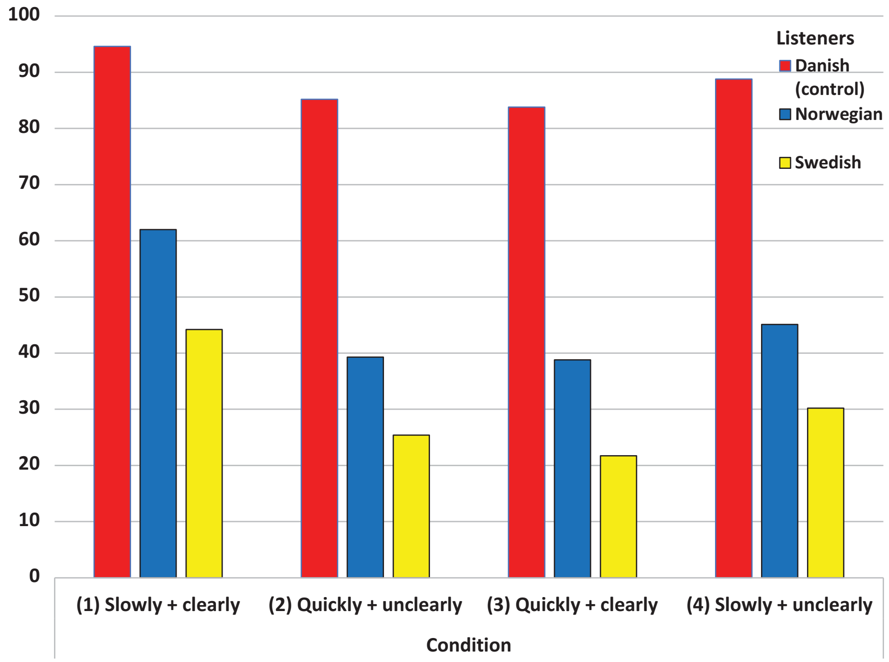

Figure 5.11: Mean intelligibility per listener group in four conditions: (1) slowly and clearly, (2) quickly and unclearly, (3) quickly and clearly, and (4) slowly and unclearly. Based on results from Schüppert, Hilton, and Gooskens (2016: 55).

tween native speakers, these two opposing demands lead to speech which contains a certain amount of reduction phenomena, but is still fairly intelligible to the listener; the speaker wants to be intelligible enough with as little effort as possible. Danish may be an example of a language where hypospeech is more dominant than hyperspeech, and this may be the explanation for the fact that it is difficult for Swedes to understand Danish and for children to acquire the native vocabulary, as shown by Bleses et al. (2008), see Section 6.2.2.

Based on the observations about the difficult segmentability and fast speech rate of Danes, it may be expected that even adult native speakers of Danish might have difficulty understanding Danish in difficult listening circumstances, for instance, if there is a lot of background noise. Gooskens et al. (2010) tested this hypothesis in an intelligibility experiment with Danish and Swedish materials presented in descending levels of noise to native listeners of Danish and Swedish, respectively. The results showed that Danish remains as intelligible to Danish listeners as Swedish does to Swedish listeners, even in (very) difficult listening circumstances; therefore no evidence was found to support the hypothesis. It seems that adult Danish native listeners can deal well enough with the characteristics of the Danish language. In this view, it could be argued that Danish pronunciation is optimal because the speaker has to do less work to remain intelligible, and it has been shown to be perfectly understandable for its own target group. Swedish and Norwegian protect the hearer and expect more work from the speaker than necessary for intelligibility (a suboptimal solution). However, there is also evidence that Danish is processed differently from other languages, even by adults. See Trecca et al. (2021) for a review of investigations demonstrating this.

Scharpff and van Heuven (1988) found that in low-quality computer-generated speech, the insertion of pauses at prosodic boundaries enhances intelligibility. Such prosodic boundaries can divide the speech stream into units of information, such as phrases and sentences. Gooskens and van Bezooijen (2014) tested the hypothesis that these findings can be generalized to a situation where listeners hear a closely related language. They presented naturally produced Danish sentences to Swedish listeners in two versions, one with and one without pauses at prosodic boundaries. The inclusion of pauses indeed led to a higher percentage of correct translations. The insertion of correctly placed speech pauses may improve intelligibility because it reduces uncertainty about word boundaries and instructs the listener to integrate the preceding words into one meaningful unit. If the same information is received in the middle of a constituent, speech understanding is compromised.

5.4.2 Visual context cues

It has been shown that visual context cues play an important role in the communication between native speakers (Drijvers and Ozyürek 2017). Drijvers (2019) distinguishes between two kinds of visual context cues that listeners use when understanding speech: co-speech gestures and visual speech.

Co-speech gestures are gestures that people spontaneously produce when speaking and often depict properties of objects, motion, action, and space. They are implicitly acquired and speakers are often unaware that they are gesturing. Even though only few studies consider variation in gesture use between different national speech communities, it is clear that co-speech gestures may vary across cultures (Kita 2009). Apart from the form and timing of gestures, cultural differences can also be found in gesture frequency (e.g., So 2010).

Visual speech is the information provided by lip and jaw movements. It can provide temporal details about the speech signal (e.g., amplitude) and some phonological information about the location of a speaker's articulators (labial/labiodental versus other) and about lip rounding. In addition, the viewer can roughly estimate whether the speaker produces a vowel or a consonant (by observing the sequences of peaks and troughs in the jaw movement). When available, such information assists a listener in processing what a speaker says. It has been observed that the amount of visual information available to a listener is the inverse of the amount of auditory information (van de Rijt et al. 2019). Vowels are recognized primarily by the second formant (F2, which leaves no trace visually), next by the first formant (F1, which is seen in jaw aperture), and last by the third formant (F3, which is seen most easily as lip rounding vs spreading), see Breeuwer and Plomp (1986).

The information conveyed by both co-speech gestures and visible speech has been shown to enhance comprehension of speech in adverse listening conditions and in non-native speech (Drijvers and Ozyürek 2017, 2020). Gooskens and Voigt (unpublished) tested the role of co-speech gestures (focusing on motion events) and visual speech for the intelligibility of spoken Spanish for Italian listeners. The experiment consisted of four conditions (see also Figure 5.12):

1) (FULL) the participants listened to the audio and saw the full upper body (gestures and face) of the speaker while she retold a story;
2) (BODY) the participants listened to the audio and saw the gestures; the face of the Spanish speaker was blocked out;
3) (FACE) the participants listened to the audio but could only see the face of the speaker;
4) (NONE) the participants listened to the audio but saw a blank screen.

The mean percentages of correct answers are presented in Figure 5.12. The results showed that when listeners had all the information at their disposal (FULL), they answered significantly more questions correctly than in the conditions where they missed information from co-speech gestures (FACE and NONE). This suggests that especially co-speech gestures can enhance intelligibility in a related language, at least in the subdomain that has to do with movements.

These results show that visual contextual cues, e.g., as in watching video films, can add to the intelligibility of a foreign language and thus should be considered as a factor in future intelligibility research. However, many spoken intelligibility tests are presented to the listeners via audio recordings only, even though technological advances make it easy to show video recordings of the speaker. Thus far, only little attention has been paid to the role of visual cues in the intelligibility of a closely related language.

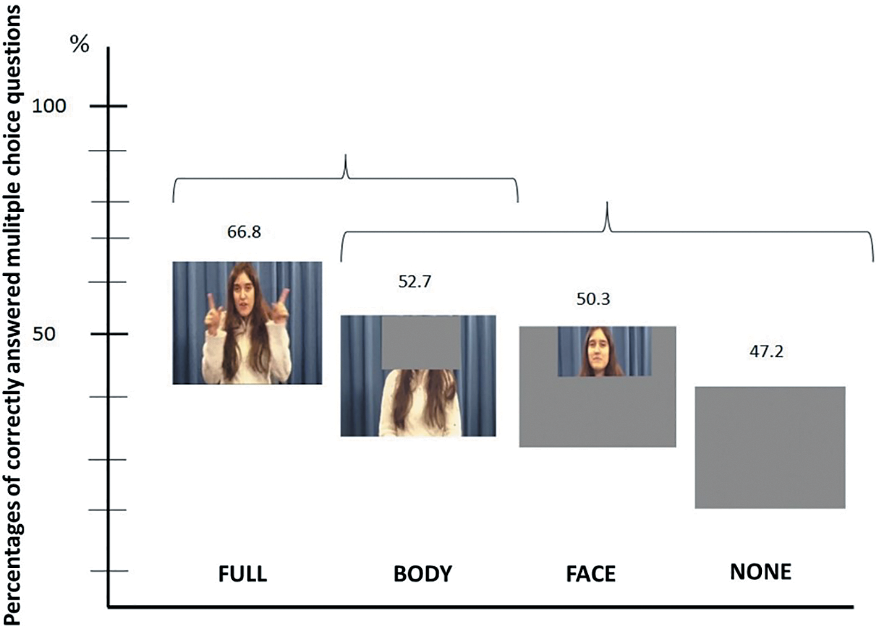

Figure 5.12: Percentage of correctly answered multiple-choice questions for four experimental conditions. Braces include conditions that do not differ from each other significantly ($p < .01$).

5.5 Conclusions

Several studies have established a connection between the intelligibility of a closely related language and linguistic distances as established by various measuring methods. Many of these studies have concentrated on lexical distance and found this to be an important predictor of intelligibility. This makes sense since without word recognition, listeners will not be able to understand a message, and the more words they recognize, the more they will understand. Therefore, the number of noncognates is a good predictor of intelligibility. However, if the pronunciation of the cognates is too deviant, the listener may also not be able to recognize them, and therefore, phonetic distance is a good predictor of intelligibility as well. Linguistic differences at other linguistic levels may also impede comprehension. This chapter has discussed methods for measuring distances at various linguistic levels that have been used as predictors of intelligibility. With the advancement of measurement techniques, there is an opportunity for further exploration of the link between intelligibility and less frequently examined linguistic features. Chapter 7 delves further into the interaction between distances at the various linguistic levels and presents simple statistical models of intelligibility.

Part II: **Further analyses**

Chapter 6
Asymmetric intelligibility

It may sound logical to assume that language A should be as intelligible to native listeners of language B as the other way around. However, this is not always the case, as it becomes clear from many observations by speakers themselves. Discussion on social media shows that many people are intrigued by the phenomenon of asymmetric intelligibility. For instance, a short YouTube video titled *Why some speakers can't understand speakers who understand them – Asymmetric Intelligibility*[18] has reached over 1.2 million views. In the comments section, almost 8000 viewers from all over the world participated in the discussion about reasons for asymmetry and provided examples of asymmetric intelligibility between languages and dialects. In many of the comments, asymmetric intelligibility is explained by extra-linguistic factors, as becomes clear from the following examples suggesting the importance of experience and motivation:

> *And then there is the exposure to "midway" languages . . . I'm Czech, I don't speak but can understand Slovak and Polish. Therefore I have an easier time understanding Ukrainian than Czechs who aren't exposed to Polish. The more slavic languages a slav is exposed to the easier it is to understand the rest.*

> *The reason I and many of my friends (that are norwegian) its because when we were young there was a lot of popular shows in danish and swedish without norwegian dubs.*

> *I remember going to legoland as a kid. This other kid there wanted to be my friend so bad but I couldn't understand him at all while he understood me perfectly. As a grownup now I don't have any particular issues with danish, but that is more through exposure than anything.*

> *I think many people in advance tell themselves that they won't understand so they might not even try. Understanding a language that is similar to your own is just a question of learning to listen properly. Learning how to decode it.*

However, many other comments are concerned with linguistic explanations for asymmetry, for example:

> *[. . .] For portuguese it's easier to understand spanish, than it is for spanish to understand portuguese. Because spanish don't understand the different endings in portuguese, while for a portuguese the endings sound easier in spanish.*

18 https://www.youtube.com/watch?v=E042GHlUgoQ&t=15s

https://doi.org/10.1515/9783111134697-006

> *danish just needs to take out the potato of their mouths so us norwegians and swedish can understand them.*

> *I, as a Slovenian, can mostly understand Croatian, Serbian and Bosnian without ever learning it, but they have a hard time understanding me. It all comes down to us moving the stressed syllable, which makes us difficult to understand . . .*

> *I as an Egyptian Arabic speaker tend not to understand Iraqis because of the obscure vocabulary they have; but the catch is that they usually have multiple words to refer to things either due to them being synonyms or having different senses for each word whereas we have one: English:* chair, *Egyptian:* korsi,

> *I'm Russian and we have two words for cat, koshka and miska. The previous one is the "official" word while miska is sometimes used in some circumstances, in Belarusian miska is the official word and I don't know if they also use koshka but if they don't Russians can understand Belarusians saying cat but not the other way around. That may be another reason for the thing.*

> *I would add that some communities speak slower and therefore they are understood more easily. This was revealed to me in my native Texas when I happened to go to a meeting and sat next to a visiting Scottish colleague. We had communicated solely through telecon in large meetings. I told him my apologies but I understood less than 50% of what he said and I relied on co-workers to translate his information after the meeting. I asked him if us Texans were as difficult to understand. I laughed and mentioned that "Texans talk real slowly."*

Since early intelligibility studies, researchers have also observed that mutual intelligibility between languages is sometimes asymmetric. Asymmetric intelligibility is fascinating in its own right. However, by understanding the mechanism behind asymmetric intelligibility, we may also advance our understanding of factors that play a role in mutual intelligibility between closely related languages.

As mentioned in Section 1.1, linguists have often assumed that the linguistic prerequisites for mutual intelligibility are symmetric, i.e., that the linguistic barriers that a speaker of language A has to deal with when listening to language B are comparable in magnitude to the barriers that speakers of language B have to deal with when listening to language A. This means that inherent intelligibility between two related languages is assumed to be symmetrical. In the early 1950s, American structuralists included reciprocity in their definition of mutual intelligibility (e.g., Casad 1974: 73). This would mean that any asymmetric intelligibility must be caused by exposure or some other extra-linguistic factor. According to Simons (1979: 81), "discrepancies larger than 10% are due to social factors rather than linguistic factors". Grimes (1992), therefore, suggested using this threshold as a means of identifying such factors and gaining a better understanding of the extra-linguistic determinants of intelligibility. However, he did not rule out that for some language combinations, specific areas of phonology may play a role in explaining asymmetry. Later, researchers found evidence that asymmetries in the different linguistic structures of the languages involved may be part of the explanation for asymmetric intelligibility in addition to extra-linguistic factors. However, most linguistic distance measures quantifying the difference in linguistic structure between languages that were described in Chapter 5 are symmetric. These measures will be inadequate for explaining the intelligibility asymmetries found for some language combinations.

One of the most well-known cases of asymmetric intelligibility is the Swedish-Danish mutual intelligibility. Danish has the reputation of being difficult to understand for Swedish listeners, while Swedish is more readily understood by Danish listeners. For instance, Maurud (1976) reported that Danes understand Swedish reasonably well, while Swedes perform more poorly when having to understand Danish, a tendency which has been reported repeatedly (e.g., Bø 1978; Börestam 1987; Delsing and Lundin Åkesson 2005). In the literature on Danish-Swedish mutual intelligibility, both linguistic and extra-linguistic explanations are proposed for asymmetric intelligibility. Many examples in this chapter are taken from research on Danish-Swedish mutual intelligibility.

Appendix B lists the percentages of correct answers in the cloze test from the MICReLa project and indicates when the intelligibility scores are asymmetrical. Most asymmetries are found in the Romance area, both when looking at the general intelligibility results and when looking at the inherent intelligibility results. In the Germanic area, there are many asymmetries in general intelligibility as well, especially between language pairs involving the school languages German and English. In the Slavic area, there is only a significant asymmetry between Croatian and Slovene.

Throughout this chapter, possible reasons for these asymmetries are discussed. Extra-linguistic explanations are discussed in Section 6.1. Section 6.2. introduces the idea of measuring linguistic variation asymmetrically. It is shown that the amount of linguistic variation between a speaker of variety A and a listener of variety B may not be equal to the amount of linguistic variation between a speaker of variety B and a listener of variety A.

6.1 Extra-linguistic factors causing asymmetric intelligibility

All the extra-linguistic factors discussed in Chapter 4 could be relevant for understanding asymmetric intelligibility. Some of them could explain the asymmetry in mutual intelligibility between individual speakers, while other factors can be generalized to whole languages rather than individuals. In this section, the extra-linguistic factors exposure, attitude, orthographic knowledge and plurilingual resources are discussed.

6.1.1 Asymmetric exposure

At the personal level, the amount of exposure may vary a lot. Still, at the language level, it has often been shown that for various reasons, some groups of speakers have more exposure to a certain related language than vice versa. Speakers of "smaller" languages are more likely to get exposed to languages with more speakers because the chance of hearing a language with many speakers is higher, but often also because larger languages generally have a higher status. A higher status may lead to more use of the language in a wider number of contexts. Exposure is a significant determinant of intelligibility (see Section 4.1), and it may explain asymmetric intelligibility in situations where there is an unequal amount of exposure between the speakers of two languages.

The amount of exposure has often been mentioned as part of the explanation for the asymmetric intelligibility between Swedish and Danish that has consistently been found in various investigations (see introduction to this chapter); Danes generally understand Swedish better than vice versa. Maurud (1976) collected self-estimations of exposure to the neighboring language by asking participants to indicate how often ("never", "seldom", "sometimes", "often") they visited their neighboring countries, listened to radio from neighboring countries, and watched television from neighboring countries. The Danes reported much more contact with Swedish than Swedes with Danish. Generally, Danes have more often listened to Swedish, for instance, through the media and interaction with Swedes, than vice versa. More recent studies have also reported this asymmetry (Jörgensen and Kärrlander 2001). The Swedish population is larger (a little more than 10 million) than the Danish (almost 6 million), and Denmark is a geographically smaller country, so for most Danes the distance to Sweden is small. The great majority of Swedes live far away from the Danish border and get little exposure to Danish, while Danes have more opportunities to get exposed to Swedish. The capital of Denmark, Copenhagen, is close to the Swedish border on the west coast of Sweden. For many Danes, the western part of Sweden is easily reached by crossing the bridge across the Øresund, while for Swedes living around the Swedish capital, Stockholm, on the east coast of Sweden, a day trip to Denmark is not within reach.

From an investigation by Delsing and Lundin Åkesson (2005), it became clear that the geographic vicinity plays an important role in the amount of exposure. They collected exposure scores (max. 100 points) from Danes in Copenhagen close to the Swedish border (30 minutes by car, contact score 21.8) and from Århus, approximately a three-and-a-half hour drive from the border (contact score 14.0) and from Swedes in Malmö just across the bridge to Denmark (contact score 35.7) and from Stockholm, a 7 hour drive to Denmark (contact score 8.3). They found significant correlations[19] between the exposure scores and scores on intelligibility tests. Listeners from Malmö performed better than listeners from Stockholm. However, unexpectedly, listeners from Copenhagen did not perform better than listeners from Århus. The researchers explain this by a lack of interest in the investigation among the listeners from Copenhagen.

In the MICReLa project, the amount of exposure the listeners had to the target language was quantified by having the listeners fill in 6 five-point scales from 1 ("never") to 5 ("every day"), see Section 4.1.1. The mean results can be found in Appendix C. A large part of the asymmetries in intelligibility (see Appendix B) can be explained by asymmetries in previous exposure to the target language. The link between asymmetric intelligibility and asymmetric exposure can be established by correlating the asymmetry scores for both measures. Asymmetric exposure scores are difference scores (also called deltas, Δ) that are calculated by subtracting the mean exposure to language A among speakers of language B (the AB exposure score) from the exposure to language B among speakers of language A (the BA exposure score). In the same way, asymmetric intelligibility scores can be calculated (the AB intelligibility score minus the BA intelligibility score). The exposure deltas can then be correlated with the intelligibility deltas. These deltas are expected to correlate positively: If speakers of two languages have different levels of exposure, they are likely also to have different levels of intelligibility.

The asymmetry scores are presented in the scattergrams in Figure 6.1. The results are arranged so that for each language pair, the lowest intelligibility score is subtracted from the highest score. In other words, when calculating the asymmetries, the highest score is chosen as the AB score and the lowest as the BA score. This results in asymmetric intelligibility scores that are always positive. The corresponding asymmetric exposure scores are mostly also positive. For example, Romanians understand Italian better (AB scores) than the other way around (BA score), and they also have more exposure to Italian. For a few languages, there are slight negative asymmetric exposure scores. However, in general, a high positive asymmetric intelligibility score corresponds to a high positive asymmetric exposure score, and a low positive asymmetric intelligibility score corresponds to a low positive asymmetric exposure score. The correlation is high ($r = .92$, $p < .001$). Even when leaving out the ten language pairs where one of the languages is a school

19 No correlation coefficients were reported.

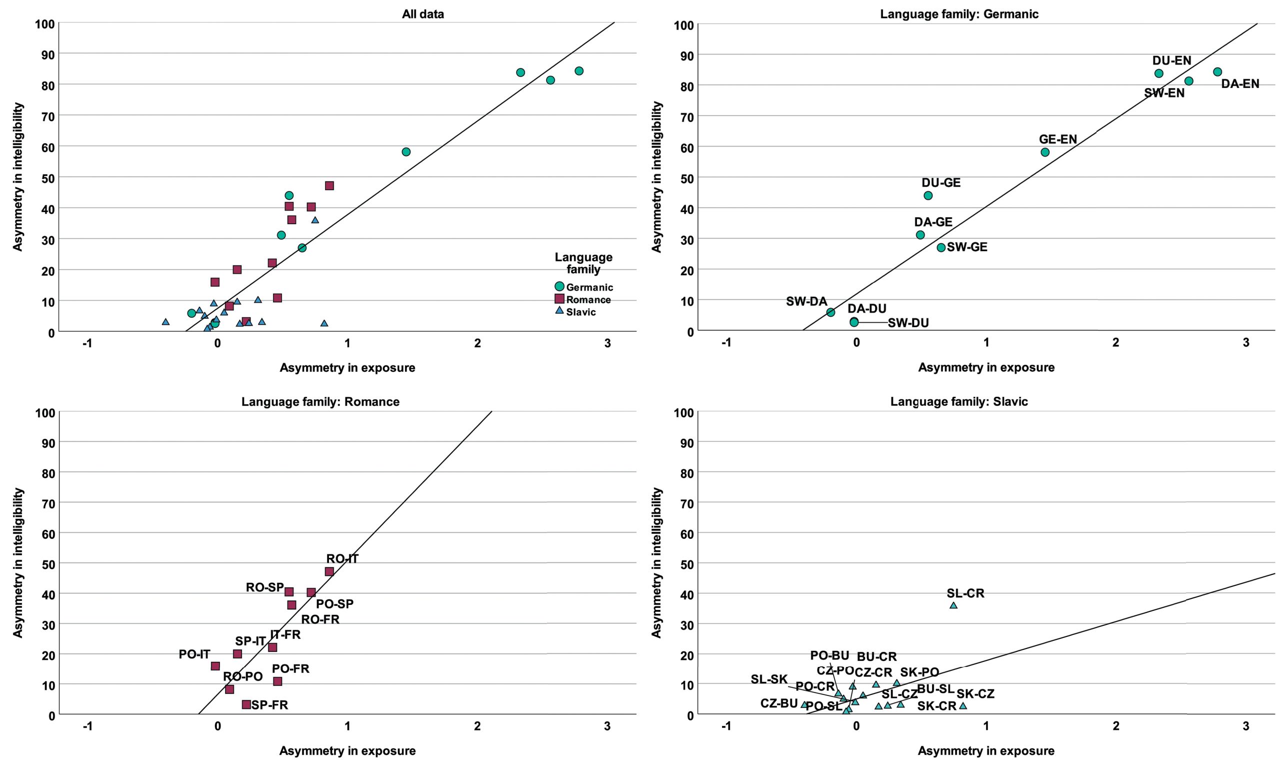

Figure 6.1: Scattergrams showing the relationship between the asymmetric intelligibility scores (AB intelligibility score minus BA intelligibility score) and the asymmetric exposure scores (AB exposure score minus BA exposure score) for 35 language pairs in the MICReLa project, $r = .92$, $p < .001$ for all language combinations, $r = .97$ for Germanic ($p < .001$), $r = .82$ for Romance ($p < .01$), and $r = .49$ for Slavic (n.s.). For explanations, see Figure 4.2.

language (English and German in the Germanic area, French in the Romance area), the correlation is high ($r = .75$, $p < .001$). This shows that asymmetric intelligibility can to a large extent be explained by asymmetric exposure even when it is gained outside a formal classroom situation.

In the Germanic language family (green markers in Figure 6.1), the highest asymmetric intelligibility and exposure scores were found in language combinations involving English. Since English and German are school languages, they attract much better scores as target languages than the other languages. Danes had more exposure to Swedish (mean exposure score 2.2) than Swedes to Danish (mean exposure score 2.0). Contrary to most previous investigations, the mutual intelligibility between Danish and Swedish was not asymmetric when measured using the cloze test.

In the Romance language area (red markers), the highest asymmetric exposure scores were found in combinations with Romanian on the one hand and Italian and Spanish on the other hand. Romanians had a mean exposure score of 2.1 to Italian and 1.8 to Spanish, while the scores were only 1.2 and 1.3 in the other direction. Romanians also understood these two languages better (57.7% and 54.0% correct answers to the cloze test) than the other way around (10.6% and 13.6%). Romanians often encounter Italian and Spanish, for instance, via television, and films are rarely dubbed in Romania. Native speakers of other Romance languages are less often exposed to the Romanian language. The Portuguese-Spanish exposure scores were also asymmetric, with Spanish listeners scoring 1.4 and Portuguese listeners 2.2. This corresponds to asymmetric intelligibility scores where Portuguese listeners score better than Spanish listeners (77.4% vs. 37.2%). Jensen (1989) found results in the same direction but with considerably smaller asymmetry for Latin American Spanish and Brazilian Portuguese. Brazilian-Portuguese listeners understood Latin-American Spanish speakers better with a mean accuracy of 58% correct answers to multiple-choice questions about four recorded text readings. Spanish listeners understood Brazilians less well with 50% correct answers.

There were few asymmetric intelligibility and exposure scores in the Slavic language group (blue markers), and the correlation is insignificant. The highest exposure delta in the Slavic group was found for Czech-Slovak. The Slovaks had more exposure to the neighboring language (mean exposure 3.6) than the Czechs (mean 2.8). The Slovaks performed better on the spoken cloze test (95.0 % vs. 92.7%), but the difference might have been larger if there had not been a ceiling effect. When asked to judge their own comprehension of the test language, Slovaks perceived their comprehension of the neighboring language to be higher (95.8%) than the Czechs (87.3%). Historical circumstances have led to more exposure to Czech and higher Czech comprehension among Slovaks than the other way around (Budovičová 1987; Berger 2003; Nábělková 2007). According to Ethnologue (Eberhard, Simons, and Fennig 2023) there are currently 9,372,000 Czech speakers in the Czech Republic but only 5,115,000 Slovak speakers in Slovakia. Written Czech texts have always played an essential role in Slovakia and their influence persists today in a diminished form. Spoken Czech texts are present in Slovakia through various mediums, such as the media and films. On the other hand, Czech contact with the Slovak language has been much scarcer, and today it is not an everyday experience for the Czech language community to be exposed to the Slovak language, especially for the younger generation. These circumstances are likely to lead to asymmetric intelligibility between the two languages.

In other language areas, asymmetric exposure has also been linked to asymmetric intelligibility. Gooskens and Schneider (2019) investigated the mutual intelligibility between related language varieties of northern Pentecost, one of the Pacific islands of Vanuatu (see Figure 6.2). They included four varieties: the Raga language in the north and the three Apma dialects, Suru Kavian, Suru Mwerani, and Suru Rabwanga, further to the south. They showed that the amount of exposure is at least part of the explanation for the asymmetric mutual intelligibility between Suru Kavian (spoken in Namaram) and other language varieties spoken on the island in the villages of Loltong, Waterfall, and Tanbok. In a word intelligibility task, Suru Kavian speakers translated more words from the three other closely related language varieties correctly (mean 90.2%) than vice versa (mean 44.1%, see the vertical axis in Figure 4.3). They also had more exposure to the other varieties, as it became clear from the higher number of non-cognates that they translated correctly (mean 85%, see the horizontal axis in Figure 4.3) than vice versa (mean 28.3%). Suru Kavian is spoken in the center of the area, and the speakers have more visitors from outside the village than the other villages do. They are, therefore, more often exposed to the other languages on the island. In addition, Suru Kavian is spoken by only a small number of speakers. For this reason, Suru Kavian speakers need to be able to understand other languages for communicative purposes. Suru Kavian speakers are locally renowned for their tendency to abandon their own variety when conversing with speakers of other languages on the island. Suru Kavian children are often taught in another variety than their own in school because teachers do not speak Suru Kavian. As a result, the number of speakers of Suru Kavian is decreasing, and the variety is on the verge of language shift. This investigation demonstrates that asymmetric exposure may result in asymmetric intelligibility, which in turn may result in language shift.

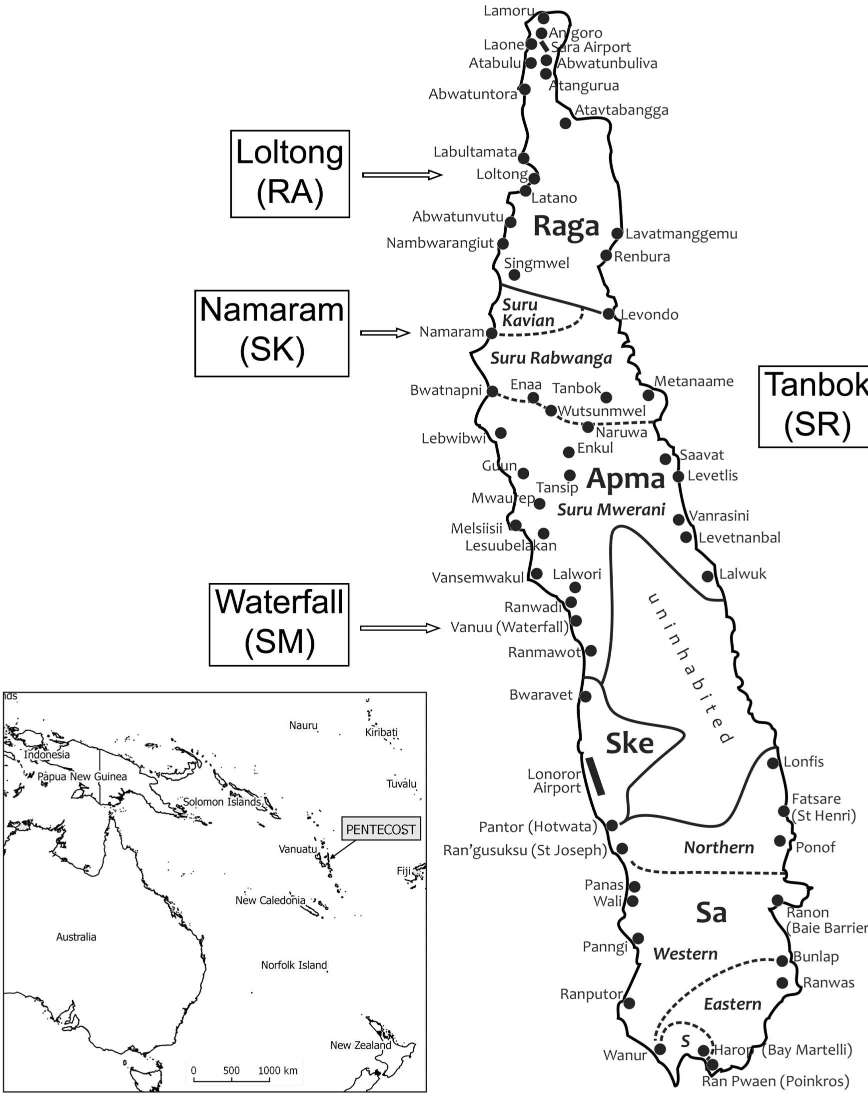

Figure 6.2: Language varieties of Pentecost, one of the islands of Vanuatu. Purported language boundaries are drawn in solid lines; purported dialect boundaries are drawn in dotted lines. For abbreviations, see Figure 4.3. Source: Schneider and Gooskens (2018: 146).

6.1.2 Asymmetric attitudes

Wolff (1959) was perhaps the first to attribute asymmetric intelligibility to the attitudes of speakers toward each other's communities, stemming from the hierarchical relations between the groups. He investigated mutual intelligibility between Kalabari and Nembe, two closely related Nigerian Ijo languages spoken in the Eastern Niger Delta. His findings revealed that Nembe speakers reported high linguistic similarity between their own language and the Kalabari language and expressed an ability to understand Kalabari. On the other hand, speakers of Kalabari judged Nembe to be linguistically distant and unintelligible to them. Wolff suggested that this asymmetry in intelligibility should be linked to an asymmetry in language attitudes. When his study was conducted, the Kalabari were arguably the most prosperous group in the Eastern Niger Delta, and they regarded other Ijo-speaking groups as inferior to them. Therefore, Nembe speakers were more positive towards Kalibari speakers and more willing to make an effort to understand their language than the other way around.

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. Consequently, if speakers of two languages have different attitudes towards each other's languages, this may result in asymmetric intelligibility. Still, the relationship between language attitudes and intelligibility is rather elusive (see Section 4.3) and has only been mentioned sporadically as an explanation for asymmetric intelligibility. Swedish and Danish form an exception. It has repeatedly been suggested that the asymmetric intelligibility between Swedes and Danes can be traced back to asymmetric attitudes. Swedes may have a less positive attitude toward the Danish language, culture, and people than Danes do towards the Swedish language, culture, and people. Sweden has been referred to as the "big brother" in Danish public opinion (Sletten et al. 2004; Thorvaldsson 2011). This draws on the fact that Sweden has historically been more influential than the other Scandinavian countries, and this may be a reason for Danes to be more positive about Swedish than Swedes about Danish and for Danes to make a greater effort to understand Swedish than for Swedes to understand Danes.

Haugen (1966) elicited language attitudes in a survey on mutual intelligibility in Scandinavia by means of the question "How do you like the sound of Y?" with the response alternatives "Compared with X (the informant's own language), Y (the target language) is: more beautiful, equally beautiful, less beautiful." He found that 42% of the 81 Danish informants thought that Swedish sounded more beautiful than their own language, while none of the 54 Swedish informants thought that Danish sounded more beautiful than Swedish. Jörgensen and Kärrlander (2001) as well as Lundin and Christensen (2001) studied the mutual intelligibility of Danish and Swedish in the Øresund region, which is the border area between Sweden and Denmark connected by the Øresund Bridge. The results showed that Danes were better at understanding Swedish than Swedes were at understanding Danish. Since the Danish listeners held more positive attitudes towards Swedish than vice versa, the researchers suggested that attitudes could be part of the explanation for the asymmetric Danish-Swedish intelligibility results. Unfortunately, they did not report correlations between attitudes and intelligibility. The same conclusions have been drawn by Delsing and Lundin Åkesson (2005), who tested the intelligibility of the three Scandinavian languages in a large number of places in the Nordic countries and asked the listeners to state their attitude towards the other languages on 5-point scales. The results showed that Danish participants considered the Swedish language more beautiful than Swedish participants considered the Danish language. They reported significant correlations between attitudes and intelligibility but did not report any correlation coefficients.

In the investigations by Haugen and by Delsing and Lundin Åkesson, listeners did not base their judgments on recordings of the target languages. Instead, they were asked to give their opinions based on their knowledge about them. The assumption here is that Danes and Swedes are likely to have sufficient experience with each other's languages to be able to form opinions about them (see Section 4.3). However, it is uncertain whether this is the case for all listeners. Schüppert, Hilton, and Gooskens (2015) carried out a matched-guise experiment (see Section 4.3.1) employing recordings of a balanced bilingual speaker proficient in both Danish and Swedish. Participants consisted of groups of Danish and Swedish children aged between 7 and 16 years, who evaluated recordings in both Swedish and Danish, along with four other languages, using 5-point semantic differential scales indicating how normal, beautiful, smart, modern, kind, and rich the speakers sounded to them (see Section 4.3.1). According to the findings, the bilingual speaker received more favorable ratings from Danish participants when speaking Swedish compared to the ratings from Swedish participants when she spoke Danish, across all seven scales. The investigation with recordings by Schüppert, Hilton, and Gooskens (2015) thus confirms the results of previous Danish-Swedish attitude investigations where the participants did not base their opinions on recordings. Since the matched-guise technique was applied, we know for sure that differences in speaker characteristics did not cause the asymmetric attitudes. The listeners heard the same text in both languages so the asymmetry cannot be due to the content of the recording, either. It was therefore assumed that the asymmetric attitudes found should be explained by imposed norms and social connotations, thus confirming the big brother hypothesis mentioned above. However, as explained in Section 4.3, there is also evidence that some languages have specific linguistic characteristics that make them sound more aesthetically attractive than others. Swedish is a tonal language resulting in a higher degree of variation in pitch contours than in Danish, with its rather monotonous intonation. This may make Swedish sound more attractive than Danish.

It is possible to calculate attitude asymmetry (or deltas, Δ) for the MICReLa-project in the same way as the exposure asymmetry in Section 6.1.1 (AB minus BA scores) and correlate them with the asymmetric intelligibility scores. In general, the language combinations with positive attitude deltas should also have positive intelligibility deltas. However, this is not always the case. For instance, Swedes understand German better than vice versa ($\Delta > 0$), but they have a more negative attitude towards German ($\Delta < 0$). The correlations are lower than for the exposure deltas but still significant ($r = .53$, $p < .001$) when including all languages. However, when examining each language family separately, the correlation is significant only in the Germanic language family. The corresponding scattergrams are presented in Figure 6.3. These results show that asymmetric intelligibility may, at least for some groups, be linked to asymmetric attitudes as suggested in previous investigations.

6.1.3 Asymmetric orthographic knowledge

As mentioned in Section 4.4, orthographic differences between Danish and Swedish may also cause asymmetric intelligibility, even in spoken language. By calculating entropies (a measurement of uncertainty, see Section 6.2.2) between phoneme-to-grapheme and grapheme-to-phoneme mappings for 16 European languages, Schüppert et al. (2017) showed that Swedish orthography is more transparent than Danish orthography. This can be explained by the fact that the Danish orthography is very conservative, while spoken Danish has changed faster than most other Germanic languages, including Swedish (Brink and Lund 1975; Grønnum 1998; Elbro 2005). Due to the large number of changes (especially lenition processes, i.e., the weakening of consonants) in the phonology of Danish, the letter-to-sound correspondences that were rather straightforward around a century ago have become more opaque during the twentieth century. Many sounds lost in pronunciation are still preserved in the orthography. In Swedish, the corresponding sounds are still preserved both in the orthography and in the pronunciation.

According to Teleman (1987: 76), the many changes in the Danish pronunciation may have made it more difficult for a Swede listening to Danish to “find the letters behind the sounds” than vice versa. Danes can often understand spoken Swedish words due to their similarity to written Danish, while Swedes cannot draw on written Swedish to the same degree when understanding spoken Danish. For example, Danes presented with the spoken Swedish word *tolv* /tɔlv/ ‘twelve’ can probably use their orthographic knowledge to match this word to the written cog-

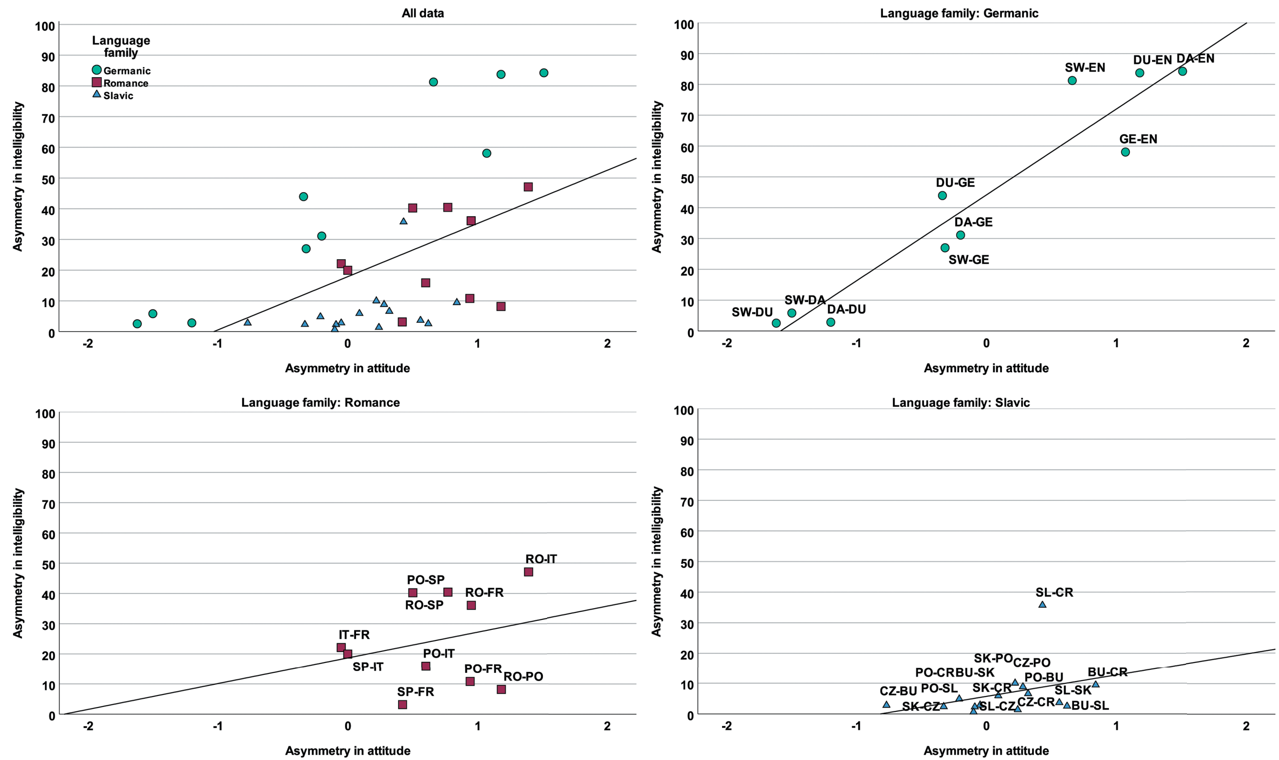

Figure 6.3: Scattergram showing the relationship between the asymmetric intelligibility scores (AB intelligibility score minus BA intelligibility score) and the asymmetric attitude scores (AB attitude score minus BA attitude score) for 35 language pairs in the MICReLa project, $r = .53$, $p < .001$ for all language combinations, $r = .95$ ($p < .001$) for Germanic, $r = .26$ for Romance (n.s.), and $r = .34$ (n.s.) for Slavic. For explanations, see Figure 4.2.

nate in their native language *tolv,* even though the Danish pronunciation is different /tɔlˀ/. On the other hand, Swedes listening to the spoken Danish word do not get help from their native orthography. The phoneme /v/ is present in Swedish pronunciation as well as in the orthography but absent in the Danish pronunciation.

Gooskens and Doetjes (2009) showed that the proportion of Swedish words that Danes can potentially understand by means of the orthography of the corresponding Danish cognates is larger than the proportion of Danish words that Swedes recognize using their own orthography. Table 5.3 in Section 5.2.1 shows how phonetic distances corrected for orthography were calculated using the Levenshtein algorithm between the Danish and Swedish words for 'pan' (*pande* and *panna*). The distance is calculated by assigning no weight to phonemes that could be understood by means of the native orthography. For Danes, the phonetic distance with corrections for orthography is smaller (25%) than that with the uncorrected Levenshtein distance (50%). This is because Danes hearing this word spoken in Swedish can recognize the initial /a/ in the word (pronounced as [a]) as one of the two possible realizations of written Danish <a> (the other realization being [æ]). Conversely, the corrected calculation has no consequences for the distance calculation for this word pair when taking the situation of Swedish listeners as a starting point (see Table 6.1). Swedish has only one way of pronouncing the <a>, and therefore, a Swede may have difficulties recognizing the Danish [æ] as a possible pronunciation of written <a>. The corrected distance is thus greater for a Swede (50%) than for a Dane (25%), which indicates that it is probably more difficult for a Swede to understand the word in the neighboring language than it is for a Dane.

Table 6.1: Phonetic distances corrected for orthography for a Dane presented with a Swedish cognate (25%) and for a Swede presented with a Danish cognate (50%). Source: Gooskens and Doetjes (2009: 214).

	Phonetic distance corrected for Danish orthography		**Phonetic distance corrected for Swedish orthography**
Danish listener:	p a/æ n ə	Swedish listener:	p a n a
Swedish target:	p a n a	Danish target:	p æ n ə
Differences:	0 0 0 1	Differences:	0 1 0 1
Distance	1/4 = 25%	Distance	2/4 = 50%

The distance representing a Swede's comprehension of Danish, as determined by phonetic distances corrected for orthography by Gooskens and Doetjes (2009), is greater, averaging 46% across all 86 cognates used for the measurements, compared to the distance representing a Dane's comprehension of Swedish, which is

30% on average. This shows that, potentially, Danes can make better use of their native orthography than Swedes when understanding the neighboring language. This argument is further supported by the fact that the corrected distances show higher correlations with the results of a word translation experiment with the 86 Swedish words among a group of Danish listeners ($r = -.63$) than the uncorrected distances ($r = -.54$). It is also supported by the finding by Schüppert and Gooskens (2012) that the intelligibility is symmetric among illiterate children, since they do not have support from the orthography when listening to the neighboring language.

Summarizing the Swedish-Danish results, we have seen that the orthographic transparency of Danish is lower than the Swedish transparency (Schüppert et al. 2017, 2022). Moreover, we know that literate Danish speakers show activation of their native orthography during word recognition tasks for spoken Swedish (see Section 4.4.1). It can, therefore, be assumed that the potential advantage that Danes have from their orthography, as demonstrated by Gooskens and Doetjes (2009) is also actually used. The asymmetric intelligibility between Swedish and Danish may be partly explained by the fact that when listening to Swedish, Danish listeners appear to use the additional information provided by the native orthography more than Swedes do when listening to Danish. This is supported by the fact that in contrast to adult Danes, Danish children did not perform better in a word-recognition task than their Swedish peers (Schüppert and Gooskens 2012).

Orthographic characteristics may also explain asymmetric mutual intelligibility between other language pairs. Schüppert et al. (2017) showed an asymmetry between phoneme-to-grapheme and grapheme-to-phoneme mappings for Spanish and Portuguese, which is comparable to the Danish-Swedish asymmetry. Spanish orthography is more transparent than Portuguese orthography. Previous research has also found asymmetric intelligibility in the same direction between these two languages (Jensen 1989 and Appendix B). Phonological lenition processes (such as consonant deletion and vowel apocope) have altered some Portuguese words to an extent that they have become unrecognizable to native speakers of other Romance languages. Portuguese listeners, however, may easily recognize the non-lenited forms in the neighboring language. To illustrate, Portuguese listeners can be expected to easily understand the spoken Spanish word for 'river' *rio* [ˈri.o] because they can recognize it from the spelling of the Portuguese equivalent *rio.* It will be much more difficult for a Spanish person to understand the spoken Portuguese word because the Portuguese pronunciation [ˈʁi.u] is rather different from the Spanish pronunciation and orthography (see Table 6.2).

Table 6.2: Example where Spanish has a stronger letter-to-sound correspondence than Portuguese.

	Portuguese	Spanish	English
Spelling	<rio>	<rio>	‘river’
Pronunciation	['ʁi.u]	['ri.o]	

6.1.4 Asymmetric plurilingual resources

As explained in Section 4.2, when people know a language other than their native language, this can sometimes be used as a “bridge language” that can help them understand a language that they have not learned before. They may understand non-cognates in the target language because they know them from this bridge language. Sometimes, knowledge of such bridge languages can result in asymmetric intelligibility between speakers of related languages.

In the Romance part of the MICReLa project, where the mutual inherent intelligibility between five Romance languages was tested, asymmetric intelligibility was found between all language combinations involving Romanian as either the target language or as the listener language. None of the Romanian listeners had more than marginal previous exposure to the target language, but still, they got higher scores when tested in Italian, Spanish, and Portuguese (inherent intelligibility scores between 20.7% and 47.2%) than Italian, Spanish, French, and Portuguese listeners who were tested in Romanian (scores between 8.7% and 14.0%). As explained in Section 6.1.1, Romanians often encounter other Romance languages, while this is less the case for native speakers of the other four languages. Therefore, even when the Romanian listeners had no exposure to the target language, they may have used their knowledge about the vocabulary of other related Romance languages to understand the target language. Unfortunately, since the listeners were not asked about their exposure to these other related languages, this explanation could not be confirmed statistically. Speakers of other Romance languages are less able to use their knowledge of related languages when listening to Romanian because the Romanian language has a very deviant vocabulary with many non-Romance words.

Past studies on inter-Scandinavian intelligibility have shown that, typically, Norwegians exhibit greater proficiency in understanding the closely related languages Danish and Swedish compared to Danes and Swedes do in understanding Norwegian. Various explanations for this asymmetric intelligibility in favor of Norwegian listeners are found in the literature. This includes higher motivation to understand the target language or more familiarity with the neighboring languages, but the high degree of experience with language variation within their own country has been especially stressed as an explanation for the superiority of the Norwegians (Börestam Uhlmann 2005). For historical reasons, dialects have a strong position in Norway, and many Norwegians are familiar with and often speak more than one variety of Norwegian. Additionally, there are two official forms of written Norwegian, Bokmål (literally 'book tongue') and Nynorsk ('new Norwegian'), each with its own variants. Conversely, the position of dialects is exceptionally weak in Denmark, and fewer people speak and hear dialects regularly (Pedersen 2003). Swedish dialects take a position between these two extremes. Both Danish and Swedish have only one standard written form. These circumstances may give Norwegians an advantage when it comes to inter-Scandinavian intelligibility. As people are more accustomed to dialectal variation in Norway, they are assumed to have a better-developed language awareness (Torp 1998), which makes them better at understanding the Danish and Swedish standard languages. They can use their linguistic experience with Norwegian dialects to guess the meaning of cognates in other related languages, in this case, Danish and Swedish (inferencing strategies, see Section 4.2). Compared with Norwegians, Danes and Swedes, in general, less often encounter their own dialects, especially if they live in the capital areas. It is often assumed that this lack of experience with language variation may be part of the explanation for their poorer understanding of the neighboring languages (Börestam Uhlmann 2005).

If it is the case that Norwegians have more language awareness, they could be expected to be better than Danes at understanding varieties with the same phonetic distance to their native variety. Gooskens and Heeringa (2014) set up an investigation to get an answer to the question of whether Norwegians are indeed better at understanding Nordic varieties (Danish, Faroese, Icelandic, Norwegian, and Swedish language varieties, see Figure 6.4) than Danes when influence of the phonetic distances between their own and the neighboring language are neutralized. The results of a sentence translation task involving 18 language varieties showed that Norwegians (from the capital of Oslo) are generally better than Danes (from Lyngby, close to Copenhagen) at understanding Nordic language varieties. While the Danish listeners translated only 39.0% of the words in the 18 language varieties correctly, the Norwegian listeners had 67.8% correct translations. At first glance, this overall result confirms the impression of Scandinavian scholars that Norwegians are better at understanding the neighboring languages. However, a closer look at the relationship between the intelligibility scores and the phonetic distances shows that the Norwegians were not better at understanding the 18 varieties *relative* to the phonetic distances between their own variety and the test varieties than the Danes (see Figure 6.5). If Norwegians indeed have higher language awareness than Danes, they can be expected to at-

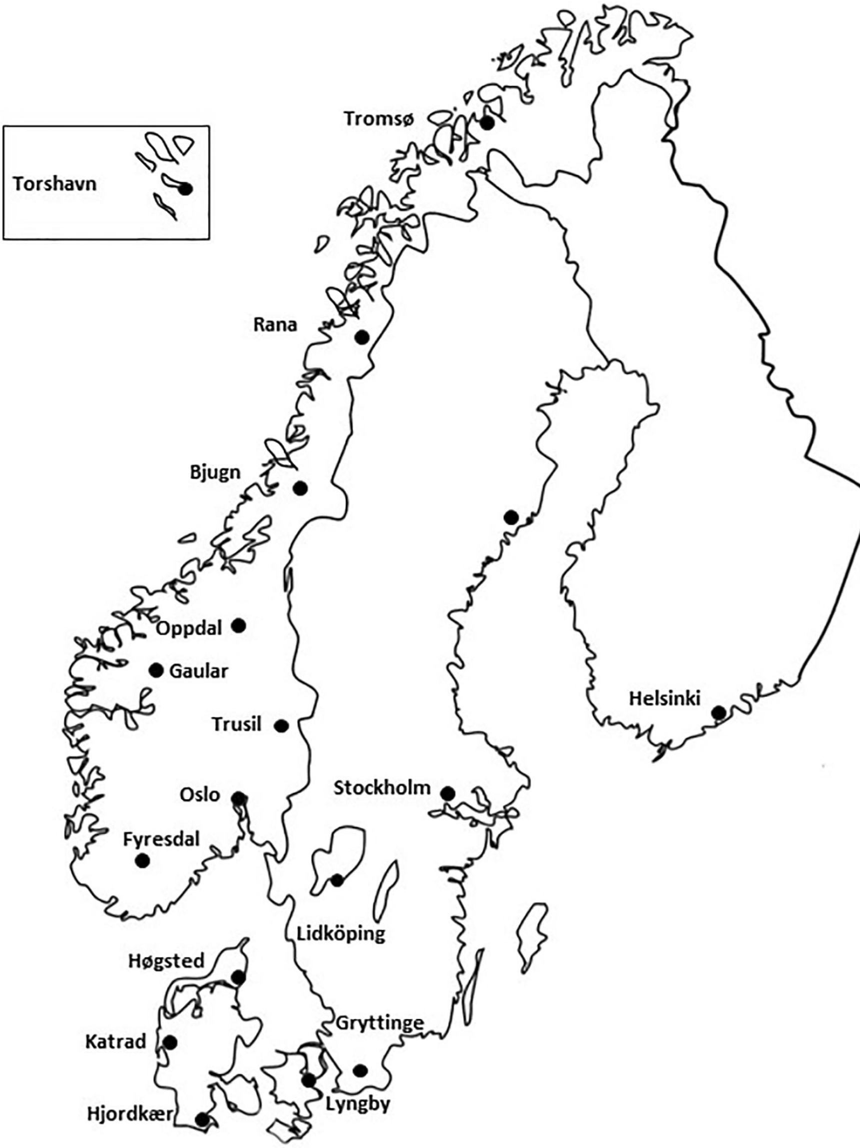

Figure 6.4: The 18 Nordic language varieties in the investigation by Gooskens and Heeringa (2014: 254).

tain higher intelligibility scores in relation to the phonetic distance. This would mean that the Norwegian regression line (blue line in Figure 6.5) would generally be situated to the right of the Danish regression line (red line). However, this is not the case, and it therefore looks as if the Danes are just as good at breaking the phonetic code as the Norwegians.[20]

The higher intelligibility scores of Norwegian listeners found in previous investigations may be due to the fact that these investigations focused on the intelligibility of the standard languages. Norwegians might have more experience with the

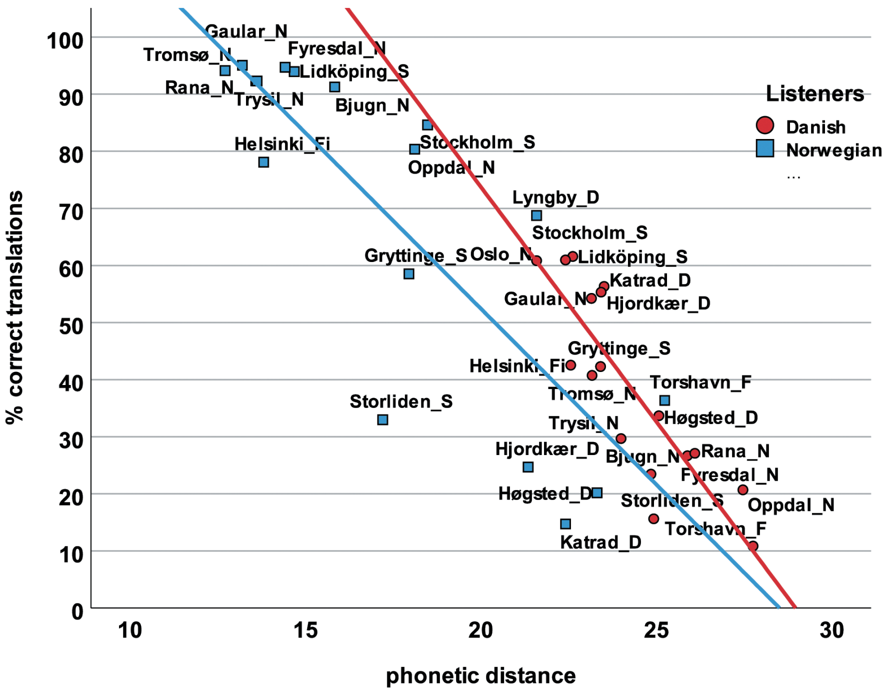

Figure 6.5: Scattergram showing the relationship between phonetic distances to the standard language of the listeners and intelligibility among Danish (red circles, $r = -.86$, $p < .01$) and Norwegian listeners (blue circles, $r = -.82$, $p < .01$). After each dialect is indicated in which country the dialect is spoken (D = Denmark, N = Norway, S = Sweden, F = Faroe Islands, Fi = Finland). Adapted from Gooskens and Heeringa (2014: 262).

20 To test whether the scores for the two listener groups were significantly different, the Norwegian residuals were compared to those of the Danes. The residuals represent the variation in the intelligibility scores that cannot be explained by phonetic distance but is determined by other factors. The results of an independent samples t-test showed that the difference between the Norwegian and the Danish residuals was not significant ($t = 1.353$, $df = 32$, $p = .185$ for all results, $t = 1.621$, $df = 22$, $p = .119$ when excluding results from the country of the listeners and $t = 1.596$, $df = 9$, $p = .149$ when including only results from the country of the listeners).

standard languages than the Danes. Danish movies and television series are popular in Norway, while Danes do not watch Norwegian programs to the same extent. Norwegian has also sometimes been referred to as “the language in the middle”. Results of phonetic (Levenshtein) distance measurements between two Norwegian, two Danish, two Swedish, and two Finland Swedish varieties showed that the mean distance from the Norwegian varieties to all other Scandinavian varieties was smaller than the distances from the Swedish and the Danish varieties (Gooskens 2007). Figure 6.5 shows a similar picture. The distances from “Standard Danish” (represented by the Lyngby variety close to Copenhagen) to the other 17 Scandinavian dialects are all above 20% (see the red circles in Figure 6.5). The distances from “Standard Norwegian” (represented by the Oslo variety) are often lower (blue circles). These smaller phonetic distances probably make it easier for Norwegians to understand Scandinavian varieties than for Danes and Swedes. It is also possible that the Norwegian listeners in the investigation by Gooskens and Heeringa (2014) had less language awareness than Norwegians in general because they were all from Oslo, the capital of Norway. This group of listeners may be less exposed to a large variety of Norwegian dialects than Norwegians from other parts of the country and, therefore, have less language awareness. Recent research has shown that even in Norway, dialects are losing ground and that the Oslo variety is becoming increasingly dominant (Røyneland 2009; Hilton 2010). To test this hypothesis, the intelligibility tests should be repeated with groups of Norwegians with more dialect exposure.

6.2 Linguistic factors causing asymmetric intelligibility

As mentioned in the introduction to this chapter, there is evidence that in some language pairs, asymmetric intelligibility has its origin in linguistic factors. Simons (1979) already noted this, and since then, experimental evidence has been amassed by various researchers. In this section, a number of these investigations are discussed.

As explained in the introduction to Chapter 5, ideally, the influence of extra-linguistic factors must be excluded to investigate the role of linguistic factors in explaining intelligibility. Therefore, asymmetric intelligibility results from the MICReLa project presented in this section are based on measurements of inherent intelligibility. Some of the other investigations also took precautions to exclude the influence of extra-linguistic factors as much as possible.

As noted by Abunasser (2015), distances cannot be asymmetric in a strictly mathematical sense. The distance from language A to language B equals the distance from language B to language A. However, to explain intelligibility, it seems reasonable to talk about asymmetric distances in a non-mathematical sense. It can be compared to geographic distances. Mathematically, the distance from place A to place B is equal to the distance from place B to place A. However, in real life, the effort that it takes to travel distances is often asymmetric due to one-way roads, going uphill or downhill, different public transportation schedules, congestion, etc. In this chapter, asymmetric distances between languages are referred to in this non-mathematical sense.

6.2.1 Lexical asymmetry

The fact that the lexical distance from language A to language B can be different from the distance from language B to language A can be illustrated with an example. In Table 6.3 (from Golubović 2016), words in five Slavic languages have been aligned to the Croatian word for ‘woman’ (*žena*). In Slovene, Bulgarian, Czech, and Slovak the words are perfect counterparts, as they are similar in form and share the same meaning. In Polish, the word *żona* means ‘wife’, but it could still be included because it points the listener in the right direction.

Table 6.3: The word for ‘woman’ in the target language (Croatian) and the corresponding cognates in five Slavic languages. Adapted from Golubović 2016: 116.

Target language	Language of listener				
Croatian	Slovene	Bulgarian	Czech	Slovak	Polish
žena /ˈʒěna/	ženska /ˈʒɛ̀ːnska/	Жена /ʒěˈna/	žena /ˈʒɛna/	žena /ˈʒena/	żona /ˈʒɔ̃na/

Table 6.4: The word for ‘woman’ in the target language (Polish). There is no corresponding cognate in the five Slavic languages. Adapted from Golubović 2016: 116.

Target language	Language of listener				
Polish kobieta	Croatian –	Slovene –	Bulgarian –	Czech –	Slovak –

Table 6.4 illustrates the same procedure, only this time with Polish as the target language. The Polish word for ‘woman’ is *kobieta,* and it does not have a cognate in any of the other five languages.

This example shows that while a word in language A (Croatian *žena*) may have a cognate with a similar meaning in language B (Polish *żona*), the word for the same concept in language B (Polish *kobieta*) may not necessarily have a cognate (with a similar meaning) in language A. This would be the case if language B uses two words for a concept (i.e., synonyms) covered with only one word in language A. By comparing the two examples in Tables 6.3 and 6.4, it is clear that a Polish speaker reading or listening to the word *žena* in Croatian has a general idea of a female person, while a Croatian speaker who has not previously been exposed to Polish would probably understand the word *żona* but would have no clue what *kobieta* might mean (see Figure 6.6). With more 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 than vice versa, which in turn might yield asymmetric intelligibility.

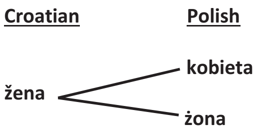

Figure 6.6: Example of language A (Croatian) that has one word žena that corresponds to two different words in language B (Polish) kobieta and żona.

When calculating the lexical distances in the MICReLa project, the words in the texts used for the cloze test in each of the target languages were translated to the corresponding cognates in the languages of the listeners if such cognates existed, in the same way as illustrated in Table 5.1. The lexical distances were expressed as the percentages of non-cognates for each combination of stimulus text in the target language and the corresponding translations in the languages of the listeners. The full matrices of distances can be found in Appendix D (see also Heeringa, Gooskens, and van Heuven 2023). The way the distances were calculated sometimes resulted in different distances from language A to language B and from language B to language A. For example, the lexical distance from Spanish (listener language) to Italian (target language) is 10.4% but larger (14.4%) in the other direction.

To investigate whether asymmetric lexical distances can predict asymmetric inherent intelligibility, the asymmetric inherent intelligibility scores were correlated with the lexical asymmetries (the lexical distances from listener language A to target language B minus the lexical distances from B to A) in the same way as for the exposure and attitude scores in Figures 6.1 and 6.3. The results presented in Figure 6.7 show that lexical asymmetries are weak predictors of asymmetric

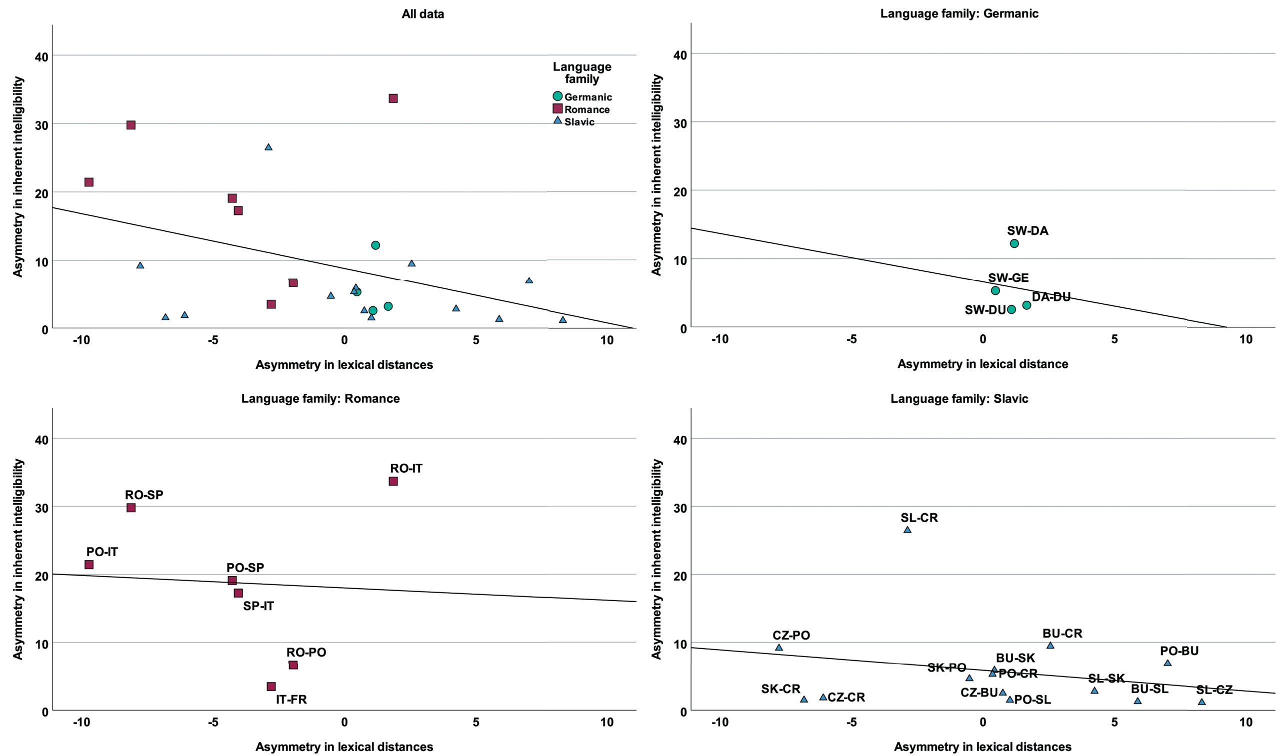

Figure 6.7: Scattergram showing the relationship between the asymmetric inherent intelligibility scores (AB intelligibility score minus BA intelligibility score) and the asymmetric lexical distances (the AB exposure score minus the BA exposure score) for 35 language pairs in the MICReLa project, $r = -.42$, $p < .05$ for all language combinations, $r = -.08$ for Germanic, $r = -.33$ for Romance, and $r = -.23$ for Slavic (all n.s.). For explanations, see Figure 4.2.

inherent intelligibility. In fact, for many languages, there even seems to be a tendency for a positive asymmetric intelligibility to be linked to negative asymmetric lexical distance combinations (the combinations on the left side of the figure). For example, the distance from Spanish to Portuguese is 1.8%, and 6.0% from Portuguese to Spanish. This would predict that Spanish is more difficult for Portuguese speakers to understand than the other way around, but this is not what was found. Portuguese listeners had 54.8% correct answers in the spoken cloze test (see Section 3.2) when tested in Spanish, while the Spanish listeners had an inherent intelligibility score of 35.7%. The weak link between inherent asymmetric intelligibility and asymmetric lexical distances is probably caused by the fact that most asymmetric inherent intelligibility scores are quite low, i.e., a bottom effect, which reduces variability and causes a restricted range effect, resulting in a weakened correlation coefficient. In addition, other linguistic asymmetries may play a more important role, and finally, it is possible that extra-linguistic factors still play a role even if an attempt was made to only include results from participants who had a minimum of previous exposure to the target language (inherent intelligibility).

6.2.2 Phonetic asymmetry

In Section 5.2, it was shown that phonetic distances measured with the Levenshtein algorithm correlate significantly with intelligibility scores. A number of potential improvements and refinements of the algorithm were discussed. Such refinements may not be universal but differ depending on the language combination. For instance, the importance of vowels compared to consonants is likely to depend on the phoneme inventories of the languages involved. Some phonetic characteristics of languages may lead to asymmetric intelligibility between language pairs. This has been argued by a number of researchers. Grimes (1992: 26) assumed that the origin of asymmetric intelligibility between Spanish and Portuguese, as well as between Chinese dialects, can be attributed to certain aspects of phonology. In several more recent investigations, additional evidence has been presented for phonetic differences causing asymmetry.

Gooskens, van Bezooijen, and van Heuven (2015) presented highly frequent Dutch and German cognate nouns to Dutch and German children between 9 and 12 years old. The words were recorded by a balanced bilingual speaker to avoid the influence of different speaker characteristics (see Section 2.1.1.5). The children had no prior knowledge of the target language or any related dialects and held equally neutral attitudes toward the target language. In this way, language exposure and attitudes were ruled out as factors in the study. The results showed that the Dutch children had a notably higher comprehension rate (50.2% correct translations) for the German cognates than the German children had for the Dutch cognates (41.9%). Since the relevant extra-linguistic factors had been excluded, the asymmetry could be attributed to linguistic factors. An examination of the 16 cognate pairs with a mean asymmetric intelligibility score greater than 20% revealed that phonetic detail and asymmetric perception of corresponding sounds played a significant role in explaining the asymmetry. For instance, the Dutch children made only a few mistakes interpreting German *Zeit* /tsait/ 'time' while the Dutch equivalent *tijd* /tɛit/ was very difficult for the German children. 89.3% of the Dutch listeners gave the correct response, while only one German child (2.9%) translated the word correctly. Dutch children had no problem relating German /ts/ to Dutch /t/. They probably considered it an allophone of /t/, or they analyzed it as a consonant cluster and disregarded the /s/. Also, they easily linked German /ai/ to Dutch /ɛi/, presumably because in avant-garde Dutch, /ai/ is used as a new form of standard Dutch /ɛi/, so that /ai/ functions as an allophone of /ɛi/ for the Dutch listeners. On the other hand, the German children interpreted Dutch /ɛi/ as /e/ or /ɛ/, presumably because the onset is more open in German. When hearing the initial /t/ in /tɛit/, the German children did not think of /ts/ but mostly translated into words starting with /d/. Dutch /t/ sounds like /d/ to a German listener because of phonetic differences in the realization of Dutch and German /t/. Dutch voiced plosives are prevoiced (negative voice onset time, VOT), whereas their German counterparts have zero VOT. Conversely, German voiceless plosives are aspirated (long positive VOT), whereas their Dutch counterparts are not.

Härmävaara and Gooskens (2019) carried out a similar investigation with adult Finnish and Estonian adult participants with minimal exposure to the neighboring language. They were tested in a word translation task with both written and spoken words. The mutual intelligibility was symmetric in the written mode, but in the spoken mode, Estonian participants translated significantly more words correctly than the Finnish participants. Since the difference between the words in the written and spoken mode is phonetic, this seems to confirm that phonetic characteristics can cause asymmetric intelligibility between related language pairs. A detailed error analysis of the spoken word pairs with asymmetric intelligibility showed that cognate word pairs with fewer syllables in Estonian than in Finnish were harder to recognize for the Finnish listeners than for the Estonian listeners. An example is the word for 'school', which is *koulu* /kou̯lu/ in Finnish and *kool* /koːl/ in Estonian. The results also showed that Finnish listeners have problems with Estonian sounds that do not exist in Finnish, such as the vowel written as *õ* and pronounced as [ɤ], for example, in the word for 'right' (Estonian *õigus*, Finnish *oikeus*). Another difficult sound for Fins is the palatalized *n* in the word for 'hour' (Estonian *tund* /tunʲd̥/, Finnish *tunti* /tunti/). Finnish lacks the feature palatalization, and therefore, the quality of the nasal is interpreted as a quality of the vowel, resulting in it being perceived as a diphthong (*oi* or *ui*).

As mentioned in Section 5.4.1, Bleses et al. (2008) have shown that Danish children exhibit slower early vocabulary development when compared to children with other native languages, including English and Swedish. Trecca et al. (2021) provide a review of recent work that shows that Danish may be processed differently from other languages, even by adults. In particular, the lenition of consonants and other reduction phenomena, such as schwa assimilation and schwa deletion, would make it difficult to segment words and utterances. This can be illustrated with the sentence *Det er hårdere at årelade* ('it is harder to bleed'). The last three words are pronounced [hɒːːːːːlɛˑð], containing one overlong vowel comprising six syllables, which makes it very difficult to segment. This example may be rather extreme, but there are many examples of this vocalization process in Danish.[21] Bleses et al. proposed that the unique features of the Danish language may also contribute to the asymmetric intelligibility between Danish and Swedish.

Schüppert, Hilton, and Gooskens (2016) found experimental evidence that confirmed this hypothesis. They showed that Danish is spoken significantly faster (in terms of the number of canonical syllables per second) than Norwegian and Swedish, resulting in assimilation and reduction phenomena. As explained in Section 5.4.1, their results suggest that a fast speech rate has a larger impact on intelligibility among non-native listeners than unclear articulation; however, both factors still influence intelligibility. Therefore, it can be concluded that these phonetic characteristics of the Danish language are likely to be part of the explanation for the Danish-Swedish asymmetric intelligibility.

In the MICReLa project, Italian and Portuguese listeners with minimal previous exposure to the target language understood Spanish better than vice versa (see inherent intelligibility scores in Appendix B). Jensen (1989) found the same asymmetry (but smaller) for Latin-American Portuguese and Spanish (see Sections 6.1.1 and 6.1.3). The presence of asymmetry even among listeners with minimal exposure to the language implies that there might be inherent characteristics of the Portuguese language itself that pose challenges for individuals with a different Romance language background. Portuguese is characterized by reduced syllables and a rich vowel inventory compared to Spanish. For example, Spanish has only five monophthongs: /i/, /e/, /a/, /o/ and /u/, while Portuguese has nine oral and five nasal monophthongs. This is likely to render Portuguese a more difficult language for Spanish listeners to understand than the other way around. That Portuguese is also more difficult for Italian listeners than vice versa lends further

21 See https://schwa.dk for more examples.

credibility to the suggestion that the linguistic characteristics of Portuguese make it difficult to understand for other Romance speakers.

The investigations discussed above show that differences in sound inventories between language pairs could be at least part of the explanation for the asymmetry found in intelligibility research. Ideally, phonetic distance measurements that aim to model intelligibility should be able to capture such asymmetries. High correlations were found between intelligibility measurements and Levenshtein distances (see Section 5.2). In its basic form, the Levenshtein algorithm cannot be used to explain asymmetry since distances from language A to language B are the same as the distance from B to A. However, in the same way that lexical differences can cause asymmetry if there are two synonyms for a concept in one language and only one word another, phonetic differences may cause asymmetry (see the example in Figure 6.8). A sound in language A may correspond to two different sounds in language B. For instance, both English /e/ (as in *bet*) and /æ/ (as in *bat*) are typically perceived and reproduced as instances of /ɛ/ by Dutch learners of English (e.g., Collins and Mees 1981; Escudero, Simon, and Mitterer 2012: 285). They will match English *bet* and *bat* onto their own word *bed*, which they pronounce as [bɛt] (with final devoicing). Conversely, a native English listener will not know whether a Dutch speaker of English saying [bɛt] means *bet* or *bat* (or even *bed* or *bad*).

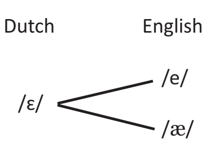

Figure 6.8: Example of phonetic asymmetry. Dutch has one sound /ɛ/ that corresponds to two different sounds in English /e/ and /æ/.

The phonetic complexity scores developed by Cheng (1997, see Section 5.2) for Chinese language varieties result in different scores between AB/BA pairs of languages and predict the corresponding asymmetries in mutual intelligibility. Cheng did not test this prediction experimentally, but Tang and van Heuven (2009) tested mutual intelligibility at the word and sentence level in 15 Chinese languages (see Section 2.2.2.2). In a follow-up analysis (Gooskens and van Heuven 2021), the influence of exposure was excluded to get inherent intelligibility scores by comparing only the nine non-Mandarin (southern) languages in the sample.[22] Correlations between

22 Standard Mandarin (the official language of the People's Republic of China) is known by all listeners, so it serves as a bridge language to all six northern (Mandarin) dialects in the sample.

the asymmetries found in Cheng's complexity scores (the AB score minus the BA score) and the asymmetries found in the word and sentence intelligibility scores by Tang and van Heuven (2009) could be computed in the same way as described for the lexical asymmetries in the MICReLa project (see Section 6.2.1). The correlations were significant (r = .454, p = .003, one-tailed for tests at the word level and r = .331, p = .024, one-tailed at the sentence level). This shows that Cheng's complexity scores reflect the asymmetric intelligibility scores to some degree.

Later investigations have used conditional entropy[23] to account for asymmetric phonetic relations. Conditional entropy measures the complexity of sound mappings and is sensitive to the frequency and regularity of sound correspondences between two languages. Specifically, it measures how predictable a given sound in language A is when mapping it to its corresponding sound in language B. The lower the entropy score, the higher the predictability, and thus, the greater the potential for intelligibility. Therefore, lower entropy scores correspond to higher intelligibility scores. One of the strengths of the entropy measure is that it can be asymmetrical: the conditional entropy between language A and language B may differ from that between language B and language A. This is an advantage compared to the Levenshtein distance, which is symmetrical in its basic form.

In the example in Figure 6.8, an English speaker who encounters an /ɛ/ when listening to a Dutch speaker has two options when trying to find the equivalent phoneme in English cognates. On the other hand, when hearing an /e/ in English, a Dutch speaker can always know that the proper correspondence is /ɛ/. The entropy in this example is, therefore, higher for speakers of English who listen to Dutch than for Dutch speakers who listen to English. Since the algorithm expresses the frequency and regularity of sound correspondences, a substantial amount of data is needed for a reliable measurement. According to Moberg et al. (2007), at least 800 words are required for reliable (stable) entropy measures. More technical details of the algorithm are explained in Frinsel et al. (2015) and Moberg et al. (2007).

To account for the asymmetric intelligibility between Danish, Norwegian, and Swedish, Moberg et al. (2007) calculated conditional entropy in the phoneme mappings for corresponding (cognate) words to express regularity and frequency correspondences in the Scandinavian languages. They measured conditional entropy between Danish, Norwegian, and Swedish based on a corpus of approximately 2200 words per language pair, of which approximately 1900 were cognates with

Listeners outside the Mandarin part of China are expected to have little knowledge of the remaining nine dialects (except their own).

23 A similar measure is word adaptation surprisal (Stenger, Avgustinova, and Marti 2017; Jágrová et al. 2019).

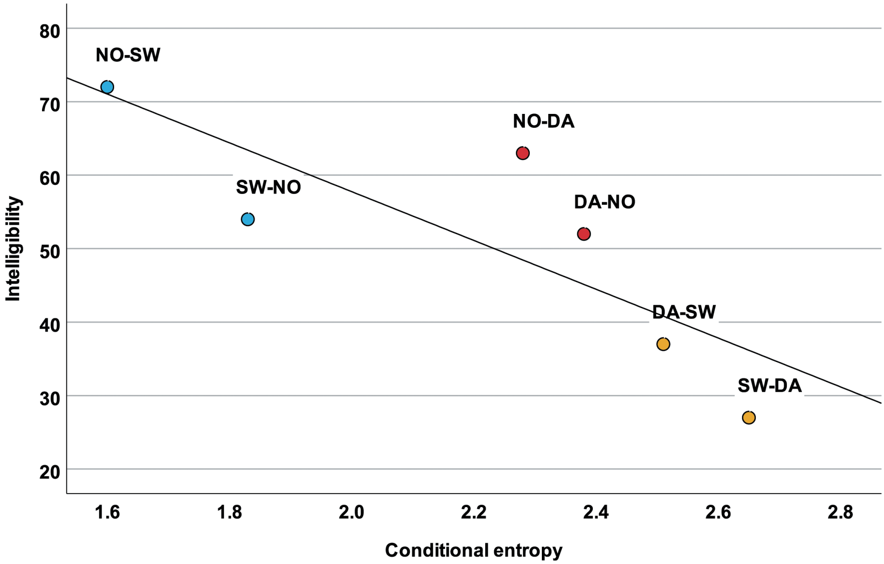

Figure 6.9: Conditional entropies between Danish (Da), Norwegian (No), and Swedish (Sw) in relation to mean percentage correct answers of three spoken intelligibility tests (Maurud 1976; Bø 1978; Delsing and Lundin Åkesson 2005). The first language in each language pair is the language of the listener, the second is the target language. DA = Danish, NO = Norwegian, SW = Swedish. Adapted from Moberg et al. (2007: 14).

each of the neighboring languages. The results are good reflections of the asymmetric intelligibility results found in the literature, $r = -.82$ ($p < .05$). For instance, the entropy was higher for Swedes listening to Danes than the other way around, corresponding to lower intelligibility scores among Swedes listening to Danes than among Danes listening to Swedes (yellow circles in Figure 6.9).

Kyjánek and Haviger (2019) measured entropy between spoken Czech, Slovak and Polish on the basis of almost 7,700 cognate words per language pair. The entropy measurements showed a large asymmetry between Slovaks and Czech, predicting that Slovaks understand Czech better than the other way around. This reflects the general observations that speakers of Slovak understand Czech better than vice versa, which has been a much-debated topic after the breakup of Czechoslovakia (Nábělková 2007). It also confirms the asymmetry found in the MICReLa project for this language combination (see Figure 3.2 and Appendix B). The entropies also predict that Slovaks understand spoken Polish better than vice versa, confirming the MICReLa results as well. Contrary to the MICReLa findings, entropy predicts that Poles understand spoken Czech better than vice versa. However, in the MICReLa results, this asymmetry was small. There is not very extensive sociolinguistic literature on the mutual intelligibility between Czech and Polish and other Slavic language combinations, so it is unclear how well the entropy results reflect observations for these languages.

Conditional entropies reflect asymmetric intelligibility as found in the literature to some degree. Conditional entropy, however, does not measure the actual similarity or distance between two languages. Instead, it measures how predictable (or straightforward) the correspondences are between the members of a particular language pair. A /p/ in one language that always corresponds to a /p/ in another language yields the same amount of entropy as when the /p/ always corresponds to an /f/. However, the latter correspondence may not be obvious for listeners who have never encountered the target language before. Some exposure to the language is necessary for the listeners to discern the patterns in the correspondences. Only then will they get an optimal advantage from low entropy, that is, high regularity. This means that the results of listeners with previous exposure to the target language are expected to show a higher correlation with entropy measures since the listeners may have discovered correspondence rules. On the other hand, listeners with no prior exposure can only recognize words in the target language based on similarities with their native language and on their intuitions of possible sound correspondences.

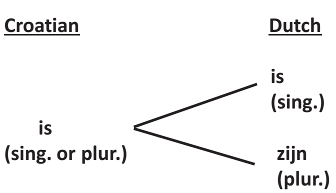

Figure 6.10: Example of language A (Afrikaans) that has one form of the third-person singular and plural of the verb *is* ('to be'), which corresponds to two different forms in language B (Dutch), *is* (third-person singular) and *zijn* (third-person plural).

6.2.3 Morpho-syntactic asymmetry

The investigations summarized in previous sections have demonstrated that asymmetries in lexical and phonetic measures may predict intelligibility asymmetry. In this section, research is presented that shows that morphological or syntactic asymmetries between language pairs may also cause asymmetric intelligibility.

In Gooskens and van Bezooijen (2006), mutual intelligibility between Dutch and Afrikaans was established using written cloze tests based on two newspaper articles. The results suggested that it is easier for Dutch participants to understand written Afrikaans than for South African participants to understand written Dutch. They showed that in addition to a correct interpretation of the lexical meaning (see Section 5.2), a correct interpretation of the grammatical meaning of a word is necessary for a good understanding of a text (see Section 5.3.1). Figure 6.10 shows an example of asymmetric grammatical transparency that may at least partly be part of the explanation for this asymmetry. The Afrikaans verb *is* is familiar to Dutch readers because they know it from their own language as the third person singular present tense of the verb *zijn* 'to be'. They will not have great trouble understanding the meaning of this word, even in cases where it refers to a plural subject. However, the Dutch form *zijn* is unfamiliar to speakers of Afrikaans, and they may have problems recognizing it as the plural of 'to be'.

As explained in Section 5.3.1, the words in the cloze test were given grammatical transparency ratings between 0 and 2. The results showed a clear asymmetry in grammatical transparency (mean 0.05 for Afrikaans participants reading Dutch texts and 0.29 for Dutch participants reading Afrikaans texts). The asymmetric transparency of the grammatical meaning is caused by the fact that many Dutch words have been preserved in Afrikaans but with a more general grammatical meaning. For example, in Dutch, there are two forms for the first person plural of the personal pronoun, namely *wij* 'we' for the subject and *ons* 'us' for the object. Afrikaans only has *ons* for both cases. Afrikaans *ons* will be familiar to Dutch readers because of its paradigmatic relationship with *wij* in their own language, whereas the South African readers will encounter Dutch *wij,* which has no related form in Afrikaans.

Similarly, the analyses of translation errors made by speakers of the two Kurdish language varieties, Kurmanji and Zazaki, when presented with each other's language (Özek, Sağlam, and Gooskens 2021), provide evidence for the role of morphology in explaining asymmetric intelligibility. Zazaki listeners have an advantage in comprehending certain Kurmanji verbs due to the language struc-

Table 6.5: The target word for 'to die' in Zazaki and Kurmaji with asymmetric percentages correct translations. Between brackets, the verb form in Zazaki that helps Zazaki listeners to understand the Kurmaji target word is shown. Source: Özek, Sağlam, and Gooskens (2021).

	Target language		
	Zazaki	Kurmanji	English
infinitive	*merdış*	*mirin*	'to die'
	2.4%	60.3%	
(3rd person plural progressive)	*(merıni)*		

ture of Zazaki, while Kurmanji listeners do not have this advantage when listening to Zazaki. In Zazaki, the infinitives of verbs can be completely different from conjugated forms. As a consequence, the Kurmanji listeners get low scores on these words. On the other hand, the conjugated forms of Zazaki and infinitives of Kurmanji are similar. Therefore Zazaki listeners could infer the meaning of some verbs because the target words were presented in the infinitive. In Table 6.5, an example is provided. The infinitive form of the Zazaki verb 'to die' is *merdış,* while the conjugated form for 3rd person progressive is *merını.* When Zazaki listeners were presented with the infinitive form of the Kurmanji verb *mirin* 'to die', they could understand its meaning (60.3% correct translations) since they associated it with the Zazaki infinitive *merını.* On the other hand, hardly any Kurmanji listeners could understand the Zazaki target word *merdış* (2.4% correct), which is quite different from the Kurmanji equivalent *mirin.*

Härmävaara and Gooskens 2019, see Section 2.2.1.1 and Section 6.2.2) provide examples of word-final elements that Finnish and Estonian listeners confuse with inflectional endings when they hear the words in the neighboring language. The Finnish listeners in the investigation often interpreted the final *-s* in some Estonian words as a marker of the inessive case corresponding to the English preposition 'in' (ending in *-ssa/-ssä* in Finnish, and sometimes shortened to *-s* in spoken language). For instance, the Finnish listeners translated Estonian *kursus* 'course' incorrectly as *korsussa* 'in a dugout' and *kirkossa* 'in a church', and some listeners translated Estonian *kohus* 'court' incorrectly as *kohdussa* 'in uterus', *kohdassa* or *paikassa* 'on a spot' or 'at a place' because they misinterpreted the Estonian *-s* for an inessive case ending. The Finnish listeners also linked the word final *-l* in Estonian and the Finnish adessive case *-lla/-llä* (corresponding to the English preposition 'at'). This ending is often represented in spoken Finnish as *-l.* That can be seen in the incorrect translations of the Estonian stimulus word *nädal* 'week' into *lähellä* 'near' or *täällä* 'here' by many listeners. The Estonian listeners, too, were sometimes guided by the word final elements in their translations (see also Kaivapalu and Martin 2014). For instance, the Finnish nouns ending in *-ma/-mä* that resembles the Estonian infinitive ending *-ma*, e.g., Finnish *elämä* 'life' and *järjestelmä* 'system', were interpreted as verbs in translations like *elama* 'to live' and *järjestama* 'to rank'. However, it seems the Finnish listeners were more often confused, which may be part of the explanation for the Finnish-Estonian asymmetric intelligibility.

Section 5.3.2 showed how the trigram measure can be used to establish syntactic distances. Syntactic distances were measured based on the texts used in the MICReLa project. These distances were measured by manually translating the texts as literally as possible into the other four (or five) languages in the same language family while ensuring that they were still syntactically correct. Therefore, the syntactic distances between the members of a language pair are computed bi-directionally, i.e., from the original text in language A to its literal translation in B, and from the original text in language B into its literal translation in A. This means that like for the lexical distances (see Section 6.2.1), two different distances per language pair are calculated, which makes it possible to express asymmetric relations.

Asymmetries occur when one language allows more syntactical variants than another. As an example, the English sentence:

The house that he has seen

can be translated into Dutch without changing the word order:

Het huis dat hij heeft gezien

However, the Dutch sentence also has an equivalent with the same meaning but a different word order than in English:

Het huis dat hij gezien heeft

Het huis dat hij gezien heeft cannot be translated into English without changing the word order in the final verb cluster *gezien heeft* into *has seen*.

Heeringa et al. (2018) calculated asymmetric syntactic distances for the Germanic languages in the MICReLa project on the basis of 66 sentences in four cloze tests (see Appendix A). The asymmetries that they found are presented in Figure 6.11. They predict that it is easier for native speakers of Dutch to understand

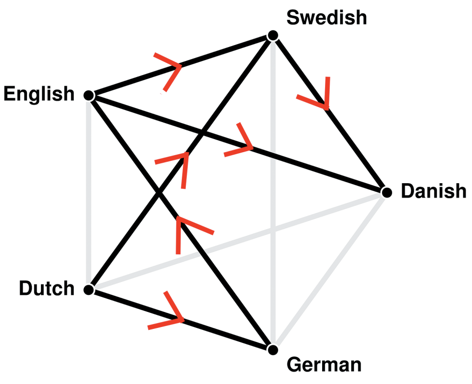

Figure 6.11: Asymmetries in syntactic trigram distance. Asymmetries were established using paired-sample t-tests. An arrow from language A to language B predicts that the native speakers of language A understand language B significantly more easily than native speakers of language B understand language A at the .05 level. Source: Heeringa et al. (2018: 291).

texts in Swedish and German than the other way around. For native speakers of English, it may be easier to understand texts in Swedish and Danish, than the other way around. Swedes may more easily understand Danish texts than Danes understand Swedish texts. Germans may more easily understand English texts than the English understand German texts. Unfortunately, it was not possible to correlate asymmetric intelligibility scores with the asymmetric syntactic scores, as all language combinations involving German or English had too few participants who had not learned the target language in school. Therefore, inherent intelligibility scores were only available in both directions (from A to B and from B to A) for four language combinations. Three of these language combinations had asymmetric intelligibility scores in the predicted direction.

6.3 Conclusions

The investigations presented in this chapter show evidence for the role of various linguistic and extra-linguistic factors in explaining the asymmetric intelligibility observed in some language combinations. Unfortunately, at this point, it is not possible to draw conclusions about which factor is most central. This depends on the socio-political circumstances and linguistic characteristics of the individual language combinations. To investigate the relative importance of the various factors, we need to quantify all factors in such a way that they can be compared in a statistical manner, and the measurements should all be based on the same material. However, the various investigations discussed in this chapter do not all lend themselves to this purpose.

The results from the MICReLa project showed statistical correlations between asymmetric intelligibility and asymmetry in extra-linguistic factors. The highest correlation was found between asymmetric intelligibility scores and asymmetric exposure scores ($r = .92$, see Figure 6.1). This seems logical as it was already concluded in Section 4.1 that exposure is the most important factor in explaining and predicting intelligibility. Asymmetric attitude scores showed low correlations with asymmetric intelligibility ($r = .53$ for all language combinations and insignificant correlations for the Romance and Slavic language families; see Figure 6.3). Even though various investigations showed that linguistic factors may also explain asymmetric intelligibility, it was only possible to establish a statistical link with the lexical asymmetries from the MICReLa project ($r = -.42$, see Figure 6.7).

Chapter 7
Towards a model of mutual intelligibility

The results of various investigations presented in the previous chapters show that we have come a long way in identifying and measuring factors that predict and explain how well speakers of closely related languages understand each other. All the linguistic dimensions presented in Figure 1.1 in the introduction (lexical, phonetic, morpho-syntactic, paralinguistic) show clear relationships with scores from functional intelligibility experiments. In addition, many extra-linguistic factors correlate with intelligibility. Since these various linguistic and extra-linguistic factors interact, it is unclear how much each factor contributes to determining mutual intelligibility. Ideally, a model of intelligibility would be developed that includes all of the relevant factors and shows their relative contribution.

Such a model could help us understand what factors *explain* intelligibility and how these factors interact. By correlating linguistic and extra-linguistic factors with intelligibility scores, we get a better understanding of the factors underlying intelligibility. Such knowledge has many theoretical and practical implications, as discussed in Chapter 8.

The model could also serve to *predict* intelligibility. Intelligibility is usually expressed as a single number (for instance, the percentage of correct answers in one of the intelligibility tests discussed in Chapter 2). However, as has become clear, it can be complicated, time-consuming, and not always feasible to carry out intelligibility tests. It is, therefore, desirable to develop a model of mutual intelligibility that can serve as a shortcut to predict intelligibility. The results of intelligibility tests can be assumed to reflect the totality of both social and linguistic factors that play a role in mutual intelligibility. By understanding the factors that contribute to intelligibility, it is possible to estimate intelligibility relations that cannot feasibly be measured or can only be measured with a large effort. Ultimately, it may even afford a more accurate estimate of intelligibility than intelligibility testing itself because of the many problems with testing. However, until we better understand the factors underlying intelligibility and their interactions, modelling may not actually replace the intelligibility testing approach. At present, the two approaches (functional testing and quantification of relevant factors) complement each other.

It is not a simple undertaking to develop a model of intelligibility. As we have seen, there is a large number of linguistic and extra-linguistic factors that should be taken into consideration when predicting and explaining mutual intelligibility. These factors all need to be quantified to include them in a statistical model. However, even if the most important factors are quantified satisfactorily, there is no

https://doi.org/10.1515/9783111134697-007

a priori way of weighing the different linguistic and extra-linguistic dimensions. The relative weight is different per language combination. It is reasonable to expect lexical distances to be the best predictors of intelligibility. Without word recognition, listeners have no way of understanding a message. For example, Gooskens and Swarte (2017) found a correlation of $r = -.95$ between inherent intelligibility scores, including 20 Germanic language combinations and lexical distances, while the correlation with phonetic distance was non-significant ($r = -.28$). Salehi and Neysani (2017) also found lexical distance to be better predictors of intelligibility of Turkish among Iranian-Azerbaijani speakers than phonetic distance. They explain this finding by the fact that the phonetic distances between Turkish and Azerbaijani are small and highly rule-governed. However, even if listeners know all the words they may not be able to recognize them if they are pronounced very differently from the corresponding words in their own language. For this reason, phonetic distances may also play an important role in predicting and explaining intelligibility. Gooskens, Heeringa, and Beijering (2008, see Example 2.15 in Section 2.2.3.4), correlated lexical and phonetic distances with the intelligibility of 18 Scandinavian language varieties among Danish listeners and found phonetic distance to be a stronger predictor of intelligibility ($r = -.86$) than lexical distance ($r = -.64$). This could be explained by the fact that there is little lexical variation between the Scandinavian languages as explained in Section 5.1.

These findings confirm the assumptions by Mckaughan (1964, summarized in Grimes 1992: 28) that it depends on the degree of divergence between languages which linguistic level best reflects intelligibility. For measuring differences between dialects that are only slightly different, McKaughan argues that methods for measuring distances at the phonetic level are more appropriate. Methods for measuring lexical distances are useful for reflecting mutual intelligibility between languages that are more divergent. Finally, widely divergent languages should be measured by structural measures as well. Therefore, instead of choosing one level for measuring linguistic distances, it would often be preferable to include all levels as predictors of intelligibility.

Within the various linguistic levels, some differences influence intelligibility more than others. To weigh the different linguistic differences, Casad (1974: 121–122) suggests ranking features at different levels in order of importance for predicting intelligibility. For instance, he assumes that phonetic differences between individual segments are less important than supra-segmental differences, such as tone or stress, and that changes in the stem of a word are more important than prefix changes. He notes that the ranking of the features must be validated by correlating with intelligibility scores. The rank order values could then serve as weighting factors. This involves summing up the total number of differences in each feature class in a data set, such as a list of sentences, and multiplying each sum by the appropriate class rank value. However, he notes that sociological factors, such as the kind, degree, and purpose of social interaction, as well as attitudes should also be considered when predicting and explaining mutual intelligibility.

There tends to be interaction between the differences at various linguistic levels. Heeringa, Gooskens, and van Heuven (2023) correlated the lexical, phonetic, and syntactic distance measures from the MICReLa project and found weak but significant correlations. The correlations among the linguistic levels were stronger when they were calculated separately for language pairs within the three language families (Germanic, Romance, and Slavic) than when all 35 language pairs were examined together. Extra-linguistic factors also interact with each other and with linguistic distances. Van Bezooijen (2002) formulated the "intelligibility-driven hypothesis" and the "similarity-driven hypothesis", which state that intelligibility and similarity are both factors that may explain attitudes toward languages. She found that, as dialects are more intelligible and closer to the standard, they receive more favorable aesthetic evaluations. An investigation involving the mainland Scandinavian languages found significant correlations between intelligibility scores on the one hand and phonetic distances ($r = -.81$) and attitudes ($r = .56$) on the other. However, only phonetic distances were included in a stepwise regression analysis with both predictors of intelligibility (Gooskens 2007). The reason that attitude did not add significantly to the prediction of intelligibility is that attitude intercorrelated highly with phonetic distance. The correlation was significant at the 1% level ($r = -.62$). Generally, the listeners were more positive about the neighboring languages if they were phonetically similar to their own variety and less positive if the phonetic distance was larger.

Other previous results show that features sometimes correlate only weakly amongst each other while significantly correlating with intelligibility scores individually and can therefore be assumed to all contribute to the explanation and prediction of intelligibility. For example, Heeringa et al. (2014) measured phonetic stem and affix differences between 20 Germanic language combinations, i.e., all possible pairings (AB and BA) of non-identical languages from the set of Danish, Dutch, English, German, and Swedish in the MICReLa project (see Chapter 3). The two linguistic distance measures (stem and affix) did not correlate significantly; when modelling intelligibility, it is therefore important to be aware of the difference between the two measures (see Section 5.3.1).

It depends on the purpose of the model which factors should be included. Only linguistic factors are relevant if the aim is to model inherent intelligibility. This means that intelligibility should be measured in such a way that extra-linguistic factors are excluded as well as possible, as discussed in Section 4.7. However, if the aim is to make a complete model of actual intelligibility in a particular area, extra-linguistic factors should also be included in the model. Figure 1.1 in Chapter 1 shows an overview of the linguistic and extra-linguistic determinants of intelligibility discussed in this book. The various investigations summarized throughout the book have shown that all the determinants can be relevant for explaining intelligibility results. A possible approach can be illustrated with data from the MICReLa project (see Chapter 3). The following section presents statistical models of general and inherent intelligibility based on measurements from this project.

7.1 Models of intelligibility

In the MICReLa project (see Chapter 3), measurements of linguistic and extra-linguistic determinants were collected to explain the results of functional intelligibility measurements. Three linguistic and three extra-linguistic determinants were quantified (see Figure 7.1). The quantifications of the measurements have been described in various sections of the book. The six determinants have been shown to be important predictors of intelligibility. Tables with the measurements can be found in Appendix C and Appendix D. Morphological, prosodic, and paralinguistic distance measures were not available, and the number of extra-linguistic determinants included was limited to three to keep the experiments within a short time limit.

The following sections present correlations between each determinant and the intelligibility results to assess how well each measure predicts intelligibility. In addition, the correlations between the determinants are included to get an impression of intercorrelations. Based on a regression analysis, a simple statistical model of intelligibility with a limited number of determinants can be developed. First, a model of general intelligibility is developed (Section 7.1.1), including data from all participants. Next, inherent intelligibility is modeled by making a selection of the data (Section 7.1.2).[24] Finally, a model is presented that demonstrates the separate contributions of consonant and vowel substitutions, insertions, and deletions (Section 7.1.3).

24 Similar analyses were presented in Gooskens & van Heuven (2020). This chapter uses a partly different set of explanatory factors and measurements.

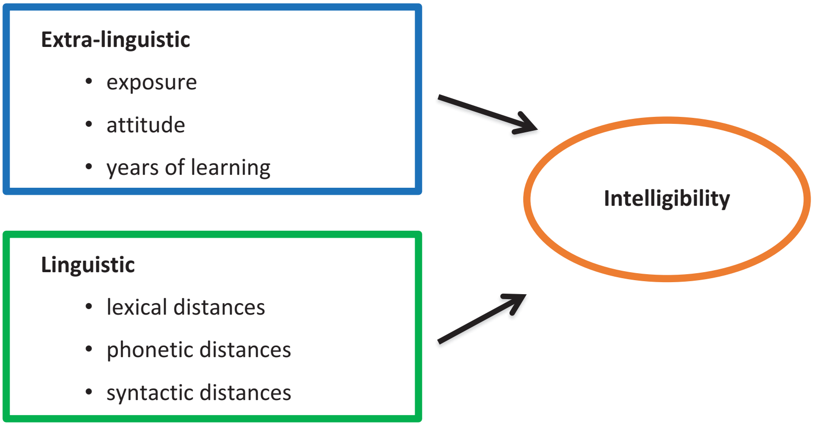

Figure 7.1: Linguistic and extra-linguistic determinants of intelligibility from the MICReLa project included in the statistical model of general intelligibility. In the model of inherent intelligibility, only linguistic determinants are included.

7.1.1 A model of general intelligibility

The model of general intelligibility presented here includes intelligibility scores of all 1833 listeners who participated in the spoken cloze test. Some of the listeners had learned the target language at school or had been exposed to it in other ways to varying degrees, while others had little or no previous exposure to the language. This means that the results reflect inherent and acquired intelligibility to various degrees.

First, the mean intelligibility results of each language combination were correlated with six linguistic and extra-linguistic determinants. The results are presented in Table 7.1. Correlations were calculated including all 70 language combinations and for each language family separately (20 language combinations for Germanic and Romance and 30 for Slavic). Table 7.2 also shows the correlations between linguistic and extra-linguistic determinants since this may help interpret the results of the regression analyses.

The correlations between intelligibility (top row) and exposure are significant and strong, ranging from r = .87 for Romance and .93 for Germanic, and reaching .90 when all language combinations are included. In the Slavic language area, there tends to be little exposure to closely related languages. Still, the correlation between intelligibility and exposure is high in the Slavic area (r = .92) due to three outliers (see Figure 4.2).

Some target languages are school subjects in the Germanic (English and German) and Romance (French) language areas. The correlation between intelligibility and the number of years listeners learned the target language is significant only for the Germanic group (r = .86). As anticipated, the correlation between exposure and years of learning is strong in the Germanic group (r = .92), but weaker in the Romance group (r = .45). This is probably because many Romance listeners learn French at school but have limited exposure to it outside of the classroom (see Appendix C). In the Slavic group, the correlation is not significant, primarily because the Slavic target languages are seldom taught in schools across the six Slavic countries involved.

Attitude shows rather high significant correlations with intelligibility scores (between r = .60 and .81 for the three language families). A strong positive correlation was found between exposure and attitude (ranging from r =.68 to .78). This implies that individuals generally hold more positive attitudes towards languages they are familiar with than to those they are less exposed to. However, the direction of cause and effect is not clear: it could also be that listeners listen to languages that they have positive attitudes towards more often.

Intelligibility in the Slavic language family shows significant correlations with all linguistic distance measures (between r = −.62 and −.85). In the Germanic and the Romance groups, the correlations between intelligibility and linguistic variables are insignificant (except for a low correlation with phonetic distances in the Romance group, r = −.51).

A noteworthy observation is the significant correlation between exposure and linguistic distances within the Slavic language family, ranging from r = −0.63 to −0.82. This suggests that Slavic individuals tend to have more exposure to languages that share similarities with their own language than to those that are linguistically distant. This can be attributed to the geographic proximity of closely related languages. This relationship between exposure and linguistic distances is not significant in the other two language families, except for a low correlation (r = −.45) with phonetic distances in the Romance language family. There are also significant negative correlations between attitude and linguistic determinants in the Slavic language family. The listeners in this family have more exposure and more positive attitudes to languages that are similar to their own language than to more deviant languages.

Table 7.1 also shows that there are many significant correlations among linguistic distances. Lexical distances correlate significantly with both other linguistic distances in all the language families (between r =.41 and .81). Phonetic and syntactic distances correlate significantly in the Romance and Slavic families (r = .59 and .69). See Heeringa, Gooskens, and van Heuven (2023) for more details.

To investigate how well intelligibility scores can be predicted from the six linguistic and extra-linguistic variables, forward stepwise regression analyses were carried out. A regression analysis can reveal the relationship between a depen-

Table 7.1: Correlation coefficients between general intelligibility scores and the six determinants across all 70 language combinations and for the three language families separately (20 Germanic, 20 Romance, and 30 Slavic language combinations).

	Exposure	Learning	Attitude	Lexical	Phonetic	Syntactic
Intelligibility						
All	.90**	.60**	.65**	−.38**	−.40**	−.40**
Germ.	.93**	.86**	.60**	−.21	−.17	−.30
Rom.	.87**	.19	.70**	−.36	−.51*	−.40
Slav.	.92**	–	.81**	−.82**	−.85**	−.62**
Exposure						
All		.76**	.61**	−.18	−.26*	−.31**
Germ.		.92**	.68**	.07	.06	−.25
Rom.		.45*	.78**	−.41	−.45*	−.40
Slav.		–	.78**	−.70**	−.82**	−.63**
Learning						
All			.32**	.19	.18	.04
Germ.			.61**	.30	.21	−.05
Rom.			.34	−.12	.19	.06
Slav.			–	–	–	–
Attitude						
All				−.38**	−.22	−.51**
Germ.				.04	−.03	−.16
Rom.				−.43	−.27	−.41
Slav.				−.69**	−.76**	−.71**
Lexical						
All					.51**	.52**
Germ.					.70**	.47*
Rom.					.48*	.81**
Slav.					.80**	.41*
Phonetic						
All						.26**
Germ.						.11
Rom.						.59**
Slav.						.69**

* $p \leq .05$; ** $p \leq .01$ (two-tailed).

dent variable (here, the mean intelligibility scores per language combination) and independent variables (the six linguistic and extra-linguistic variables presented in Figure 7.1). In the stepwise regression analyses presented in the rest of this chapter, the variable that explains the highest percentage of variance (expressed as the squared multiple correlation coefficient R^2) is presented in the top row. In the following rows, the remaining variables are added one by one, starting with

Table 7.2: Stepwise regression analyses with mean general intelligibility score as the criterion and six linguistic and extra-linguistic predictors.

Language combination	Predictors	R^2	ΔR^2	t	p
All	Exposure	.82		18.962	< .001
	Lexical distance	.87	.05	−4.994	< .001
Germanic	Exposure	.85		2.316	.034
	Lexical distance	.92	.07	−9.130	< .001
	Years of learning	.97	.05	5.740	< .001
Romance	Exposure	.76		7.566	< .001
Slavic	Exposure	.85		8.437	< .001
	Lexical distance	.91	.06	−4.250	< .001

the variable whose inclusion gives the most statistically significant improvement until the percentage of explained variance is no longer improved significantly.

The results of a regression analysis for all language combinations and separately for each of the language families are presented in Table 7.2. As expected, exposure is the most important predictor of intelligibility for all three language families. Lexical distances are also included in Germanic and Slavic models but add little to the predictive power. The Germanic model incorporates the number of years that the listeners have been studying the target language. However, this extra-linguistic variable adds little to the model, likely due to the high intercorrelation with exposure.

It is evident that individuals who acquire a language through exposure or formal education will understand it better than those with limited prior exposure, irrespective of linguistic differences. Hence, exposure supersedes linguistic factors. However, it is also interesting to know how well intelligibility can be predicted from linguistic distances only. Additional analyses were therefore carried out where the extra-linguistic predictors were left out. The results are presented in Table 7.3.

In the case of the Germanic language family, linguistic distances have no predictive power (no predictors are included in the model), as could be expected from the fact that none of the linguistic determinants correlate significantly with intelligibility measures. As discussed above, high intercorrelations between extra-linguistic and linguistic variables were found for the Slavic family (see Table 7.1). This may explain why Slavic linguistic (specifically phonetic and lexical) distances are rather good predictors of intelligibility: R^2 is .77, which means that 77% of the variance is explained by phonetic and lexical distances together. In the Romance language family, 27% of the variance is explained by phonetic distances. Overall, it can be concluded that linguistic distance measurements are generally uncertain predictors of general intelligibility.

Table 7.3: Stepwise regression analyses with mean general intelligibility scores as criterion variable and three linguistic predictors only.

Language combinations	Predictors (distances)	R^2	ΔR^2	t	p
All	Phonetic distance	.16		−2.940	.004
	Syntactic distance	.25	.09	−2.889	.005
Germanic	–	–		–	–
Romance	Phonetic distance	.27		−2.545	.020
Slavic	Phonetic distance	.71		−6.467	.002
	Lexical distance	.77	.06	−2.592	.018

7.1.2 A model of inherent intelligibility

The regression analyses presented in Table 7.3 demonstrate that linguistic distances are not always good predictors of general intelligibility because they are "drowned out" by extra-linguistically determined variance. However, this does not mean that linguistic distances should be ignored as determinants of mutual intelligibility. To predict inherent intelligibility, it makes sense to include linguistic distances only since extra-linguistic factors should not play a role. To assess the predictive power of linguistic distances on inherent intelligibility, it is preferable to establish a correlation between linguistic distances and intelligibility scores from listeners who have had no prior exposure to the target language. This enables us to evaluate how well listeners comprehend the target language based solely on its resemblance to their native language. The predictive power of linguistic distances in determining intelligibility scores is expected to be more pronounced within this subset of listeners devoid of prior exposure to the target language, in contrast to the broader cohort, which includes individuals with previous exposure. As explained in Section 3.1.4, a subset of listeners was chosen based on their self-reported mean exposure ratings of below 2.0 (with 1 for "no exposure") on six five-point scales and on their absence of formal education in the target language. The intelligibility scores of this group are likely to be based mainly (or even exclusively) on the similarity between the listener language and the target language.

The correlations with linguistic distances obtained with the reduced data set are higher than those obtained for the whole data set (compare top rows in Tables 7.1 and 7.4). In the Germanic language family, the correlation with lexical distances has increased significantly, from $r = -.21$ to $-.95$. Likewise, in the Romance language family, the correlation with lexical distances increased substantially, from $r = -.36$ to $-.69$. The correlations with syntactic distances show similar improvements for the Germanic and the Romance language families. The correlations in the Slavic language family were already high when the results from all listeners were included, and they remained high when filtered for exposure and years of learning the language. This is likely because listeners from this language family had limited prior exposure to the target languages, so that exposure played a minor role in determining general intelligibility.

The only insignificant correlation in Table 7.4 is the correlation with phonetic distances in the Germanic language family. The intelligibility scores for Danish-Swedish and Swedish-Danish language pairs stand out as outliers (see Figure 5.5). The listeners in these two language combinations understand each other well owing to the minimal lexical differences between the two languages (a distance of 4.6% for Danish-Swedish and 5.8% for Swedish-Danish, see Figure 5.2 and Appendix D), while in other language combinations, lexical distances are a greater obstacle. Excluding these two language combinations results in a significant correlation with phonetic distance in the Germanic language family as well (an increase from $r = -.53$ to $-.73$, $p < .01$).

Table 7.4: Correlations between inherent intelligibility and three linguistic distances for all 57 language combinations and for the three language families separately.

	Lexical	**Phonetic**	**Syntactic**
All	−.76**	−.66**	−.56**
Germ.	−.95**	−.53	−.68**
Rom.	−.69**	−.68**	−.77**
Slav.	−.80**	−.80**	−.52**

* $p \leq .05$; ** $p \leq .01$ (two-tailed).

Table 7.5 presents the results of stepwise regression analyses with the three linguistic variables as independent variables (predictors) and inherent intelligibility scores as the dependent variable (criterion). The results show that a combination of three linguistic distance measurements can predict inherent intelligibility to a relatively high extent. When comparing the results to those of the regression analyses with general intelligibility (Table 7.3), we observe a markedly higher predictive power of the linguistic variables, particularly in the Germanic language family, where lexical distances alone account for 90% of the variance. Given that this pertains to spoken language, we might expect phonetic distances to be a contributing factor. However, as discussed earlier, the correlations between intelligibility and phonetic distances are low in the Germanic language family, mainly due to the Danish-Swedish outliers. In the Slavic family, the percentage of explained variance is also substantial, with phonetic and lexical distances together accounting for 73% of the variance. While the predictive power is high for the Germanic and the Slavic families, the situation is not as straightforward for the Romance family. Surprisingly, syntactic distance is the only variable that needs to be included in the model. This can be explained by intercorrelations between the three linguistic distance measurements (see Table 7.1). For instance, there is a high correlation between syntactic and lexical distances (r = .81). The predictive power of the optimal model is rather low (59%).

Table 7.5: Stepwise regression analyses with mean inherent intelligibility scores per language combination as the criterion and three linguistic distances as predictors.

Languages	Predictors (distances)	R^2	ΔR^2	t	p
All	Lexical distance	.57		−4.743	< .001
	Phonetic distance	.70	.13	−5.154	< .001
	Syntactic distance	.73	.03	−2.290	.026
Germanic	Lexical distance	.90		−10.567	< .001
Romance	Syntactic distance	.59		−4.329	.001
Slavic	Phonetic distance	.65		−2.918	.007
	Lexical distance	.73	.08	−2.805	.009

7.1.3 Consonant and vowel substitutions, insertions and deletions

Section 5.2.1 discussed the nature of edit operations (substitutions, insertions, or deletions) and their contribution to intelligibility. There is evidence that when listeners hear a target word in a closely related language, they first look for a word that differs from the test word in one segment (substitutions) and next for correspondences that have one segment more or one segment less (insertions/deletions) in their native variety. This is explained by the fact that segment insertions or deletions may change the structure of words, while in the case of a substitution, the segmental framework of a word remains unaltered. This effect is likely to be stronger for consonants than for vowels.

To investigate the role of consonant and vowel substitutions and insertion/deletions (indels), a separate analysis was carried out with measurements for vowel indels, consonant indels, vowel substitutions, and consonant substitutions. The measurements were calculated in a similar manner as for the overall measurements. First, the full phonetic strings of the cognate pairs per language combination were aligned to each other using the Levenshtein algorithm (see Section 5.2). Once the optimal alignment was found, the distance was based on the alignment slots in which only the relevant operation was involved, i.e., slots with either vowel indels, consonant indels, vowel substitutions, or consonant substitutions. This distance was divided by the length of the full alignment. In the example in Table 5.2, there are two vowel operations (a substitution and an insertion), one consonant operation (a deletion) and 8 alignment slots. Each operation counts for 1/8 = 0.13 or 13%.

Table 7.6 shows the correlations between intelligibility scores and all four phonetic distances. In the Germanic language family, the distances based on consonant substitutions (Csubst) and vowel substitutions (Vsubst) correlate significantly with intelligibility (r = −.66 and −.65), while insertions/deletions (Vindel

Table 7.6: Correlation coefficients between inherent intelligibility scores and six linguistic predictors for 61 language combinations and for the three language families separately. Vindel = vowel insertions/deletions, Vsubst = vowel substitutions, Cindel = consonant insertions/deletions, Csubst = consonant substitutions.

	Phonetic	Vindel	Vsubst	Cindel	Csubst
Intelligibility					
All	−.66**	−.17	−.34**	−.45**	−.39**
Germ.	−.53	.12	−.66*	−.18	−.65*
Rom.	−.68**	−.37	.03	−.74**	.21
Slav.	−.80**	−.25	−.48**	−.70**	−.51**
Phonetic					
All		.56**	.51**	.74**	.37**
Germ.		.26	.64*	.73**	.29
Rom.		.76**	−.15	.54*	−.23
Slav.		.41*	.65**	.82**	.42*
Vindel					
All			.18	.42**	−.10
Germ.			−.23	.14	−.12
Rom.			−.20	.35	−.55*
Slav.			.31	.22	−.28
Vsubst					
All				.26*	.08
Germ.				.19	.15
Rom.				−.03	.09
Slav.				.36	−.02
Cindel					
All					.30*
Germ.					−.12
Rom.					−.54*
Slav.					.37*

* $p \leq .05$; ** $p \leq .01$ (two-tailed).

and Cindel) do not. On the other hand, consonant indels correlate highly with intelligibility in both the Romance and the Slavic language family (r = −.74 and −.70), and in the Slavic family, vowel and consonant substitutions also show significant correlations with intelligibility (r = −.48 and −.51).

There is generally not a lot of intercorrelation between the four phonetic sublevels. In particular, there are no intercorrelations in the Germanic language family. Therefore, in a stepwise regression analysis (see Table 7.7), all four levels are included in the model for this language family. The explained variance R^2 is very high (94%). It seems counterintuitive that the combination of the four phonetic levels results in such a high percentage of explained variance, while the overall phonetic distances have a lower percentage of explained variance (28%). A closer look at the data shows that there are only a few vowel and consonant substitutions between Swedish and Danish. The small distances and high intelligibility scores for this language combination cause the overall correlations to be very high. When leaving out the language combinations with Swedish and Danish (see the extra rows in Table 7.7), the Germanic and the overall results change considerably. Now only consonant indels remain as predictors.

In the Slavic model, three phonetic levels are included (65% explained variance). In contrast, only consonants indels are included in the Romance model (55%). Importantly, consonant indels are included in all three models, thus confirming the special role of consonants and indels in predicting intelligibility.

Table 7.7: Stepwise regression analyses with mean inherent intelligibility scores per language combination as the criterion and four phonetic distances as predictors.

Languages	Predictors (distances)	R^2	ΔR^2	t	p
All	Consonant indel	.20		−2.520	.015
	Consonant substitution	.27	.07	−2.340	.023
	Vowel substitution	.32	.05	−2.034	.047
All without Danish and Swedish	Consonant indel	.35		−5.349	< .001
Germanic	Vowel substitution	.44		−8.379	< .001
	Consonant substitution	.75	.31	−6.642	< .001
	Consonant indel	.86	.11	−4.676	.001
	Vowel indel	.94	.08	3.577	.006
Germanic without Danish and Swedish	Consonant indel	.34		−2.243	.049
Romance	Consonant indel	.55		−3.971	.002
Slavic	Consonant indel	.49		−3.267	.003
	Consonant substitution	.56	.07	−2.720	.012
	Vowel substitution	.65	.09	−2.508	.019

7.2 Conclusions

The statistical models presented in this chapter exemplify how intelligibility modeling can be approached. For several reasons, developing a model that can predict intelligibility perfectly is not realistic at this time. Such a model would have to be highly language-dependent and involve many linguistic and extra-linguistic factors (see Chapters 4 and 5). We have seen that even within the same Indo-European language family, there are large differences in the relative importance of the various determinants of intelligibility. In future research, to make better generalizations about the relative weight of the determinants, it is therefore important to include data from various contexts and from many different languages and language families. Additionally, an improved model may include more factors and subfactors. As discussed in Chapters 4 and 5, the various measurements should also be refined to better predict intelligibility since each determinant can vary in many dimensions.

However, even incomplete models like the ones presented in this chapter can often explain and predict intelligibility to a high degree and may provide valuable insights. They can show us what factors are central in predicting intelligibility and how they intercorrelate, advancing our knowledge of the mechanisms behind intelligibility. They may also be helpful in situations where we need information about intelligibility but where it is not feasible to conduct intelligibility testing.

Part III: **Applications and conclusions**

Chapter 8
Applications of mutual intelligibility studies

As mentioned in Chapter 1, there are various reasons to investigate the mutual intelligibility between languages and dialects. These range from theoretical and scientific interest to more practical issues such as language planning and policy and cover cultural, communicative, educational, and economic matters. This highlights the broad significance of mutual intelligibility across various domains of human interaction.

8.1 Fundamental knowledge

Intelligibility research bears on central questions of linguistics, and it adds to our fundamental knowledge of language. One of the main functions of language is communication (Crystal and Robins 2021). Human speech is a remarkably robust system of communication that generally succeeds even in difficult listening conditions. Yet, pinpointing exactly how and why people manage to communicate successfully, even in the face of communication barriers, remains an ongoing challenge for language researchers. Even in situations where the interactants have the same native language background, they often have to deal with speech in non-optimal situations, such as speech produced with background noise, competing speech input from other speakers, low amplitudes, hearing or speech disorders, and a bad internet or telephone connection. Another non-optimal situation where some level of communication can still be achieved is when listeners encounter a language that has only some partly overlap with their native language. Intelligibility research investigates such non-optimal situations and tries to understand what factors make mutual intelligibility more difficult, what factors are less obstructive to intelligibility, and how these factors interact. In doing so, we gain fundamental knowledge about the human language faculty and its limitations.

It is presently not possible to define a breakdown point where languages are so different that they are no longer mutually intelligible. Moreover, we do not know whether the relationship between the number and magnitude of deviations and intelligibility is linear or non-linear. It could be hypothesized that recognition of words in a closely related language remains good if they are similar to the corresponding words in the native language of the listener but abruptly breaks down when the differences become too large. However, at this point, it is not possible to define at which point this breakdown takes place and how various linguistic factors interact in different languages. The investigations summarized in

https://doi.org/10.1515/9783111134697-008

this book have made a beginning. However, it is complicated to develop a complete model of intelligibility since so many linguistic and extra-linguistic factors play a role, and each language and language combination represents a unique situation.

Intelligibility research is also fundamental to sociolinguistic theories. When different language varieties are in contact through interaction, they are likely to become more alike with time (linguistic leveling). Since there is a clear link between intelligibility and the amount of exposure (see Sections 4.1 and 7.1.1), intelligibility will likely play a crucial role in this process. Therefore, intelligibility is also critical for understanding mechanisms behind language shift. Trudgill (1986: 21–23) explains how mutual intelligibility plays a role in accommodation processes in addition to socio-psychological factors, such as the desire not to be different from other speakers in our social group. Misunderstandings may cause speakers to accommodate their pronunciation to the language of the listeners and eventually lead to language shift. An example of misunderstandings caused by pronunciation differences between British English and American English provided by Trudgill is the following:

> *I can attest that one factor that without doubt precipitated the introduction of flaps* (of intervocalic [t]) *into my own speech in America was the number of people who thought, for example, if only for a second, that I wanted a pizza rather than that my name was Peter* (Trudgill 1986: 23).

This example shows how an outsider's deviant pronunciation can lead to miscommunication and as a consequence cause the outsider to adapt his speech to that of the locals.

8.2 Defining "language" versus "dialect"

Linguists are commonly asked how many languages there are worldwide (Anderson 2012). This may sound like a simple question, but it is not easy to answer. Ethnologue's classified list as of 2023 includes 7,168 distinct languages (Eberhard, Simons, and Fennig 2023). This number is constantly changing because many languages have only a few speakers and are, therefore, in danger of extinction. Linguists are also gaining an increased understanding of how many languages are actually spoken in areas that have previously not been described in detail. However, a significant challenge arises from the fact that the definition of "language" as opposed to "dialect" is not clear-cut. To be able to quantify the number of languages it is necessary to agree on a way to distinguish the two.

Also at the language planning and policy level, it may be crucial to establish criteria that define a language variety as either a language or a dialect. A language often represents a community and is tightly linked to standardization processes and the development of orthographies. The choice or development of a single shared standard variety and a single orthography for a number of closely related language varieties may save time and effort. It will make it possible to develop teaching resources and orthographies that can be used in a larger area with a larger number of people rather than creating the material for each variety independently. To decide on orthographic forms it is necessary to gain knowledge about which language varieties are similar enough to share an orthography. For example, during the fifties of the previous century, there was a strong interest in establishing the mutual intelligibility of American Indian languages (Hickerson, Turner, and Hickerton 1952; Pierce 1952; Voegelin and Harris 1951). The aim was to investigate the genetic relationship between language varieties and to develop or adopt a single orthography for multiple closely related language varieties within literacy programs. Similar projects are still being carried out worldwide, particularly by SIL International in the context of language development, language planning and education.[25] It is not entirely clear when differences between varieties become too great to be bridged with one and the same orthography. By means of recorded text testing (RTT, see Section 2.2.3.2), Brye and Brye (2002) tested the mutual intelligibility between speakers of the Eastern Beboid varieties Nooni, Ncane, and Nsari, spoken in Cameroon. They aimed to investigate the potential for extendibility of Nooni literature to the other speech varieties. For the choice of a threshold, they refer to Grimes (1992: 22), who states that a score of 85% is necessary to speak of a dialect cluster as a single language. In addition, the standard deviation should be below 15%, as a higher deviation would probably indicate that some individuals have had exposure to the variety. Grimes also notes that the final threshold depends on the importance of communicating well and on extra-linguistic criteria. For example, some of the participants in Brye and Brye's investigation showed an unwillingness to pursue a written form based on other varieties than their own.

For individuals fighting for the rights of a specific language variety, it holds immense significance that the variety is acknowledged as a language rather than a dialect, as the official recognition can give the variety a stronger position. For instance, in Part 1 of the European Charter for Regional or Minority Languages (Council of Europe 1992), the entitlement to use one's variety in public life (e.g., in educational, juridical, administrative, or media contexts) is contingent on the sta-

25 https://www.sil.org/language-development

tus of this variety as an official, regional or minority language. Those who speak varieties classified as “dialects” do not have these rights under the Charter. Conversely, there are also instances where clearly distinct languages are labeled as dialects of a single language because it is desirable to maintain the perceived unity of the region. Often mentioned examples are Cantonese and Mandarin. These two Chinese varieties are not mutually intelligible but are still often referred to as dialects of the same language, probably because of the shared writing system and to stress the political, social, and cultural unity of the country (Kurpaska 2010; Wardhaugh and Fuller 2021).

In search of criteria for deciding whether a language variety should be considered a language or a dialect, Kloss (1967) introduced the terms *Ausbausprache* (language by development) and *Abstandsprache* (language by distance). *Abstand* languages (including dialects) are “intrinsically distant to others” (Kloss 1967: 29). One language is distinguished from another based on linguistic criteria, i.e., because they differ substantially in pronunciation, vocabulary, and grammar. There is “a definite break” between the two languages, and “the linguist would not hesitate for a single minute to list the two separately” (Kloss 1967: 30). Two dialects within one language are less different in structure than two dialects of different languages. Varieties from different language families, such as French (a Romance language) and German (Germanic), are two different *Abstand* languages. However, languages from the same family, such as German and Dutch (both Germanic), can also be considered *Abstand* languages. An *abstand* language is not necessarily required to have a standardized form. This circumstance frequently arises with minority languages, which are often only used in private life, while official functions are conducted in the majority language.

According to the *Ausbau* definition, languages and dialects are social constructs that depend on their socio-political and cultural status and breadth of use, and structural characteristics of the language varieties may be of less importance. An *Ausbausprache* generally refers to a language variety with its own standardized form taught in schools and used in various socio-political contexts. It is typically an official national language. The Scandinavian languages (Danish, Norwegian, and Swedish) are examples of *Ausbausprache* that are quite similar. In another context, they might therefore have been classified as dialects. However, they are still regarded as separate languages because they are spoken in separate nation-states and have distinct standardized forms with their own orthographies, grammar books, and literary works. A similar case can be observed in Czech and Slovak, which are closely related, possibly even closer than the Scandinavian languages. The linguistic distance measurements in the MICReLa project (see Appendix D) showed smaller distances at the phonetic and syntactic levels between Czech and Slovak (11.5 and 10.9%) than between Danish and Swedish (34.5 and 14.7%). The lexical distances between Czech and Slovak are only slightly higher (mean 8.1%) than between Danish and Swedish (5.2%). However, despite the small distances, Czech and Slovak are still considered two different languages. Since the dissolution of Czechoslovakia in 1993, the Czech and Slovak written standards have been the official languages of the Czech and Slovak Republics, respectively, and standardized forms of the two languages are distinguishable and recognizable. There are numerous other situations where two varieties are considered different languages even though they are very similar and mutually intelligible, such as Serbian and Croatian (Bailyn 2010) and Hindi and Urdu (Tripathi 2016). For historical reasons, some language varieties that are not separated by state borders may be recognized as distinct regional languages, even if they are closely related to the state’s primary language or official language. An example is West Frisian, an official language in the Dutch province of Friesland. Like Dutch, it belongs to the West Germanic language family, and the two languages are mutually intelligible to some extent (see van Bezooijen and van den Berg 1999; Gooskens and Heeringa 2004b; van Bezooijen and Gooskens 2005b). Within the borders of the Province of Friesland, anyone has the right to address both the local and the national administration in Frisian. It has an official written form and is taught at primary and secondary schools in Friesland.

In terms of language rights, it may have large consequences for the vitality of a language variety, whether it is considered a language or a dialect. Tamburelli (2014) points out that the *Ausbau* definition may result in a circularity effect. As previously mentioned, according to the Council of Europe (1992; see also Tamburelli 2014), the use of a linguistic variety in official contexts is a right reserved for varieties with language status. However, for a variety to achieve the “language” label, it must have a certain socio-political status. This means that the endangered language varieties that language legislation is intended to protect may, in fact, be excluded from this protection a priori. Tamburelli (2021), therefore, warns that over-reliance on socio-political criteria (the *Ausbau* definition of languages) in the linguistic literature can lead to discrimination of speakers of contested languages at both the social and the institutional levels. He also argues that an approach based on *Ausbau* criteria leads to recognizing as few languages as possible because languages that exhibit considerable linguistic distance from each other are often still considered dialects of the same language. This is ultimately detrimental to the maintenance of linguistic diversity. Tamburelli advocates for a stronger reliance on *Abstand* criteria, which, for instance, would result in the classification of distinct Chinese dialects as separate languages.

When asked to define languages, linguists generally prefer to apply the *Abstand* criterion rather than the *Ausbau* criterion. Kloss (1967) did not specify precisely how the differences between two language varieties can be measured objectively, likely due to a lack of appropriate tools at the time. However, as discussed in Chapter 5, dialectometrists have now developed methods for measuring linguistic distances objectively. Despite these advancements, a fundamental issue remains: languages vary across multiple dimensions, such as lexicon, phonetics, phonology, morphology, and syntax, and there is no straightforward way to assign weights to these dimensions a priori (van Heuven 2008).

Trudgill (2000) introduced the intelligibility criterion, which has since become the primary criterion among many linguists. According to this criterion, dialects are regarded as mutually intelligible varieties, whereas languages are distinguished by linguistic differences significant enough to hinder mutual comprehension among speakers. From this, it follows that a language can be defined as a collection of mutually intelligible dialects. Through an intelligibility experiment, Yang et al. (2020) demonstrated that Jejueo, the traditional variety of speech used on Jeju Island, long classified as a nonstandard dialect of Korean, is not comprehensible to monolingual speakers of Korean. The researchers, therefore, concluded that Jejueo should be treated as a distinct language rather than a dialect of Korean. Similarly, Özek, Sağlam, and Gooskens (2021) presented experimental evidence supporting the categorization of Kurmanji and Zazaki dialects spoken in Eastern Anatolia as separate languages rather than as dialects of Kurdish.

The compilers of Ethnologue also rely on intelligibility as the main criterion for distinguishing between languages (Lewis, Simons, and Fennig 2013). Ethnologue has been criticized for mixing linguistic factors with language-external factors, such as common literature and ethnolinguistic identity (van Rooy 2020: 258) and for dividing language varieties into too many distinct languages (Wichmann 2019: 829). However, Hammarström (2005) asserts that Ethnologue is generally consistent with expert opinions. Glottolog (Hammarström et al. 2023) provides a comprehensive list of languoids (language families, languages, and dialects) and uses mutual intelligibility as its sole criterion to distinguish languages from dialects. Glottolog's classified list as of 2024 includes 8,595 entries of languages, more than the list provided by Ethnologue (7.168, see above).

The intelligibility criterion for defining languages has been problematized by some linguists (e.g., Hammarström 2008). Intelligibility can be measured using functional tests or opinion tests (see Chapter 2). Intelligibility testing is often very labor intensive, and the diversity of tests, test conditions, and listener backgrounds makes it hard to compare results. Intelligibility measurements may be influenced by non-linguistic factors, such as attitude, exposure, and personal characteristics of the listeners (see Chapter 4). Therefore, great care should be taken to exclude such factors (see Section 4.7). In addition, both opinion testing and functional testing produce measurements that express how well subjects can understand a language (variety), implying that intelligibility is a matter of degree. This reflects the situation in many dialect areas. Traditional dialectologists draw isoglosses (lines on a map separating features) to show the geographic distribution of dialectal features, such as a particular word form or pronunciation. A dialect boundary is considered major if several isoglosses coincide (isogloss bundles). An example of an isogloss bundle is presented in Figure 8.1. (Shibatani 1990: 197). It shows a bundle of isoglosses that

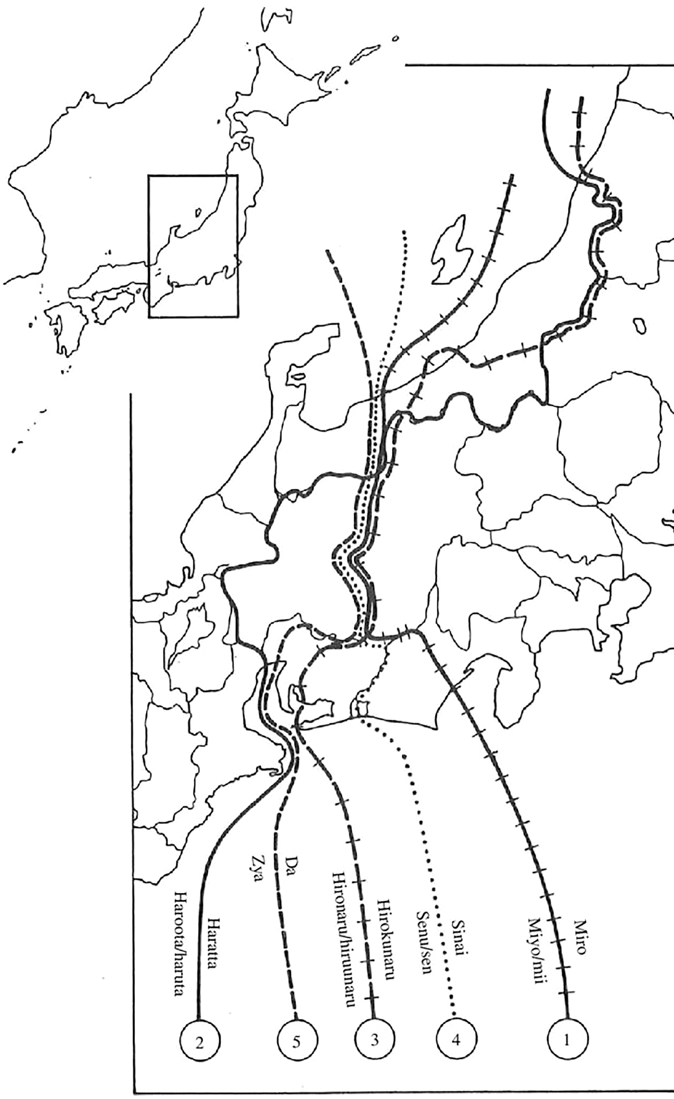

Figure 8.1: Isogloss bundle separating Western Japanese dialects from Eastern Japanese dialects. Source: Shibatani (1990: 197).

cut through the middle of Japan and divide it into Eastern and Western Japanese. The isoglosses show overlap but are spread over a larger area and only coincide approximately, resulting in different so-called dialect continua. This is a range of dialects spoken across a geographic area and differing only slightly between geographically close areas. They gradually decrease in mutual intelligibility and increase in linguistic differences as the distances become greater.

The Dutch-German dialect area known as the Rhenish fan provides another prominent instance of a dialect continuum. In the 19th century, it was possible to travel from the southernmost region of the German-speaking area to the westernmost part of the Dutch-speaking area without encountering any abrupt linguistic barrier that hindered mutual comprehension. Despite this gradual transition, the speech varieties at either end of this chain were so different from one another that they were not mutually intelligible. Today, the internal dialect continua of both Dutch and German remain largely intact. However, the continuum which historically connected the Dutch and German dialects has mostly disintegrated due to leveling towards their respective standard varieties, migration, education, and decreasing use of the dialects. Throughout Europe, various other dialect continua can be found. One example is the Romance continuum, which extends across the Iberian peninsula, parts of Belgium and France, and reaches as far east as the Black Sea, encompassing languages such as Portuguese, Spanish, Catalan, French, Italian, and Romanian. Similarly, many dialect continua outside of Europe are observed, such as in the Chinese, Arabic, Indic, Turkic, and Algonquian language areas.

The presence of dialect continua and their accompanying gradient intelligibility make it challenging to use the criterion of mutual intelligibility to determine the number of languages spoken in a particular region. To tackle this issue, Hammarström (2008) took an abstract approach to demonstrate that it might be feasible to compute such figures. In a scenario where there are three dialects, A, B, and C, in a language region, where the neighboring dialects (A and B, or B and C) are mutually intelligible, but the non-neighboring dialects (A and C) are not, Hammarström used the intelligibility principle to determine that we must be dealing with two languages (A/B and C, or A and B/C). However, although Hammarström shows how to count the number of languages in a continuum, he fails to define languages uniquely using this line of reasoning. Tamburelli (2014) suggests that a choice between the two possible options in the example above can be made by conducting intelligibility tests or measuring objective linguistic distances. The two language varieties that show the highest level of mutual intelligibility or are linguistically closest should be considered dialects of a single language.

From the above discussion, it becomes clear that it is not unproblematic to use intelligibility as a way of objectively substantiating the *Abstand* criterion for distinguishing languages and dialects. Further investigation is required before we can determine when two varieties are so different that they are no longer mutually intelligible. As explained, the alternative *Ausbau* criterion also has drawbacks. So, for the time being, the most workable solution is to keep using a combination of the two criteria and to keep in mind that the definition is not straightforward as it is influenced by several objective and subjective criteria.

8.3 Intelligibility as a distance measurement

In Chapters 5 and 7, it was shown how (combinations of) distance measurements at different linguistic levels can predict and explain the results of intelligibility measurements. The results can provide a greater understanding of what linguistic factors play a role in intelligibility.

Intelligibility measurements can also serve as a means to validate linguistic distance measurements. Recently, dialectometry has advanced significantly in quantifying linguistic distances through diverse algorithms. These advancements have yielded various methods for distinguishing between language varieties. With advancements in such objective techniques for measuring linguistic distances, many researchers have recognized the growing importance of validating these methods by conducting functional tests (Heeringa et al. 2006). As mentioned in Section 8.2, intelligibility measurements can be an adequate way of establishing how different two languages or language varieties are (van Heuven 2008). Unless some extra-linguistic factor interferes (see Chapter 4), a high degree of intelligibility between two language varieties can be expected if the linguistic distance is small, whereas larger distances likely result in lower degrees of mutual intelligibility. The results of various investigations generally show that experimental intelligibility results reflect objective distance measures to a large extent (see Chapter 5).

So, on the one hand, linguistic distance measurements can serve as a shortcut to measuring intelligibility without having to carry out experiments, and objective distance measurements can help explain intelligibility results. On the other hand, intelligibility measurements can aid in validating and refining algorithms for measuring communicatively relevant linguistic distances.

8.4 Genetic relationship

Language varieties change over time and diverge due to innovations in the varieties. Such innovations may be driven by internal linguistic forces (such as ease of pronunciation) or social factors (such as accommodation and speakers' use of identity markers to show that they belong to a certain group). Over the centuries, languages that were originally homogeneous have diversified into many varieties. In historical linguistics, this process is often represented by language trees. A language tree represents a language family, and the different varieties are represented by branches. Generally, the closer the branches are the more recent the innovations and the more similar the varieties are. For example, in the Germanic language tree presented in Figure 8.2, the closely related Scandinavian languages Danish and Swedish belong to the same branch (North), while English, Dutch, and German are more distant from these two languages and belong to another branch (West). Within the West branch, Dutch and German are more closely related to each other than to English.

Distances between languages in the historical language trees can be quantified using so-called cophenetic distance measurements (Sokal and Rohlf 1962; Jain and Dubes 1988). The languages in a tree are terminal nodes, which are gathered hierarchically into higher-order nodes. The cophenetic distance between any two terminal nodes within such a tree is defined as the number of nodes needed to go up from language A to the lowest common node shared between the pair, and then down again to language B. In Figure 8.2, the tree distance between English and Dutch would be 4: (i) from English to North Sea, (ii) from North Sea to West (which is the lowest common node), (iii) from West down to Continental and then (iv) from Continental to Dutch.

Intelligibility measurements can be used as a criterion to determine the genetic relationship between language varieties. The tree distances between members of pairs of languages can be correlated straightforwardly with the mean intelligibility scores per language combination, averaging out any asymmetries that may exist within the language pairs AB and BA. Gooskens et al. (2018) used the inherent intelligibility measurements from the MICReLa project (see Appendix B) to determine how well intelligibility measures for the three language families fit proposals made by linguists concerning the closeness of the languages from a historical-linguistic perspective. The cophenetic tree distances were correlated with the data set, which was reduced to encompass only listeners with no schooling or minimal exposure to the target language (inherent intelligibility). As shown in Figure 3.2, certain cells remain empty when filtering for exposure. For instance, since all Germanic listeners have studied English in school, there are no outcomes available for combinations with English as the target language. In these cases, only the results of the English listeners taking the test in the other four languages were used, rather than averaging the distance for the two languages in a language combination. Regarding Czech and Slovak, both groups had to be omitted due to the exposure criterion, resulting in an empty cell.

The Germanic, Romance, and Slavic family trees in Figure 8.2 only include the languages in the MICReLa project. The correlation between the inherent intelligibility scores and the Germanic tree structure distances is $r = -.75$ ($p < .05$), i.e., the smaller the cophenetic distance between two languages, the higher their mutual intelligibility (and vice versa). The Romance tree, structured according to Hall's categorization (Hall 1974), shows a non-significant and lower correlation with intelligibility ($r = -.41$) than the Germanic data, possibly reflecting the substantial disagreement among scholars on the sub-grouping of the Romance language family, see Posner (1996). The genealogic characterization of the Slavic area is straightforward (see, e.g., Sussex and Cubberley 2006), and results in the tree depicted in Figure 8.2. The tree distances exhibit a high correlation with scores from the intelligibility test ($r = -.86$, $p < .01$), confirming the traditional genealogic characterization.

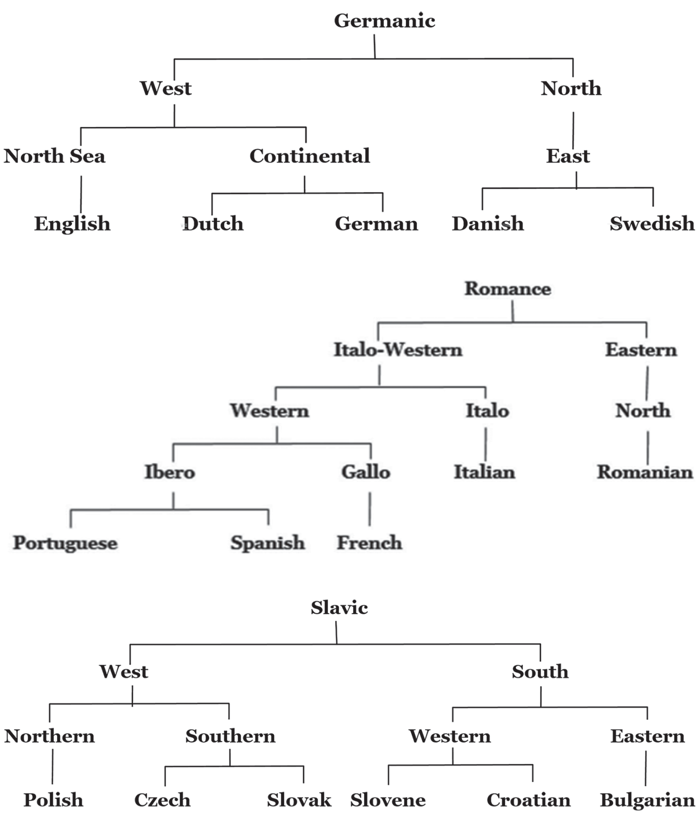

Figure 8.2: The traditional Germanic, Romance, and Slavic language trees showing the languages included in the MICReLa project. Source: Gooskens et al. (2018: 185 and 186).

It is also possible to create language trees on the basis of intelligibility scores, so-called dendrograms. Based on a matrix with intelligibility scores, language varieties can be classified using a hierarchical cluster analysis. The results are binary tree structures (dendrograms) in which the languages are the leaves, and the length of the branches reflects the distance between the leaves (Jain and Dubes 1988). The goal of a cluster analysis is to identify the main groups. The groups are called clusters. Clusters may consist of subclusters, which may, in turn, consist of subsubclusters, and so on. In an investigation by Tang and van Heuven (2008, 2009, 2015), the relationship between functional intelligibility rates and family relationships among Sinitic varieties was established. They generated dendrograms from the intelligibility matrices and compared these with the traditional linguistic taxonomy proposed by Chinese dialectologists. Figure 8.3 present the dendrogram resulting from the functional sentence-level intelligibility scores (SPIN test, see Example 2.7). This language tree corresponds well with the traditional classification by dialectologists. For instance, it shows the primary split in the trees between Mandarin dialects on the one hand and Southern (non-Mandarin) dialects on the other, and it also reflects the substructures of some clusters. The authors proposed to use the results to settle disputes among linguists concerning the characterization of the Taiyuan dialect. Some dialectologists classify this dialect as a (Northern) Mandarin variety, while others classify it as a Southern variety. The intelligibility measurements consistently yielded tree structures in which Taiyuan is grouped with Mandarin varieties.

Feleke, Gooskens, and Rabanus (2020) conducted intelligibility research to classify Ethiosemitic languages. To measure the degree of intelligibility, a word categorization test was adopted from Tang and van Heuven (2009). In addition, structural distances were determined by computing the lexical and phonetic differences. The results of cluster analyses on all measurements showed that the selected South Ethiosemitic languages can be classified into five groups (indicated by braces): {Chaha, Ezha, Gura, Gumer}, {Endegagn, Inor}, {Muher, Mesqan}, {Kistane} and {Silt'e}.

The various investigations discussed in this section thus showed how mutual intelligibility can be used as a unidimensional, experimentally grounded criterion based on communicative principles. This criterion makes it possible to classify related language varieties and establish relationships among them.

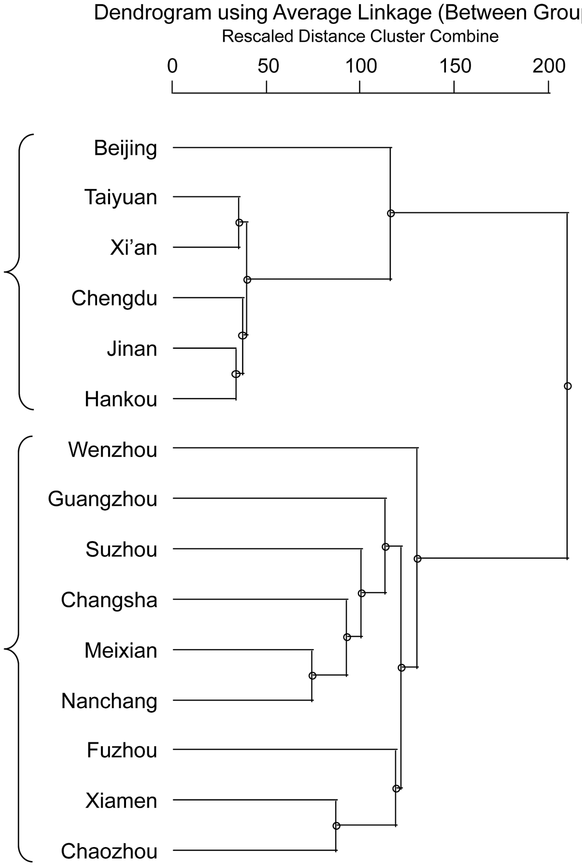

Figure 8.3: Dendrogram based on sentence-level intelligibility scores obtained for 225 combinations of 15 speaker and 15 listener dialects. Mandarin (upper brace) and non-Mandarin (lower brace) dialects are in different main branches of the tree. Source: Tang and van Heuven (2015: 292).

8.5 Improving communication

Language serves as a means of communication, not only between speakers of the same variety but also between people using different accents, dialects or closely related languages. Milliken and Milliken (1996: 17) note that, partly due to possible asymmetric intelligibility between dialects in an area (see Chapter 6), particular dialects in a group are more suitable as centers of communication than others. Simons (1979) showed how groups of dialects can be clustered into optimal communication networks based on intelligibility tests. He notes that such groupings can, for instance, be used to match interpreters with immigrant clients from different dialect backgrounds in high-stakes, time-sensitive environments, such as court, school, and hospital systems.

As mentioned several times throughout this book, in situations where languages are closely related, speakers of different varieties often understand enough of each other's language to be able to communicate while using their own varieties. This kind of communication is often referred to as receptive multilingualism. Communication using receptive multilingualism typically involves closely related languages (inherent receptive multilingualism). It should be noted that this mode of communication involves more than just linguistic overlap between two languages. As discussed in Section 4.6, speakers and listeners can develop strategies to improve communication. For one thing, there is convincing evidence that listeners quickly adapt to non-native speakers (e.g., Cutler 2012), quickly discovering how perceptual categories should be adapted and which atypical features should be disregarded. Additionally, it is well-established that native speakers tend to change their speaking style to accommodate the needs of non-native recipients. Speakers using this so-called "foreigner talk" typically speak at a reduced pace, with greater loudness and expanded pitch range. They also tend to avoid the use of complex grammatical structures and contracted forms, choose easy (i.e., short and frequent) words, or circumscribe less frequent words (e.g., Ferguson 1971; Hatch 1983; Wesche 1994). The listeners also play an important role in reaching mutual understanding. They can signal to the speakers that something is unclear and provide feedback to show that they have understood the speaker.

In circumstances where languages are less closely related, receptive multilingualism can still be employed as long as the parties involved have attained a satisfactory level of passive proficiency in each other's languages (acquired receptive multilingualism, Kluge 2007). For example, an English and a French person may have learned to understand each other's languages at school but have not learned them well enough to speak them. They may then be able to communicate, each speaking their own native language.

Branets, Bahtina, and Verschik (2020) introduced the term mediated receptive multilingualism to describe situations where speakers use third language intervention to communicate. An example is Estonian native speakers who understand and speak Russian. They are able to communicate with Ukrainian speakers by means of receptive multilingualism, since Russian and Ukrainian are closely related. In educational contexts, the EuroCom method (Hufeisen and Marx 2007) centers on achieving comprehension through mediated receptive multilingualism, offering techniques to understand Germanic, Romance, or Slavic languages when the learner is already familiar with another closely related language acting as a bridge language.

One benefit of utilizing receptive multilingualism as a mode of communication is that, for many speakers, it is simpler and more effective to articulate their thoughts and ideas in their mother tongue rather than in English or any other foreign language. In addition, receptive language skills are much broader than productive skills. Language users understand many more words and have much more experience with infrequent or unusual structures than they actively use themselves. Receptive multilingualism's success stems from its receptive resources. The fact that both participants in a conversation can speak the language they know best, their native language, and both have to make an effort to understand the other language results in inherent fairness and equality between the speakers. Additionally, since language is an essential part of identity, many individuals place great importance on using their native language when interacting with others. The motivation for communicating using this mode may not necessarily stem from an inability to speak the other language but rather from socio-political factors that emphasize the affiliation with a particular cultural or ethnic group. Bilaniuk (2010) demonstrated how Ukrainian and Russian speakers show an attitude of purism and resistance to linguistic accommodation when talking to each other even when they speak both languages. They, therefore, tend to use their own language when communicating.

At the level of the individual language user, participating in receptive multilingualism can be viewed as a means of cultivating extensive communicative proficiency and cognitive linguistic adaptability (Doyé 2005; Melo-Pfeifer 2014). Even though there is some disagreement among scholars about the efficiency of learning a second language by first acquiring passive knowledge (Swain 2000), it seems that receptive multilingualism can also be a stepping stone to an active command of a second language. Once second language learners have achieved a passive command of a language, the step towards actively using it will perhaps be smaller than if they have to produce words and sentences themselves from the beginning (Ringbom 2007).

The use of receptive multilingualism often goes against the natural tendency to accommodate to the speaker (Giles and Ogay 2007), and therefore, for many speakers, it may feel rude or impolite at first to use receptive multilingualism. However, receptive multilingualism has been an important, and often the only, means of communication throughout history. During the late Middle Ages, receptive multilingualism served as a common mode of communication for trade and political consultations between speakers of Low German and Scandinavian languages (Braunmüller 2007) as well as in the Romance language area (Blanche-Benveniste 2008; Carlucci 2020). However, the emergence of nationalism, linguistic standardization, and the subsequent ideal of monolingualism, resulted in a more limited use of this form of communication. Little is known about the extent of receptive multilingualism in other past contexts and we have little information about the number of languages or speakers involved in receptive multilingualism today. However, it is reasonable to assume that receptive multilingualism was often the only possible manner of communication in the past and still is today in situations where the speakers have not acquired any other language than their native language.

Scandinavia provides one of the best-documented examples of communication by means of receptive multilingualism and has received the most attention from linguists (e.g., Haugen 1966; Maurud 1976; Zeevaert 2004; Delsing and Lundin Åkesson 2005; Schüppert 2011). Many people from Denmark, Sweden, and Norway use receptive multilingualism rather than a lingua franca (e.g., English or German) when communicating with individuals from their neighboring countries. For instance, when Danes visit Sweden, they often speak Danish to the Swedes they meet, and the Swedes respond in Swedish. Nordic meetings generally take place in this mode, too (Börestam Uhlmann 1994). Receptive multilingualism in Scandinavia has been subject to official language planning and has become a fundamental part of inter-Nordic identity and cooperation. There is a special board, Nordplus, dealing with quality and innovation in the educational systems to strengthen the language comprehension between the Nordic languages, especially among children and youth.[26] However, in daily practice, inter-Scandinavian communication often fails, and many young people tend to prefer English to receptive multilingualism as a means of communication (Skjold Frøshaug and Stende 2021). Some research has been carried out on receptive multilingualism in the rest of the Germanic language area (e.g., Ház 2005; Beerkens 2010) and other Indo-European languages, particularly the Romance language area (Blanche-Benveniste 2008, Conti and Grin 2008) and the Slavic language area (Nábělková 2007; Rehbein and Romaniuk 2014). Other regions of the world have not received as much scholarly attention regarding receptive multilingualism. For a more extensive overview of the occurrence of receptive multilingualism, see ten Thije (2018).

A level of mutual intelligibility sufficient for successful communication does not automatically imply that speakers of the involved languages actually engage in receptive multilingualism. While research has been conducted to determine the extent of mutual intelligibility between specific language pairs, quantitative data regarding the utilization of receptive multilingualism is limited to only a few specific language combinations. For instance, the results of a survey among 252 Dutch and German professionals working in governmental or civil society organizations in the Dutch-German border area (Beerkens 2010) showed that 27% of the

26 https://www.nordplusonline.org

respondents reported using receptive multilingualism in at least some situations at work (e.g., e-mail, face-to-face, telephone, etc.). Receptive multilingualism was used less often than other modes of communication, such as one or both participants speaking Dutch or German as a foreign language or using English as a lingua franca.

The choice to use receptive multilingualism as a mode of communication often depends on the speakers involved and the specific context and domain in which it is used (Beerkens 2010). Receptive multilingualism is commonly used in families where parents have different language backgrounds and between younger and older generations in immigrant families. For instance, the children of Turkish immigrants in Germany may speak German to their parents, who then respond in Turkish (Herkenrath 2012). Such children often have a high level of fluency in two languages, and their language choice can vary depending on various factors, such as the content and context of the conversation and the presence of outsiders who may not understand one of the languages. Receptive multilingualism is also frequently encountered among adult immigrants to the U.S. who do not speak English as a native language but whose children speak fluent English, usually because their education has been conducted in English. The immigrant parents can often understand both their native language and English but only speak their native language to their children, while the children frequently prefer to speak English to their parents. The use of receptive multilingualism may also depend on the language policy of particular institutions, such as educational institutions (Vetter 2012), governmental organizations (Ribbert and ten Thije 2007), the army (Berthele and Wittlin 2013), or the work place (Lüdi 2013).

The European Commission promotes receptive multilingualism as a means to enhance the mobility of European citizens and maintain linguistic diversity (European Commission 2007). It is important to gain more knowledge about when it makes sense to focus on receptive multilingualism and when English as a lingua franca (ELF) should be promoted. Van Mulken and Hendriks (2015, 2017) used a find-the-differences task to compare the effectiveness of communication between native German and Dutch speakers by means of different communication modes (ELF, both speaking either Dutch or German, or receptive multilingualism). They found receptive multilingualism to be most effective while ELF was least effective. Bulatović, Schüppert, and Gooskens (2019) tested how well Slovene native speakers understand Croatian and English as produced by Croatian speakers. Overall, the level of comprehension was found to be high for both languages. The authors conclude that receptive multilingualism and English as a lingua franca are both likely to be successful modes of communication in interactions between Croatian and Slovene speakers.

If we want to introduce receptive multilingualism in a wider context, more knowledge should be gained about the factors that determine how well speakers with different language backgrounds understand each other. In cases where mutual intelligibility is inadequate, it is important to identify factors that are contributing to the communication difficulties. Based on this knowledge, advice can be given for improving mutual intelligibility in a receptive multilingualism situation. For instance, we saw in Section 5.4.1 that the Danish language is spoken faster and with more reduction than Swedish, and this may at least partly explain why Swedes have more difficulties understanding Danes than vice versa. Danes could be advised to speak more slowly, insert pauses at prosodic boundaries, and pronounce words more carefully when communicating with a Swede. Materials have been developed for inhabitants of the Nordic countries with advice on how to communicate (e.g., Grünbaum and Reuter 2013). This includes advice for the speaker to speak slowly, articulate well, avoid unnecessary phrases and words that can be expected to be unknown to the listener and explain them where necessary, and for the listener not to pay too much attention to unknown words but rather focus on the overall message and to ask for an explanation when needed.

It is also possible to set up teaching programs to introduce receptive multilingualism and help students overcome difficulties. Often minimal training is needed to improve intelligibility to such an extent that participants can communicate by means of receptive multilingualism. As mentioned in Section 4.1, a short course of a few hours focusing on the differences and similarities between the native and target language has resulted in a considerable improvement in intelligibility among Dutch students learning Swedish (Hedquist 1985) and among Czech students learning Croatian (Golubović 2016). Rehbein, ten Thije, and Verschik (2012: 252) refer to claims that within the family of Romance languages, sufficient comprehension could be established with 30 to 50 hours of explicit language training (Blanche-Benveniste 2008; Conti and Grin 2008; Janin 2008).

In various parts of Europe, educational programs have already been developed to teach receptive multilingualism, often in the written modality (e.g., Dylan;[27] EuroCom;[28] Europa IC;[29] GalaNet;[30] GalaPro;[31] Linee;[32] Magicc;[33] Miriadi[34]). They

27 http://www.dylan-project.org/

28 http://www.eurocomprehension.eu/; Hufeisen and Marx 2007; Klein and Stegmann 2000

29 https://www.europaic.eu/

30 https://aulaintercultural.org/

31 https://www.ua.pt/pt/cidtff/page/12042

32 https://cordis.europa.eu/project/id/28388

33 https://www.unil.ch/magicc/home.html

34 https://www.miriadi.net/; Benavente Ferrera et al. (2022)

teach learners to reflect on language differences and to develop strategic and metacognitive knowledge and competencies by raising their awareness of comprehension processes (Bonvino, Fiorenza, and Velásquez Diego 2018). Marx (2012) and ten Thije (2018) discuss some of these projects in more detail. However, they also note that more research needs to be conducted to investigate the effects and implementation of these programs. Didactic material can be improved if it is built on scientific findings, and more knowledge may also contribute to higher awareness among speakers of the possibility of using their own native language to communicate with speakers of closely related languages. To establish the potential for improved communication it is important to carry out intervention studies that test the effect of teaching material. Experimental studies should test the intelligibility models and establish whether it makes it easier to overcome language barriers if teaching programs focus on linguistic and extra-linguistic factors that have been shown to correlate best with intelligibility measures. Receptive multilingualism may form a stepping stone to productive skills in the target language, and therefore, intelligibility research is also valuable for language acquisition research in general.

8.6 Conclusion

This chapter has made clear that there are many practical and theoretical motivations for conducting intelligibility research. Accordingly, intelligibility research involves researchers and practitioners from many different backgrounds. The various purposes of an intelligibility investigation are linked to the many different approaches used to quantify intelligibility, as well as to the linguistic and extra-linguistic determinants of intelligibility discussed in this book.

Chapter 9
Conclusions and next steps

9.1 Summary

This book has provided an overview of research on mutual intelligibility between closely related language varieties. It has discussed various methods for quantifying intelligibility as well as linguistic and extra-linguistic determinants of intelligibility. From this overview, it became clear that the choice of method for measuring the level of intelligibility depends on the purpose of the investigation and many practical factors. At the same time, the overview has shown that many different linguistic and extra-linguistic factors determine how well speakers of one language understand another related language and that all these factors often intercorrelate.

As far as extra-linguistic factors are concerned, the amount of exposure is a very important determinant of intelligibility. Several investigations found high correlations between intelligibility scores and the amount of exposure to the target language. Consequently, an important recommendation to language teachers, language policy makers and language users who want to improve receptive language skills is to create situations where the language learners are exposed to the target language. These situations can be created through personal contacts, school exchange programs, or the media. The studies discussed in this book regarding the relationship between exposure and intelligibility also underscore the importance of exposure for strengthening the position of lesser-used languages. Increased exposure to a language enhances comprehension, thereby expanding opportunities for speakers of smaller languages to use their native tongue across a wider range of contexts. However, it should be noted that it is difficult to establish cause and effect, and it may sometimes be a vicious cycle. Typically, people have limited exposure to minor languages, leading to poorer comprehension. Consequently, difficulties in understanding can discourage further contact with the language, perpetuating the cycle.

Lexical differences between listener language and target language are the most important linguistic determinant of inherent intelligibility. If listeners recognize too few words in a text or message, there is no way they will be able to understand the message (van Heuven 2008). As long as the listeners recognize the words in a message, the word order is less important. This line of reasoning is confirmed by the high correlations that have been found between inherent intelligibility scores and lexical distances in various investigations. Lexical distances explain most variance in the model based on measurements from the MICReLa project involving Germanic, Romance and Slavic language families. Receptive multilingual-

https://doi.org/10.1515/9783111134697-009

ism is, therefore, most likely to be successful if the listeners recognize a large proportion of the words. This means that receptive multilingualism can be improved if interactants learn non-cognates they do not recognize from their own language. However, even if words are cognates, there is no guarantee that the listeners will recognize them since the pronunciation of target words may be too different from the corresponding words in the language of the listeners. The importance of pronunciation differences in predicting and explaining intelligibility became clear from the research that was presented in Chapter 5, where high correlations were found between intelligibility measures and phonetic distances. Therefore, intelligibility and receptive multilingualism will improve if listeners become aware of sound correspondences between cognates in the target language and their own language. Listeners can become aware of such correspondences by being exposed to the language and by formal teaching programs that focus on the most frequent and regular correspondences between cognates in their own language and the language of the speaker.

By looking at the relationship between intelligibility scores and their determinants we gain a better understanding of the mechanisms behind the intelligibility of a closely related language. As shown in the simple statistical models presented in Chapter 7, the level of intelligibility can, to a large extent, be predicted and explained by a limited number of linguistic and extra-linguistic factors. However, the correlations and the models of intelligibility still leave room for improvement and, as it became clear from the discussions of linguistic and extra-linguistic determinants of intelligibility, there are many factors that play a role in determining intelligibility. Continued research will advance our understanding of the processes involved in understanding a closely related language. The following sections will identify suggestions for future research and gaps in our knowledge.

9.2 Improving measurements of intelligibility

Many different methods for measuring the level of intelligibility of closely related languages have been used for various purposes. Some of these methods have been developed within other disciplines, such as speech technology, second language acquisition, and speech pathology. New technical advancements are contributing to ongoing developments in all these areas. It is important to keep an eye on such developments, as these may also contribute to research in the area of mutual intelligibility of closely related languages. Some of these new techniques may be easier to use, while providing us with more detailed and refined measurements. They may also solve some of the disadvantages and weaknesses of traditional methods, as discussed in Chapter 2.

It is important to note that different methods may lead to different results. For example, the percentage of correct translations cannot be compared directly to the percentage of correct answers to open questions about a text. Even if the same method is used, the results may not be comparable if variables such as test material, background of the listeners, and quality of recordings are different (see Section 2.1). Therefore, using different methods for each new investigation makes it difficult to put results into perspective by comparing them to the results of other investigations. For example, Delsing (2007) compared the results of an investigation of inter-Scandinavian intelligibility (see Example 2.13) to results from a previous investigation by Maurud (1976) and concluded that comprehension of neighboring languages seems to have deteriorated during the past 30 years. However, this conclusion should be drawn cautiously since the methods and materials used in the two investigations are different.

Therefore, it would significantly advance research on mutual intelligibility between closely related languages if one uniform method for measuring intelligibility were developed, preferably with user-friendly software where speech material from new languages can easily be added. A step in this direction was taken in the MICReLa project. The word recognition task from the MICReLa project was later used with translations of the same words to investigate the mutual intelligibility between Finnish and Estonian (Härmävaara and Gooskens 2019, see Section 6.2.2). Finnish listeners could recognize 61.3% of the Estonian words, and the Estonian listeners could recognize 44.6% of the Finnish words. In isolation, the intelligibility results from the Finnish-Estonian investigation were not very informative about the level of mutual intelligibility between the two languages. However, by comparing these results to the results for 70 European language combinations with participants who had the same background as far as age and education are concerned, the results could be put into perspective. The comparison to the MICReLa results showed that Estonian participants understand Finnish rather well in comparison to Europeans from the Germanic, Romance, and Slavic language families. Only 21% of the results involving the 70 MICReLa language combinations were higher than the Estonian listeners' results. However, the results of the Finnish listeners were rather low compared to the MICReLa results. Here, 59% of the MICReLa listeners performed better when listening to a closely related language. The intelligibility results of the Estonian listeners were closest to the results of Slovak participants tested in Polish. The results of the Finnish listeners were most similar to those of Romanian participants tested in Italian.

The word recognition task used for the MICReLa project could relatively easily be used for the Finnish-Estonian project. However, to develop a uniform intelligibility test that can be used cross-culturally and with different kinds of listener groups is not straightforward, as becomes clear from the considerations in Chapter 2. For example, some listeners may be illiterate, some texts may not be suitable for all listeners, or words that are frequent in one language may be less frequent in another. A uniform method for collecting data from many different languages and dialects from worldwide would be a powerful tool for researchers and language policymakers.

Most of the methods discussed in Chapter 2 measure intelligibility in one direction, e.g., the intelligibility of language A among speakers of language B. In addition, knowledge about mutual intelligibility has primarily been collected from controlled laboratory conditions and tasks. Such measurements are generally low in ecological validity. Much of our experience with closely related languages comes from real-life communicative situations. Further research in more realistic situations using conversational tasks may advance our understanding of cognitive processes during receptive multilingualism. During a conversation in the receptive multilingualism mode, listeners speak one language while listening to another, which takes considerable effort. Little is known about the cognitive processes involved in this kind of language switching (Declerck and Phillipp 2015). Some languages have a stronger representation in the mental lexicon than others, and several languages can be activated in parallel (see, e.g., Szubko-Sitarek 2011). They influence each other during language processing, and this is especially relevant in the context of receptive multilingualism. Psychological experiments using techniques to measure processing effort, such as eye tracking and EEG measurements, may add significantly to our understanding of processes involved in receptive multilingualism.

Often, researchers are primarily interested in determining inherent intelligibility, see Section 4.7. However, in reality, inherent intelligibility is almost a theoretical construct since most listeners have been exposed to the language of the speaker or some related language to some extent. In addition, as the overview of methods for measuring intelligibility by means of functional tests in Chapter 2 made clear, it is not a straightforward task to quantify the level of intelligibility between two languages. Preparing and administering experiments and the wide varieties of tests, test material, and test situations is often labor-intensive. Also, the variation in the personal backgrounds of listeners involved in intelligibility research makes it hard to compare levels of intelligibility between different language pairs. Problems related to functional testing may be circumvented by measuring objective linguistic distances that have been shown to correlate with intelligibility scores. New technologies within speech recognition based on deep neural networking techniques may prompt new objective approaches to measuring inherent intelligibility. For example, it would be interesting to test whether Swedish and Danish speech recognition systems would understand the neighboring languages equally well or whether there would be the same asymmetric intelligibility as repeatedly found in previous research, with Swedes having more difficulties understanding Danish than vice versa. If an asymmetry is found in automatic speech recognition, this may confirm the evidence that asymmetric inherent intelligibility has a linguistic basis.

9.3 Improving measurements of determinants to better reflect intelligibility

In this book, we have focused on measurements of determinants that have been shown to correlate with intelligibility. As mentioned above, the extra-linguistic factor that predicts general intelligibility best is exposure. However, as became clear from Chapter 4, many extra-linguistic factors play a role and some interact with each other and with linguistic factors. Advanced methods have recently been developed to measure some of these extra-linguistic factors in more refined ways within psycholinguistics. For example, advances have been made to elicit implicit attitudes that are difficult to capture with traditional methods. Attitudes have mostly been found to show rather low correlations with intelligibility, even though intuition expects attitudes to be important for the level of intelligibility. Positive attitudes may motivate the listeners to make an effort to understand the speaker, and negative attitudes may have the opposite effect. Future research is likely to find measures of attitudes that better reflect the kind of attitudes that matter for intelligibility.

Even if great care is taken to select a homogenous group of listeners for an intelligibility experiment, there may still be large differences between the performances of individual listeners. In Section 4.5, some individual characteristics were mentioned that may determine how well listeners can understand a closely related language. More research along these lines with input from psychology is likely to improve our understanding of the processing of such languages. The correlational analyses presented throughout the book mostly focused on predicting and explaining the mutual intelligibility of whole languages. However, to gain a greater understanding of processes involved in understanding a closely related language, it is necessary to correlate measurements at the level of the listener rather than working with the means across language combinations. It may be necessary to use more advanced statistical analyses, such as mixed modeling, to analyze such data sets.

As far as the linguistic determinants of intelligibility are concerned, many of the measurements introduced in this book were initially developed for measuring distances between dialects to characterize dialect areas and draw dialect maps. In Chapter 5, research was presented that has used such measurements to predict and explain the level of mutual intelligibility between languages and dialects. Especially lexical distances and phonetic distance measurements have been successful determinants of inherent intelligibility measures. Measurements of syntactic and morphological distances have received less attention from dialectometrists and the correlations with intelligibility are lower. There may be room for improvement of measurements at these linguistic levels, making them suitable to better predict and explain intelligibility.

Even when combining measurements at different linguistic levels in a statistical model of inherent intelligibility, there is still a large amount of unexplained variance. This means that there is room for improvement of communicatively relevant linguistic distance measures. In Sections 5.1.1 and 5.2.1, a number of improvements were discussed that may better reflect communicatively relevant linguistic distances.

It remains to be investigated how such improvements can be incorporated into distance measurements. For example, it seems important to gain more knowledge of listeners intuitions about correspondences between sounds in the target language and their own native language. Based on such intuitions, communicatively relevant weights can be added to algorithms that measure phonetic distance, such as the Levenshtein algorithm. Intuitions may be universal to some extent, but they are also likely to depend on the native language of the listeners. To advance our knowledge of the processes in the mind of the listener, when listening to a closely related language, it may, therefore, be advisable to first focus on a particular language combination before pursuing the goal of developing a universal measure of communicatively relevant linguistic distances.

To be better able to pinpoint individual features of a language that can predict the intelligibility of closely related languages, knowledge about word recognition and language processing from work by psycholinguists should be implemented. An experimental approach may advance our understanding of the weights that should be given to various linguistic differences between target language and listener language. For example, it is still unclear how disturbing differences in word order are for the listener. Experiments could be developed that systematically change the word order in sentences while keeping all other linguistic levels constant. By comparing the intelligibility of manipulated sentences to sentences with native word order, we can draw conclusions about the importance of word order for mutual intelligibility in particular language combinations. Similarly, we may gain more knowledge about the weighing of various other characteristics by experimental testing within linguistic level.

There may be a point beyond which the model cannot be further enhanced. It may be unrealistic to implement all relevant linguistic differences between the target language and listener language into a model of mutual intelligibility. Various factors interact, and the combination of factors plays a role. In addition, certain factors may only apply to a restricted set of words. As a result, listeners might need to employ different strategies for each word to find its equivalent in their own language.

9.4 Conclusions

To conclude, the research discussed in this book has shown that we have come a long way in developing methods for research on mutual intelligibility between closely related languages. In particular, interdisciplinary research with insights from different disciplines has been fruitful. At the same time, new techniques and technology have advanced intelligibility research, allowing us to gain more knowledge about the limits and possibilities for mutual understanding between languages worldwide.

As discussed in Chapter 8, the results of intelligibility investigations have served various theoretical and practical purposes in the past. Future investigations will likely provide important input for further advancements and developments in various areas within and outside of linguistics. If we collect sufficient intelligibility data, we may come closer to the point where we can define a threshold below which there is an intelligibility breakdown. Such a threshold would make it easier to define when a language variety should be considered a language and when a dialect from a linguistic perspective. This is an important question for people fighting for language rights and to protect minority languages and lesser-used languages.

Appendices

Appendix A: The four texts used for the cloze test in Chapter 3 and for linguistic distance measurements in Chapters 5 and 7. The words that were removed from the texts in the cloze test are underlined

Child athletes

Parents whose children show a special interest in a particular sport have a difficult decision to make. Should they allow their children to train to <u>become</u> top sportsmen and women? For many <u>children</u> this means starting very young. School work, going out with friends and other interests have to take <u>second</u> place. It's very <u>difficult</u> to explain to young children why they have to train for five hours a day. That includes even the weekend, when most of their friends are <u>playing</u>. Another problem is of course money. In many <u>countries</u> money for training is available from the government for the very best young athletes. If this help cannot be given the parents have to find the time and <u>money</u> to support their children. Sports clothes, transport to competitions, special equipment et cetera, can all be very <u>expensive</u>. Many parents are understandably worried that it's dangerous to start serious training in a sport at an early age. Some doctors agree that <u>young</u> muscles may be damaged by training before they are properly developed. Trainers, however, <u>believe</u> that you can only reach the top as a sports person when you start young. What is clear is that very few people do <u>reach</u> the top. So both parents and children should be prepared for failure. It happens even after many years of <u>training</u>.

Catching a cold

Hello, my name is Christina and I give advice to people with questions about their health. I get a lot of <u>letters</u> at this time of year. People <u>complain</u> that they have a cold which won't go away. There are so many different stories about how to <u>prevent</u> or cure a cold. So it's often difficult to <u>know</u> what to do. Colds are rarely dangerous, except for people who are already <u>weak</u>, such as the elderly or young babies. Still, colds are always <u>uncomfortable</u> and usually most (unpleasant). Of course you can buy lots of <u>medicines</u> which will help to make your cold less unpleasant. But remember that nothing can actually <u>cure</u> a cold or make it go away faster. Another thing is that any medicine which is strong enough to make you feel better

https://doi.org/10.1515/9783111134697-010

could be dangerous. If you are already taking medicine for other illnesses always check with your doctor if that's all right. And remember that it could happen that they might make you sleepy. Please don't try to drive if they do! Lastly, there are no magic foods or drinks. The best answer is to keep strong and healthy. You'll have less chance of catching a cold, and if you do, it shouldn't be so bad!

Driving in winter

Winter is dangerous because it's so difficult to know what is going to happen. Accidents take place so easily. Fog can be waiting to meet you over the top of a hill. Ice might be hiding beneath the melting snow, waiting to send you off the road. The car coming towards you may suddenly slide across the road. Rule number one for driving on icy roads is to drive smoothly. Uneven movements can make a car suddenly very difficult to control. Every time you turn the wheel, brake or increase speed, you must be gentle and as slow as possible. Imagine you are driving with a full cup of hot coffee on the seat next to you. Drive so that you wouldn't spill it. Rule Number Two is to pay attention to what might happen. The more ice there is, the further down the road you have to look. Test how long it takes to stop by gently braking. Remember that you may be driving more quickly than you think. In general, allow double your normal stopping distance when the road is wet. Use three times this distance on snow, and even more on ice. Try to stay in control of your car at all times and you will avoid trouble.

Riding a bike

Getting enough exercise is part of a healthy lifestyle. Along with jogging and swimming, riding a bike is one of the best all-round forms of exercise. It can help to increase your strength and energy. Also it gives you more efficient muscles and a stronger heart. But increasing your strength is not the only advantage of riding a bike. You're not carrying the weight of your body on your feet. That's why riding a bike is a good form of exercise for people with painful feet or backs. However, with all forms of exercise it's important to start slowly and build up gently. Doing too much too quickly can damage muscles that aren't used to working. If you have any doubts about taking up riding a bike for health reasons, talk to your doctor. Ideally you should be riding a bike at least two or three times a week. For the exercise to be doing you good, you should get a little out of breath. Don't worry that if you begin to lose your breath, it could be dangerous. This is simply not true. Shortness of breath shows that the exercise is having the right effect. However, if you find you are in pain then you should stop and take a rest. After a while it will get easier.

Appendix B: General and inherent intelligibility results (% correct) from the spoken cloze test in the MICReLa project. In parentheses the results for listeners with minimal exposure (inherent intelligibility). Scores indicated in bold are significantly different (asymmetrical) within a language pair at the .01 level (Bonferroni test). For abbreviations see Figure 3.1

Table B.1: Results of cloze tests for the Germanic language family.

	Target language					
Listener language	**DA**	**DU**	**EN**	**GE**	**SW**	**Total**
DA	–	13.3	**92.1**	**47.8**	56.7	52.5
	–	(13.1)	–	–	(43.8)	(34.7)
DU	10.5	–	**94.0**	**75.0**	10.4	47.5
	(9.9)	–	–	–	(10.4)	(10.2)
EN	**7.9**	**10.3**	–	**27.7**	**8.3**	13.6
	(7.9)	(9.7)	–	(9.5)	(8.7)	(8.9)
GE	**16.7**	**31.1**	**85.7**	–	**10.0**	35.9
	(12.5)	(25.5)	–	–	(10.0)	(16.0)
SW	62.5	13.0	**89.6**	**37.0**	–	50.5
	(56.0)	(13.0)	–	(15.3)	–	(29.2)
Total	24.4	16.9	90.4	46.9	21.4	40.0
	(23.0)	(15.4)	–	(11.3)	(21.3)	(24.7)

Table B.2: Results of cloze tests for the Romance language family.

	Target language					
Listener language	**FR**	**IT**	**PT**	**RO**	**SP**	**Total**
FR	– –	**24.2** (22.2)	23.5 18.2	**11.0** 10.7	31.5 11.5	22.6 (22.9)
IT	**46.3** (18.8)	– –	**33.5** **(22.4)**	**10.6** **(8.7)**	**65.7** **(56.5)**	36.6 (29.4)
PT	34.3 –	**49.4** **(43.8)**	– –	14.7 (14.0)	**77.4** **(54.8)**	47.2 (40.3)
RO	**47.1** –	**57.7** **(47.2)**	22.9 (20.7)	– –	**54.0** **(46.6)**	44.9 (38.2)
SP	28.2 –	**45.7** **(38.2)**	**37.2** **(35.7)**	**13.6** **(13.7)**	– –	32.2 (29.2)
Total	39.0 (18.6)	44.3 (38.1)	29.3 (26.6)	12.5 (12.4)	57.2 (54.9)	36.7 (32.0)

Table B.3: Results of cloze tests for the Slavic language family.

	Target language						
Listener language	**BU**	**CR**	**CZ**	**PO**	**SK**	**SL**	**Total**
BU	– –	29.2 (29.2)	10.7 (10.9)	7.1 (6.9)	16.0 (16.0)	20.6 (19.9)	16.7 (16.7)
CR	19.7 (19.7)	– –	18.1 (18.1)	9.5 (9.5)	23.0 (23.0)	**43.7** **(40.2)**	22.8 (22.3)
CZ	13.5 (13.5)	19.4 (19.9)	– –	35.4 (33.3)	92.7 (87.5)	15.7 (16.7)	35.3 (34.4)
PO	13.7 (13.8)	14.4 (14.8)	26.6 (24.2)	– –	40.7 (41.7)	13.4 (13.4)	21.8 (21.3)
SK	10.1 (10.1)	25.9 (24.5)	95.0 –	50.7 (46.3)	– –	15.1 (16.0)	39.4 (24.8)
SL	18.0 (18.6)	**79.4** **(66.7)**	18.0 (17.8)	12.8 (11.9)	18.8 (18.8)	– –	29.4 (28.0)
Total	15.0 (15.1)	33.6 (32.0)	33.7 (17.8)	23.1 (22.4)	38.2 (37.2)	21.7 (21.5)	27.6 (24.6)

Appendix C: Extra-linguistic measurements from the MICReLa project. For abbreviations see Figure 3.1

Table C.1: Mean exposure scores across six scales from 1 ("never") to 5 ("every day") for the Germanic language family.

	Target language					
Listener language	**DA**	**DU**	**EN**	**GE**	**SW**	**Total**
DA	–	1.17	4.07	1.76	2.16	2.29
DU	1.19	–	3.80	1.95	1.24	2.05
EN	1.29	1.47	–	1.70	1.14	1.40
GE	1.27	1.40	3.15	–	1.08	1.73
SW	1.96	1.22	3.70	1.73	–	2.15
Total	1.43	1.32	3.68	1.79	1.40	1.92

Table C.2: Mean exposure scores for the Romance language family.

	Target language					
Listener language	**FR**	**IT**	**PT**	**RO**	**SP**	**Total**
FR	–	1.63	1.36	1.11	1.89	1.50
IT	2.05	–	1.60	1.23	2.05	1.73
PT	1.82	1.58	–	1.23	2.16	1.70
RO	1.68	2.09	1.32	–	1.80	1.72
SP	1.50	1.90	1.44	1.25	–	1.52
Total	1.76	1.80	1.43	1.21	1.98	1.63

Table C.3: Mean exposure scores for the Slavic language family.

	Target language						
Listener language	**BU**	**CR**	**CZ**	**PO**	**SK**	**SL**	**Total**
BU	–	1.25	1.46	1.32	1.13	1.40	1.22
CR	1.10	–	1.28	1.18	1.08	1.57	1.16
CZ	1.06	1.22	–	1.45	2.78	1.18	1.63
PO	1.18	1.25	1.48	–	1.40	1.08	1.33
SK	1.08	1.42	3.60	1.71	–	1.21	1.95
SL	1.16	2.32	1.35	1.33	1.20	–	1.47
Total	1.11	1.29	1.88	1.41	1.60	1.29	1.46

Table C.4: Attitude scores on a scale from 1 ("ugly") to 5 ("beautiful") for the Germanic language family.

	Target language					
Listener language	**DA**	**DU**	**EN**	**GE**	**SW**	**Total**
DA	–	1.90	3.94	2.30	3.80	2.99
DU	2.76	–	3.39	2.88	3.77	3.20
EN	2.50	2.67	–	2.93	2.74	2.71
GE	2.50	2.93	4.14	–	3.05	3.22
SW	2.50	2.17	3.75	2.91	–	2.83
Total	2.57	2.42	3.87	2.76	3.34	2.99

Table C.5: Attitude scores for the Romance language family.

	Target language					
Listener language	**FR**	**IT**	**PT**	**RO**	**SP**	**Total**
FR	–	4.19	2.79	2.64	3.97	3.40
IT	4.15	–	3.65	2.22	4.15	3.54
PT	3.50	4.13	–	1.83	3.58	3.26
RO	3.88	3.96	3.41	–	3.76	3.75
SP	3.36	4.23	3.38	2.39	–	3.34
Total	3.72	4.13	3.31	2.27	3.87	3.46

Table C.6: Attitude scores for the Slavic language family.

	Target language						
Listener language	**BU**	**CR**	**CZ**	**PO**	**SK**	**SL**	**Total**
BU	–	2.93	2.94	1.93	2.67	3.10	2.62
CR	2.28	–	2.77	2.77	2.67	3.03	2.52
CZ	2.06	2.98	–	3.21	4.38	2.42	3.16
PO	2.19	2.38	3.16	–	3.16	2.50	2.72
SK	2.29	3.29	3.98	3.27	–	2.44	3.21
SL	2.42	3.46	3.04	3.20	3.10	–	3.04
Total	2.21	2.90	3.21	2.80	3.12	2.70	2.85

Table C.7: Number of years of learning the target language for the Germanic language family.

	Target language					
Listener language	**DA**	**DU**	**EN**	**GE**	**SW**	**Total**
DA	–	.00	10.83	4.80	.00	3.91
DU	.00	–	7.51	3.97	.00	2.87
EN	.00	.00	–	2.18	.00	0.55
GE	.00	.00	9.11	–	.00	2.28
SW	.00	.00	11.25	2.50	–	3.44
Total	.00	.00	9.68	3.36	.00	2.61

Table C.8: Number of years of learning the target language for the Romance language family.

	Target language					
Listener language	**FR**	**IT**	**PT**	**RO**	**SP**	**Total**
FR	–	.62	.18	.00	1.50	.58
IT	1.95	–	.23	.00	.71	.72
PT	3.00	.10	–	.00	.75	.96
RO	2.29	.24	.14	–	.44	.78
SP	2.56	.32	.04	.04	–	.74
Total	2.45	.32	.15	.01	.85	.76

No data are presented for the Slavic language family, since in the Slavic language area no listeners learned any of the target languages at school.

Appendix D: Linguistic measurements from the MICReLa project. For abbreviations see Figure 3.1

Table D.1: Lexical distances for the Germanic language family.

	Target language					
Listener language	**DA**	**DU**	**EN**	**GE**	**SW**	**Total**
DA	–	43.94	49.17	38.95	5.82	34.47
DU	45.60	–	39.02	21.52	47.83	38.49
EN	48.54	38.07	–	44.47	49.48	45.14
GE	38.97	19.80	46.41	–	40.35	36.38
SW	4.63	46.75	51.38	39.88	–	35.66
Total	34.44	37.14	46.50	36.21	35.87	38.03

Table D.2: Lexical distances for the Romance language family.

	Target language					
Listener language	**FR**	**IT**	**PT**	**RO**	**SP**	**Total**
FR	–	21.63	17.76	49.01	19.96	27.09
IT	24.39	–	11.53	54.00	14.44	26.09
PT	21.02	21.25	–	51.95	6.02	25.06
RO	57.81	52.14	53.90	–	54.03	54.47
SP	24.30	10.40	1.75	45.91	–	20.59
Total	31.88	26.36	21.24	50.22	23.61	30.66

Table D.3: Lexical distances for the Slavic language family.

	Target language						
Listener language	**BU**	**CR**	**CZ**	**PO**	**SK**	**SL**	**Total**
BU	–	21.32	39.53	44.64	39.03	30.42	36.13
CR	23.88	–	33.10	44.27	31.15	22.91	33.10
CZ	38.78	39.18	–	25.44	5.99	45.17	27.35
PO	37.62	43.91	17.67	–	23.43	46.85	30.66
SK	39.46	37.96	10.20	23.95	–	40.68	27.89
SL	36.30	25.80	36.84	47.87	36.44	–	36.65
Total	34.94	35.59	25.13	34.58	24.90	37.21	31.03

Table D.4: Phonetic distances for the Germanic language family.

	Target language					
Listener language	**DA**	**DU**	**EN**	**GE**	**SW**	**Total**
DA	–	39.65	41.54	40.19	34.45	38.96
DU	39.65	–	33.85	29.56	37.20	35.07
EN	41.54	33.85	–	37.64	42.41	38.86
GE	40.19	29.56	37.64	–	37.36	36.19
SW	34.45	37.20	42.41	37.36	–	37.86
Total	38.96	35.07	38.86	36.19	37.86	37.39

Table D.5: Phonetic distances for the Romance language family.

	Target language					
Listener language	**FR**	**IT**	**PT**	**RO**	**SP**	**Total**
FR	–	40.94	43.27	44.21	38.58	41.75
IT	40.94	–	41.23	39.19	33.28	38.66
PT	43.27	41.23	–	42.83	33.88	40.30
RO	44.21	39.19	42.83	–	36.99	40.81
SP	38.58	33.28	33.88	36.99	–	35.68
Total	41.75	38.66	40.30	40.81	35.68	39.44

Table D.6: Phonetic distances for the Slavic language family.

	Target language						
Listener language	**BU**	**CR**	**CZ**	**PO**	**SK**	**SL**	**Total**
BU	–	31.90	40.22	38.84	42.29	33.45	38.31
CR	31.90	–	28.14	35.02	30.15	29.75	31.30
CZ	40.22	28.14	–	30.85	11.48	37.11	27.67
PO	38.84	35.02	30.85	–	31.52	39.29	34.06
SK	42.29	30.15	11.48	31.52	–	36.57	28.86
SL	33.45	29.75	37.11	39.29	36.57	–	35.23
Total	38.31	31.30	27.67	34.06	28.86	35.23	32.04

Table D.7: Syntactic distances for the Germanic language family.

	Target language					
Listener language	**DA**	**DU**	**EN**	**GE**	**SW**	**Total**
DA	–	36.80	19.10	40.50	14.65	27.76
DU	36.80	–	34.41	23.77	36.87	32.96
EN	19.10	34.41	–	43.64	22.16	29.96
GE	40.50	23.77	43.64	–	40.65	37.14
SW	14.65	36.87	22.16	40.65	–	28.58
Total	27.76	32.96	29.83	37.14	28.58	31.26

Table D.8: Syntactic distances for the Romance language family.

	Target language					
Listener language	**FR**	**IT**	**PT**	**RO**	**SP**	**Total**
FR	–	21.17	23.26	31.65	16.64	23.18
IT	21.17	–	13.81	26.50	11.13	18.15
PT	23.26	13.81	–	20.79	14.22	18.02
RO	31.65	26.50	20.79	–	26.87	26.45
SP	16.64	11.13	14.22	26.87	–	17.22
Total	23.18	18.15	18.02	26.45	17.22	20.60

Table D.9: Syntactic distances for the Slavic language family.

	Target language						
Listener language	**BU**	**CR**	**CZ**	**PO**	**SK**	**SL**	**Total**
BU	–	41.12	40.10	36.71	41.99	38.66	39.98
CR	41.12	–	29.11	23.99	25.04	29.49	29.82
CZ	40.10	29.11	–	21.70	10.92	27.84	25.46
PO	36.71	23.99	21.70	–	23.14	24.41	26.39
SK	41.99	25.04	10.92	23.14	–	31.03	25.27
SL	38.66	29.49	27.84	24.41	31.03	–	30.29
Total	39.98	29.82	25.46	26.39	25.27	30.29	29.38

References

Abraham, Roberta G. & Carol A. Chapelle. 1992. The meaning of cloze test scores: An item difficulty perspective. *The Modern Language Journal* 76(4). 468–479.

Abunasser, Mahmoud A. K. 2015. *Computational measures of linguistic variation: A study of Arabic varieties*. Urbana-Champaign: University of Illinois at Urbana-Champaign dissertation.

Anderson, Anne H., Miles Bader, Ellen G. Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson & Regina Weinert. 1991. The hcrc map task corpus. *Language and Speech* 34(4). 351–366.

Anderson, Stephen R. 2012. *Languages: A very short introduction*. Oxford University Press.

Anwyl-Irvine, Alex, Edwin Dalmaijer, Nick Hodges & Jo K. Evershed. 2021. Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods* 53. 1407–1425.

Ashby, Michael & John Maidment. 2005. *Introducing phonetic science*. Cambridge: Cambridge University Press.

Bahtina, Daria & Jan D. ten Thije. 2012. Receptive multilingualism. In Carol A. Chapelle (ed.), *The encyclopedia of applied linguistics*, 1–6. Wiley Online Library.

Bahtina, Daria, Jan D. ten Thije & Frank Wijnen. 2013. Combining cognitive and interactive approaches to lingua receptiva. *International Journal of Multilingualism* 10(2). 159–180.

Bailyn, John F. 2010. To what degree are Croatian and Serbian the same language? Evidence from a translation study. *Journal of Slavic Linguistics* 18(2). 181–219.

Baker, Rachel & Valerie Hazan. 2011 DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods* 43, 761–770.

Barcroft, Joe. 2004. Second language vocabulary acquisition: A lexical input processing approach. *Foreign Language Annals* 37(2). 200–208.

Bardel, Camilla & Ylva Falk. 2007. The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research* 23(4). 459–484.

Beerkens, Roos. 2010. *Receptive multilingualism as a language mode in the Dutch-German border area*. Münster: Waxmann.

Beijering, Karin, Charlotte Gooskens & Wilbert Heeringa. 2008. Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm. In Marjo van Koppen & Bert Botma (eds.), *Linguistics in the Netherlands*, 13–24. Amsterdam: John Benjamins.

Benavente Ferrera, Susana, Francisco Calvo del Olmo, Erika H. Frisan, Veronica Manole, Karine Marielly, Rocha da Cunha & Hugues Sheeren. 2022. *PanromanIC. Manuale di intercomprensione fra lingue romanze* [PanromanIC. Manual of Intercomprehension among Romance languages]. Bologna: Zanichelli Editore.

Bender, Marvin L. & Robert L. Cooper. 1971. Mutual intelligibility within Sidamo. *Lingua* 27. 32–52.

Benoît, Christian. 1990. An intelligibility test using semantically unpredictable sentences: Towards the quantification of linguistic complexity. *Speech Communication* 9(4). 293–304.

Benoît, Christian, Martine Grice & Valérie Hazan. 1996. The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using semantically unpredictable sentences. *Speech Communication* 18(4). 381–392.

Bent, Tessa, Adam Buchwald & Wesley Alford. 2007. Inter-talker differences in intelligibility for two types of degraded speech. In *Research of spoken language processing progress report no. 28*, 316–330. *Bloomington, IN*: *Speech Research Laboratory, Indiana University*.

https://doi.org/10.1515/9783111134697-011

Berger, Tilman. 2003. Slovaks in Czechia – Czechs in Slovakia. *International Journal of the Sociology of Language* (162). 19–39.
Berthele, Raphael. 2008. Dialekt-Standard Situationen als embryonale Mehrsprachigkeit. Erkenntnisse zum interlingualen Potenzial des Provinzlerdaseins. *Sociolinguistica 22(1)*. 87–107.
Berthele, Raphael. 2011. On abduction in receptive multilingualism. Evidence from cognate guessing tasks. *Applied Linguistics Review* 2(2011). 191–220.
Berthele, Raphael & Gabriele Wittlin. 2013. Receptive multilingualism in the Swiss Army. *International Journal of Multilingualism* 10(2). 181–195.
Best, Catherine T. 1995. A direct realist perspective on cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language research*, 171–204. Timonium, MD: York Press.
Best, Catherine T., Gerald W. McRoberts & Elizabeth Goodell. 2001. Discrimination of nonnative consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109(2). 775–794.
Bezooijen, Reneé van. 1995. Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech* 38(3). 253–265.
Bezooijen, Renée van. 2000. Een toontje hoger of lager. Wat zegt de hoogte van onze stem? [A tone higher or lower. What does the pitch of our voice say?]. *Onze Taal* 69(12). 325–330.
Bezooijen, Renée van. 2002. Aesthetic evaluation of Dutch: Comparison across dialects, accents, and languages. In Daniel D. Long & Dennis R. Preston (eds.), *Handbook of perceptual dialectology: Volume 2*, 13–31. Amsterdam: John Benjamins.
Bezooijen, Renée van & Charlotte Gooskens. 2005a. How easy is it for speakers of Dutch to understand Frisian and Afrikaans, and why? In Jenny Doetjes & Jeroen van de Weijer (eds.), *Linguistics in the Netherlands*, 13–24. Amsterdam: John Benjamins.
Bezooijen, Renée van & Charlotte Gooskens. 2005b. Intertalig tekstbegrip. De begrijpelijkheid van Friese en Afrikaanse teksten voor Nederlandse lezers [Interlingual text comprehension. The comprehensibility of Frisian and African texts for Dutch readers]. *Nederlandse Taalkunde* 10(2). 129–152.
Bezooijen, Renée van & Charlotte Gooskens. 2006. Waarom is geschreven Afrikaans makkelijker voor Nederlandstaligen dan andersom? [Why is written Afrikaans easier for Dutch speakers than the other way around?]. In Tom Koole, Jacomine Nortier & Bert Tahitu (eds.), *Artikelen van de vijfde sociolinguïstische conferentie in Lunteren*, 68–76. Delft: Eburon.
Bezooijen, Renée van & Charlotte Gooskens. 2007. Interlingual text comprehension: Linguistic and extralinguistic determinants. In Jan D. ten Thije & Ludger Zeevaert (eds.), *Receptive Multilingualism and intercultural communication: Linguistic analyses, language policies and didactic concepts*, 249–264. Amsterdam: John Benjamins.
Bezooijen, Renée van & Rob van den Berg. 1999. Word intelligibility of language varieties in the Netherlands and Flanders under minimal conditions. In Renée van Bezooijen & René Kager (eds.), *Linguistics in the Netherlands*, 1–12. Amsterdam: John Benjamins.
Bezooijen, Renée van & Rob van den Berg. 2000. Hoe verstaanbaar is het Fries voor niet-Friestaligen? [How comprehensible is Frisian to non-Frisian speakers?]. In Piter Boersma, Pieter Breuker & Lammert G. Jansma (eds.), *Philologia Frisica anno 1999*, 9–26. Leeuwarden: Fryske Akademy.
Bezooijen, Renée van & Vincent J. van Heuven. 1997. Assessment of speech synthesis. In Dafydd Gibbon, Roger Moore & Richard Winksi (eds.), *Handbook of standards and resources for spoken language systems*, 481–653. Berlin/New York: Mouton de Gruyter.
Biggs, Bruce. 1957. Testing intelligibility among Yuman languages. *International Journal of American Linguistics* 23(2). 57–67.

Bilaniuk, Laada. 2010. Language in the balance: The politics of non-accommodation on bilingual Ukrainian-Russian television shows. *International Journal of the Sociology of Language* 2010(201). 105–133.

Blanche-Benveniste, Claire. 2008. Comment retrouver l'expérience des anciens voyageurs en terres de langues romanes? In Virginie Conti & François Grin (eds.), *S'entendre entre langues voisines: Vers l'intercompréhension*, 33–51. Chêne-Bourg: Georg.

Blees, Gerda J. & Jan D. ten Thije. 2017. Receptive multilingualism and awareness. In Jasone Cenoz, Durk Gorter & Stephen May (eds.), *Language awareness and multilingualism*, Encyclopedia of language and education, 333–345. Springer.

Bleses, Dorthe, Werner Vach, Marlene Slott, Sonja Wehberg, Pia Thomsen, Thomas O. Madsen & Hans Basbøll. 2008. Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language* 35(3). 619–650.

BNC Consortium. 2007. BNC: The British national corpus, version 3, BNC XML edition. http://www.natcorp.ox.ac.uk/XMLedition/ (accessed 7 April 2024).

Bø, Inge. 1978. *Ungdom og naboland. En undersøkelse av skolens og fjernsynets betydning for nabospraksforståelsen* [Youth and neighboring country: An investigation of the influence of school and TV on inter-Scandinavian comprehension]. Stavanger: Rogalandsforskning.

Boets, Herman & Georges de Schutter. 1977. Verstaanbaarheid en appreciatie. Nederlandse dialecten uit België zoals inwoners van Duffel die ervaren [Intelligibility and appreciation. Dutch dialects from Belgium as experienced by inhabitants of Duffel]. *Taal en Tongval* 29. 156–177.

Bohn, Ocke-Schwen & Trine Askjær-Jørgensen. 2017. A comparison of Danish listeners' processing cost in judging the truth value of Norwegian, Swedish, and English sentences. *Proceedings of Interspeech 2017*, 1741–1744.

Bonvino, Elisabetta, Elisa Fiorenza & Cortés Velásquez Diego. 2018. Observing strategies in intercomprehension reading. Some clues for assessment in plurilingual settings. *Frontiers in Communication* 3(29).

Börestam Uhlmann, Ulla. 1994. *Skandinaver samtalar. Språkliga och interaktionella strategier i samtal mellem danskar, norrmän och svenskar* [Scandinavians converse. Linguistic and interactional strategies in conversations between Danes, Norwegians and Swedes]. Uppsala: Uppsala University dissertation.

Börestam Uhlmann, Ulla. 2005. Interscandinavian language contact I: Internal communication and comprehensibility problems. In Oskar Bandle, Kurt Braunmüller, Lennart Elmevik, Ernst H. Jahr, Allan Karker, Hans-Peter Naumann, Ulf Telemann & Gun Widmark (eds.), *The Nordic languages: An international handbook of the history of the North Germanic languages. Vol. 2*, 2025–2031. Berlin, Boston: De Gruyter Mouton.

Börestam, Ulla. 1987. *Dansk-svensk språkgemenskap på undantag* [Danish-Swedish language community as a special case].Uppsala: Uppsala University.

Börestam, Ulla. 2015. Excuse me, but can you tell me where the Nordic house is located? Linguistic strategies in inter-Nordic communication in Iceland illustrated through participant observation. *Linguistics* 53(2). 219–254.

Börestam, Ulla & Lise Horneman Hansen. 2021. Dynamic comprehension between neighboring languages: An experiment on Danish and Swedish. In Éric Castagne & Philippe Monneret (eds.), *Intercompréhension et analogie*, 95–109. Louvain-la-Neuve: De Boeck Supérieur.

Borghini, Giulia & Valerie Hazan. 2018. Listening effort during sentence processing is increased for non-native listeners: A pupillometry study. *Frontiers in Neuroscience* 12(152). 1–13.

Bot, Kees de. 2004. The multilingual lexicon: Modelling selection and control. *International Journal of Multilingualism* 1(1). 17–32.

Bot, Kees de, Wander Lowie & Marjolijn Verspoor. 2005. *Second language acquisition: An advanced resource book*. London: Routledge.
Bradlow, Ann R., Gina M. Torretta & David B. Pisoni. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20(3). 255–272.
Branets, Anna, Daria Bahtina & Anna Verschik. 2020. Mediated receptive multilingualism: Estonian-Russian-Ukrainian case study. *Linguistic Approaches to Bilingualism* 10(3). 380–411.
Braunmüller, Kurt. 2006. Vorbild Skandinavien? Zur Relevanz der rezeptiven Mehrsprachigkeit in Europa. In Konrad Ehlich & Antonie Hornung (eds.), *Praxen der Mehrsprachigkeit*, 11–31. Münster: Waxmann Verlag.
Braunmüller, Kurt. 2007. Receptive multilingualism in northern Europe in the Middle Ages. In Jan Ten Thije & Ludger Zeevaert (eds.), *Receptive multilingualism: Linguistic analyses, language policies and didactic concepts*, 25–47. Amsterdam: John Benjamins.
Breeuwer, Marcel & Reinier Plomp. 1986. Speechreading supplemented with auditorily presented speech parameters. *The Journal of the Acoustical Society of America* 79(2). 481–499.
Brink, Lars & Jørn Lund. 1975. *Dansk rigsmål: Lydudviklingen siden 1840 med særligt henblik på sociolekterne i København* [Danish standard language: Sound development since 1840 with special attention to sociolects in Copenhagen]. Copenhagen: Gyldendal.
Brodeur, Mathieu B., Katherine Guérard & Maria Bouras. 2014. Bank of Standardized Stimuli (BOSS) Phase II: 930 new normative photos. PLoS ONE 9(9): e106953.
Broeders, A. P. A., Tina Cambier-Langeveld & Jos Vermeulen. 2002. Arranging a voice lineup in a foreign language. *The International Journal of Speech, Language and the Law* 9(1). 1350–1771.
Brown, Gillian, Anne Anderson, Richard Shillcock & George Yule. 1984. *Teaching talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.
Brye, Edward & Elizabeth Brye. 2002. *Rapid appraisal and intelligibility testing surveys of the Eastern Beboid group of languages (Northwest Province)*. Yaoundé: SIL Electronic Survey Reports: SILESR, 19.
Brysbaert, Marc, Pawel Mandera & Emmanuel Keuleers. 2018. The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science* 27(1). 45–50.
Brysbaert, Marc, Michaël Stevens, Simon De Deyne, Wouter Voorspoels & Gert Storms. 2014. Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica* 150. 80–84.
Budovičová, Viera. 1987. Literary languages in contact (A sociolinguistic approach to the relation between Slovak and Czech today). In Jan Chloupek & Jiří Nekvapil (eds.), *Reader in Czech sociolinguistics*, 156–175. Amsterdam/Philadelphia: John Benjamins.
Bulatović, Stefan, Anja Schüppert & Charlotte Gooskens. 2019. Receptive multilingualism versus Elf: How well do Slovenes understand Croatian compared to Croatian speakers' English? *Journal of English as a Lingua Franca* 8(1). 37–65.
Campbell-Kibler, Kathryn. 2012. The Implicit Association Test and sociolinguistic meaning. *Lingua* 122 (7). 753–763.
Carlucci, Alessandro. 2020. How did Italians communicate when there was no Italian? Italo-Romance intercomprehension in the Late Middle Ages. *The Italianist* 40(1). 19–43.
Casad, Eugene H. 1974. *Dialect intelligibility testing*. Norman, OK: Summer Institute of Linguistics of the University of Oklahoma.
Cate, Abraham P. ten & Peter Jordens. 1990. *Phonetik des Deutschen: Eine Kontrastiv Deutsch-Niederländische Beschreibung für den Zweitspracherwerb*. Dordrecht: Foris.
Cattell, Raymond B. 1963. Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology* 54(1). 1–22.

Čéplö, Slavomír, Ján Bátora, Adam Benkato, Jiří Milička, Christophe Pereira & Petr Zemánek. 2016. Mutual intelligibility of spoken Maltese, Libyan Arabic, and Tunisian Arabic functionally tested: A pilot study. *Folia Linguistica* 50(2). 583–628.

Cheng, Chin-Chuan. 1997. Measuring relationship among dialects: DOC and related resources. *Computational Linguistics and Chinese Language Processing* 2(1). 41–72.

Chiswick, Barry R. & Paul W. Miller. 2005. Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development* 26(1). 1–11.

Ciobanu, Alina M. & Liviu P. Dinu. 2020. Automatic identification and production of related words for historical linguistics. *Computational Linguistics* 45(4). 667–704.

Claridge, Gordon. 1977. Manual of the Eysenck personality questionnaire (junior and adult). *Behaviour Research and Therapy* 15(4). 369–370.

Collins, Beverley & Inger Mees. 1981. *The sounds of English and Dutch*. The Hague: Leiden University Press.

Conti, Virginie & François Grin. 2008. *S'entendre entre langues voisines: Vers l'intercompréhension*. Chêne-Bourg: Georg.

Costa, Paul T. & Robert R. McCrae. 1992. Normal personality assessment in clinical practice: The NEO personality inventory. *Psychological Assessment* 4(1). 5–13.

Council of Europe. 1992. *European charter for regional or minority languages*. Strasbourg: Council of Europe.

Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Crystal, David. 1975. Paralinguistics. In Jonathan Benthall & Ted Polhemus (eds.), *The body as a medium of expression*, 162–174. London: Institute of Contemporary Arts.

Crystal, David. 2009. *Dictionary of linguistics and phonetics*. Chichester: John Wiley Sons.

Crystal, David & Robert H. Robins. 2021. Language. *Encyclopedia Britannica* https://www.britannica.com/topic/language (accessed 7 April 2024).

Cutler, Anne. 2012. *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.

Cutler, Anne, Nuria Sebastián-Gallés, Olga Soler-Vilageliu & Brit Van Ooijen. 2000. Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition* 28 (5). 746–755.

Declerck, Mathieu & Andrea M. Phillipp. 2015. A review of control processes and their locus in language switching. *Psychonomic Bulletin Review* 22. 1630–1645.

Delsing, Lars-Olof. 2007. Scandinavian intercomprehension today. In Jan D. Ten Thije & Ludger Zeevaert (eds.), *Receptive multilingualism*, 231–246. Amsterdam/Philadelphia: John Benjamins.

Delsing, Lars-Olof & Katarina Lundin-Åkesson. 2005. *Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska* [Does the language keep together the Nordic countries? A research report of mutual comprehension between young speakers of Danish, Swedish and Norwegian]. Copenhagen: Nordic Council of Ministers.

Derwing, Tracey M. & Murray J. Munro. 1997. Accent, intelligibility and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition* 19(1). 1–16.

Dewaele, Jean-Marc. 1998. Lexical inventions: French interlanguage as L2 versus L3. *Applied Linguistics* 19(4). 471–90.

Dijkstra, Ton. 2003. Lexical processing in bilinguals and multilinguals. In Ulrike Jasone, Ulrike Jessner & Britta Hufeisen (eds.), *The Multilingual Lexicon*, 11–26. Dordrecht: Springer.

Doetjes, Gerald. 2007. Understanding differences in inter-Scandinavian language understanding. In Jan D. Ten Thije & Ludger Zeevaert (eds.), *Receptive multilingualism. Linguistic analyses, language policies and didactic concepts*, 217–230. Amsterdam: John Benjamins.

Dörnyei, Zoltán. 2005. *The psychology of the language learner: Individual differences in second language acquisition*. New York: Routledge.

Doyé, Peter. 2005. *Intercomprehension. Guide for the development of language education policies in Europe: From linguistic diversity to plurilingual education*. Strasbourg: Council of Europe.

Drager, Katie. 2018 *Experimental research methods in sociolinguistics*. London: Bloomsbury Academic.

Dragojevic, Marko & Howard Giles. 2016. I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research* 42 (3). 396–420.

Drijvers, Linda. 2019. *On the oscillatory dynamics underlying speech-gesture integration in clear and adverse listening conditions*. Nijmegen: Radboud University Nijmegen dissertation.

Drijvers, Linda & Asli Ozyürek. 2017. Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *Journal of Speech, Language & Hearing Research* 60(1). 212–222.

Drijvers, Linda & Asli Özyürek. 2020. Non-native listeners benefit less from gestures and visible speech than native listeners during degraded speech comprehension. *Language and Speech* 63 (2). 209–220.

Dunn, L. M., & Dunn, D. M. 2007. *Peabody picture vocabulary test* (4th ed.). Minneapolis, MN: Pearson Assessments.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2023. *Ethnologue: Languages of the World*. Twenty-sixth edition. Dallas, TX: SIL International. http://www.ethnologue.com (accessed 7 April 2024).

Edwards, John. 1999. Refining our understanding of language attitudes. *Journal of Language and Social Psychology* 18(1). 101–110.

EF EPI. 2023. EF English proficiency index. *EF Education First*. http://www.ef.nl/epi/ (accessed 7 April 2024).

Elbro, Carsten. 2005. Literacy acquisition in Danish: A deep orthography in cross-linguistic light. In R. Malatesha Joshi & P. G. Aaron (eds.), *Handbook of orthography and literacy*, 31–45. New York, NY: Routledge.

Engen, Kristin J. van & Drew J. McLaughlin. 2018. Eyes and ears: Using eye tracking and pupillometry to understand challenges to speech recognition. *Hearing Research* 369. 56–66.

Escudero, Paola, Ellen Simon & Holger Mitterer. 2012. The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics* 40(2). 280–288.

European Commission. 2007. *Final report high level group on multilingualism*. Luxembourg: Office for Official Publications of the European Communities.

Feleke, Tekabe L., Charlotte Gooskens & Stefan Rabanus. 2020. Mapping the dimensions of linguistic distance: A study on South Ethiosemitic languages. *Lingua* 243. 102893.

Ferguson, Charles A. 1971. Absence of copula and the notion of simplicity: A study of normal speech, baby talk, foreigner talk and pidgins. In D. Hymes (ed.), *Pidginization and creolization of languages*, 141–150. London: Cambridge University Press.

Field, John. 2019. Second language listening: Current ideas, current issues. In John W. Schwieter & Alessandro Benati (eds.), *The Cambridge handbook of language learning*, 283–319. Cambridge: Cambridge University Press.

Frinsel, Felicity, Anne Kingma, Femke Swarte & Charlotte Gooskens. 2015. Predicting the asymmetric intelligibility between spoken Danish and Swedish using conditional entropy. *Tijdschrift voor Scandinavistiek* 34(2). 120–138.

Fuller, Janet M. 1999. Between three languages: Composite structure in interlanguage. *Applied Linguistics* 20(4). 534–561.

Gass, Susan & Evangeline Marlos Varonis. 1984. The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning* 34(1). 65–87.

Giles, Howard & Nancy Niedzielski. 1998. Italian is beautiful, German is ugly. In L. Bauer & Peter Trudgill (eds.), *Language myths*, 85–93. London: Penguin Books.

Giles, Howard & Tania Ogay. 2007. Communication accommodation theory. In Bryan B. Whaley & Wendy Samter (eds.), *Explaining communication: Contemporary theories and exemplars*, 293–310. Mahwah, NJ: Lawrence Erlbaum.

Godfroid, Aline. 2020. *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. New York, NY: Routledge.

Goldberg, Lewis R. 1993. The structure of phenotypic personality traits. *The American Psychologist* 48 (1). 26–34.

Golubović, Jelena. 2016. *Mutual intelligibility in the Slavic language area*. Groningen: University of Groningen dissertation.

Gooskens, Charlotte. 2005. Traveling time as a predictor of linguistic distance. *Dialectologia et Geolinguistica* 13. 38–62.

Gooskens, Charlotte. 2007. The contribution of linguistic factors to the intelligibility of closely related languages. *Journal of Multilingual and Multicultural Development* 28(6). 445–467.

Gooskens, Charlotte & Gerald Doetjes. 2009. Skriftsprogets rolle i den dansk-svenske talesprogsforståelse [The role of orthography in the Danish-Swedish mutual intelligibility]. *Språk och Stil* 19. 105–123.

Gooskens, Charlotte & Wilbert Heeringa. 2004a. Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language Variation and Change* 16(3). 189–207.

Gooskens, Charlotte & Wilbert Heeringa. 2004b. The position of Frisian in the Germanic language area. In Dicky Gilbers, Maartje Schreuder & Nienke Knevel (eds.), *On the boundaries of phonology and phonetics*, 61–87. Groningen: University of Groningen.

Gooskens, Charlotte & Wilbert Heeringa. 2014. The role of dialect exposure in receptive multilingualism. *Applied Linguistics Review* 5(1). 247–271.

Gooskens, Charlotte, Wilbert Heeringa & Karin Beijering. 2008. Phonetic and lexical predictors of intelligibility. *International Journal of Humanities and Arts Computing* 2(1–2). 63–81.

Gooskens, Charlotte, Sebastian Kürschner & Renée van Bezooijen. 2012. Intelligibility of Swedish for Danes: Loan words compared with inherited words. In Henk van der Liet & Muriel Norde (eds.), *Language for its own sake*, 435–455. Amsterdam: Amsterdam University Press.

Gooskens, Charlotte, Sebastian Kürschner & Vincent J. van Heuven. 2022. The role of loanwords in the intelligibility of written Danish among Swedes. *Nordic Journal of Linguistics* 45(1). 4–29.

Gooskens, Charlotte & Cindy Schneider. 2016. Testing mutual intelligibility between closely related languages in an oral society. *Language Documentation & Conservation* 10. 278–305.

Gooskens, Charlotte & Cindy Schneider. 2019. Linguistic and non-linguistic factors affecting intelligibility across closely related varieties in Pentecost Island, Vanuatu. *Dialectologia* 23. 61–85.

Gooskens, Charlotte & Femke Swarte. 2017. Linguistic and extra-linguistic predictors of mutual intelligibility between Germanic languages. *Nordic Journal of Linguistics* 40(2). 123–147.

Gooskens, Charlotte & Renée van Bezooijen. 2006. Mutual comprehensibility of written Afrikaans and Dutch: Symmetrical or asymmetrical? *Literary and Linguistic Computing* 21(4). 543–557.

Gooskens, Charlotte & Renée van Bezooijen. 2013. Explaining Danish-Swedish asymmetric word intelligibility – An error analysis. In Charlotte Gooskens & Renée van Bezooijen (eds.), *Phonetics in Europe: Perception and production*, 59–82. Frankfurt a. M.: Peter Lang.

Gooskens, Charlotte & Renée van Bezooijen. 2014. The effect of pause insertion on the intelligibility of Danish among Swedes. In Johanneke Caspers, Yiya Chen, Willemijn Heren, Jos Pacilly, Niels O. Schiller & Ellen van Zanten (eds.), *Above and beyond the segments. Experimental linguistics and phonetics*, 96–108. Amsterdam: John Benjamins.

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

Gooskens, Charlotte & Vincent J. van Heuven. 2017. Measuring cross-linguistic intelligibility in the Germanic, Romance and Slavic language groups. *Speech Communication* 89. 25–36.

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

Gooskens, Charlotte & Vincent J. van Heuven. 2021. Mutual intelligibility. In Marcos Zampieri & Preslav Nakov (eds.), *Similar languages, varieties, and dialects: A computational perspective*, 51–95. Cambridge: Cambridge University Press.

Gooskens, Charlotte, Vincent J. van Heuven, Jelena Golubović, Anja Schuppert, Femke Swarte & Stefanie Voigt. 2018. Mutual intelligibility between closely related languages in Europe. *International Journal of Multilingualism* 15(2). 169–193.

Gooskens, Charlotte, Vincent J. van Heuven, Renée van Bezooijen & Jos J. A. Pacilly. 2010. Is spoken Danish less intelligible than Swedish? *Speech Communication* 52(11–12). 1022–1037.

Goslin, Jeremy, Hester Duffy & Caroline Floccia. 2012. An ERP investigation of regional and foreign accent processing. *Brain and Language* 122(2). 92–102.

Grimes, Joseph E. 1992. Correlations between vocabulary similarity and intelligibility. In Eugene H. Casad (ed.), *Windows on bilingualism*, 17–32. Dallas, TX: Summer Institute of Linguistics and the University of Texas at Arlington.

Grønnum, Nina. 1998. *Fonetik og fonologi: Almen og dansk* [Phonetics and phonology: General and Danish].Copenhagen: Akademisk Forlag.

Grønnum, Nina. 2009. A Danish phonetically annotated spontaneous speech corpus (DanPASS). *Speech Communication* 51(7). 594–603.

Grünbaum, Catharina & Mikael Reuter. 2013. *Att förstå varandra i Norden – språkråd til nordbor i nordisk samarbete* [To understand one and another in the Nordic – language advice for inhabitants of the Nordic countries].Helsinki: Nordiska ministerrådet.

Hall, Robert A. 1974. *External history of the Romance languages*. New York, NY: Elsevier.

Hammarström, Harald. 2005. Review of Ethnologue: Languages of the world, 15th edn. *Linguist List* 16.2637.

Hammarström, Harald. 2008. Counting languages in dialect continua using the criterion of mutual intelligibility. *Journal of Quantitative Linguistics* 15(1). 34–45.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2023. *Glottolog 4.8*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://glottolog.org (accessed 7 April 2024).

Härmävaara, Hanna-Ilona & Charlotte Gooskens. 2019. Mutual intelligibility of Finnish and Estonian vocabulary. *Lähivõrdlusi. Lähivertailuja* 29. 13–56.

Hatch, Evelyn. 1983. *Psycholinguistics: A second language perspective*. Rowley, MA: Newbury House.

Haugen, Einar. 1966. Semicommunication: The language gap in Scandinavia. *Sociological Inquiry* 36 (2). 280–297.

Ház, Eva. 2005. *Deutsche und Niederländer. Untersuchungen zur Möglichkeit einer unmittelbaren Verständigung*. Hamburg: Kovač.

Hedquist, Rolf. 1985. *Nederländares förståelse av danska och svenska. En språkpedagogisk undersökning med utnyttjande av likheterna mellan språken* [Dutchmen's understanding of Danish and Swedish. A language pedagogical investigation with the use of similarities between the languages].Umeå: Institutionerna för fonetik och nordiska språk, Umeå university.

Heeringa, Wilbert. 2004. *Measuring dialect pronunciation differences using Levenshtein distance*. Groningen: University of Groningen dissertation.

Heeringa, Wilbert, Charlotte Gooskens & Vincent J. van Heuven. 2023. Comparing Germanic, Romance and Slavic: Relationships among linguistic distances. *Lingua* 287. 103512.

Heeringa, Wilbert, Peter Kleiweg, Charlotte Gooskens & John Nerbonne. 2006. Evaluation of string distance algorithms for dialectology. In John Nerbonne & Erhard Hinrichs (eds.), *Linguistic distances workshop at the joint conference of international committee on computational linguistics and the sssociation for computational linguistics*, *Sydney*, *July*, *2006*, 51–62. Stroudsburg, PA: Association for Computational Linguistics.

Heeringa, Wilbert, Femke Swarte, Anja Schüppert & Charlotte Gooskens. 2014. Modeling intelligibility of written Germanic languages: Do we need to distinguish between orthographic stem and affix variation? *Journal of Germanic Linguistics* 26(4). 361–394.

Heeringa, Wilbert, Femke Swarte, Anja Schüppert & Charlotte Gooskens. 2018. Measuring syntactical variation in Germanic texts. *Digital Scholarship in the Humanities* 33(2). 279–296.

Heeringa, Wilbert & Vincent J. van Heuven & Hans van de Velde. 2023. LED-A: Levenshtein Edit Distance App [computer program]. https://www.led-a.org/(accessed 7 April 2024).

Heinz, Bettina. 2003. Backchannel responses as strategic responses in bilingual speakers' conversations. *Journal of Pragmatics* 35(7). 1113–1142.

Helms-Park, Rena & Vedran Dronjic. 2016. Crosslinguistic lexical influence: Cognate facilitation. In Rosa Alonso Alonso (ed.), *Crosslinguistic influence in second language acquisition*, 71–92. Bristol: Multilingual Matters.

Herkenrath, Annette. 2012. Receptive multilingualism in an immigrant constellation: Examples from Turkish-German children's language. *International Journal of Bilingualism*. 16(3). 287–314.

Heuven, Vincent J. van 2008. Making sense of strange sounds: (Mutual) intelligibility of related language varieties. A review. *International Journal of Humanities and Arts Computing* 2(1–2). 39–62.

Heuven, Vincent J. van 2022. Resolving the prosody paradox. In Pablo Arantes & Amanda Post da Silveira (eds.), *Prosody and bilingualism*, 168–204. Letraria: Araraquara.

Heuven, Vincent J. van & Jan W. de Vries. 1981. Begrijpelijkheid van buitenlanders: De rol van fonische versus niet-fonische factoren [Comprehensibility of foreigners: The role of phonic versus non-phonic factors]. *Forum der Letteren* 22. 309–320.

Heuven, Vincent J. van & Klaske van Leyden. 1996. Lexical stress and spoken word recognition, Dutch versus English. In Marcel den Dikken & Crit Cremers (eds.), *Linguistics in the Netherlands*, 159–170. Amsterdam: John Benjamins.

Hickerson, Harold, Glen D. Turner & Nancy P. Hickerson. 1952. Testing procedures for estimation transfer of information among Iroquois dialects and languages. *International Journal of American Linguistics* 18(1). 1–8.

Hilton, Nanna H. 2010. *Regional dialect levelling and language standards: Changes in the Hønefoss dialect*. York: University of York dissertation.

Hilton, Nanna H., Charlotte Gooskens & Anja Schüppert. 2013. The influence of non-native morphosyntax on the intelligibility of a closely related language. *Lingua* 137. 1–18.

Hilton, Nanna H., Charlotte Gooskens, Anja Schüppert & Chaoju Tang. 2022. Is Swedish more beautiful than Danish? Matched guise investigations with unknown languages. *Nordic Journal of Linguistics* 45(1). 30–48.

Hilton, Nanna H., Anja Schüppert & Charlotte Gooskens. 2011. Syllable reduction and articulation rates in Danish, Norwegian and Swedish. *Nordic Journal of Linguistics* 34(2). 215–237.

Houwer, Jan de, Sarah Teige-Mocigemba, Adriaan Spruyt & Agnes Moors 2009. Implicit measures: A normative analysis and review. *Psychological Bulletin* 135(3). 347–368.

Hufeisen, Britta & Nicole Marx. 2007. How can DaFnE and EuroComGerm contribute to the concept of receptive multilingualism? Theoretical and practical considerations. In Jan D. ten Thije & Ludger Zeevaert (eds.), *Receptive Multilingualism*, 307–321. Amsterdam, Philadelphia: John Benjamins.

Huisman, John L. A., Roeland van Hout & Asifa Majid. 2021. Cross-linguistic constraints and lineage-specific developments in the semantics of cutting and breaking in Japonic and Germanic. *Linguistic Typology* 27(1). 41–75.

Impe, Leen. 2010. *Mutual intelligibility of national and regional varieties of Dutch in the Low Countries*. Leuven: University of Leuven dissertation.

Impe, Leen, Dirk Geeraerts & Dirk Speelman. 2008. Mutual intelligibility of standard and regional Dutch language varieties. *International Journal of Humanities and Arts Computing* 2(1–2). 101–117.

Iwasaki, Junko & Rhonda Oliver. 2003. Chat-line interaction and negative feedback. *Australian Review of Applied Linguistics. Supplement Series*, 17(1). 60–73.

Jágrová, Klára, Tania Avgustinova, Irina Stenger & Andrea Fischer. 2019. Language models, surprisal and fantasy in Slavic intercomprehension. *Computer Speech and Language* 53. 242–275.

Jain, Anil K. & Richard C. Dubes. 1988. *Algorithms for clustering data*. Upper Saddle River, NJ: Prentice-Hall.

Janin, P.ierre 2008. L'intercompréhension dans une politique francophone du plurilinguisme. In V. irginie Conti & F.rançois Grin (eds.), *S'entendre entre langues voisines: Vers l'intercompréhension*, 53–77. Chêne-Bourg: Georg.

Janse, Esther, Sieb G. Nooteboom & Hugo Quené. 2003. Word-level intelligibility of time-compressed speech: Prosodic and segmental factors. *Speech Communication* 41 (2–3). 287–301.

Jensen, John B. 1989. On the mutual intelligibility of Spanish and Portuguese. *Hispania* 72(4). 848–852.

Jensen, Kristian T. H. 2009. Indicators of text complexity. *Copenhagen Studies in Language* 37. 61–80.

Jörgensen, Nils & Eva Kärrlander. 2001. *Grannspraksförståelse i Öresundsregionen år 2000. Gymnasisters hörförståelse* [Mutual intelligibility in the Oresundsregion year 2000: The listening comprehension of secondary school students]. Lund: University of Lund.

Kaivapalu, Annekatrin. 2015. Eesti ja soome keele vastastikune mõistmine üksiksõna- ja tekstitasandil: Lingvistilised tegurid, mõistmisprotsess ja sümmeetria [Mutual intelligibility of Estonian and Finnish context-free words and texts: Linguistic determinants, comprehension process and symmetry]. *Eesti Rakenduslingvistika Ühingu Aastaraamat* 11. 55–74.

Kaivapalu, Annekatrin & Maisa Martin. 2014. Symmetry of similarity: Definition, perception, measurement: Finnish and Estonian noun morphology as a testing ground. In Heli Paulasto, Lea Meriläinen, Helka Riionheimo & Maria Kok (eds.), *Language contacts at the crossroads of disciplines*, 283–318. Newcastle upon Tyne: Cambridge Scholars.

Kaivapalu, Annekatrin & Maisa Martin. 2017. Perceived similarity between written Estonian and Finnish: Strings of letters or morphological units? *Nordic Journal of Linguistics* 40(2). 149–174.

Kalikow, Daniel N., Kenneth N. Stevens & Lois L. Elliott. 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America* 61(5). 1337–1351.

Kang, Okim & Donald Rubin. 2009. Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology* 28(4). 441–456.
Karl, John R. & David B. Pisoni. 1994. The role of talker-specific information in memory for spoken sentences. *The Journal of the Acoustical Society of America* 95(5). 2873.
Kaushanskaya, Margarita, Henrike K. Blumenfeld & Viorica Marian. 2020. The language experience and proficiency questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition* 23 (5). 945–950.
Kellerman, E. 1978. Giving learners a break: Native language intuitions as a source of predictions about transferability. *Working Papers on Bilingualism* 15. 59–92.
Keshavarz, Mohammad H. & Hossein Salimi. 2007. Collocational competence and cloze test performance: A study of Iranian EFL learners. *International Journal of Applied Linguistics* 17(1). 81–92.
Kessler, Brett. 1995. Computational dialectology in Irish Gaelic. In Steven P. Abney, Erhard W. Hinrichs (ed.), *Proceedings of the seventh conference on European chapter of the association for computational linguistics*, Dublin, Ireland. 60–66. San Francisco, CA: Morgan Kaufmann.
Kirchner, Ruth. 2022. Questionnaires to elicit quantitative data. In Ruth Kirchner & Lena Zipp (eds.), *Research methods in language attitudes*, 129–144. Cambridge: Cambridge University Press.
Kita, Sotaro. 2009. Cross-cultural variation of speech-accompanying gesture: A review. *Language and Cognitive Processes* 24(2). 145–167.
Klein, Horst G. & Tilbert D. Stegmann. 2000. *EuroComROM, die sieben Siebe: Romanische Sprachen sofort lesen können*. Aachen: Shaker.
Kloss, Heinz. 1967. “Abstand languages” and “ausbau languages”. *Anthropological Linguistics* 9(7). 29–41.
Kluge, Angela. 2006. *RTT retelling method: An alternative approach to intelligibility testing*. Dallas, TX: SIL International.
Kluge, Angela. 2007. *“Sorry, could you repeat that, please?!” – Where does one language end and the next begin?* Jakarta: SIL International-Indonesia.
Krug, Manfred G. & Julia Schlüter. 2013. *Research methods in language variation and change*. New York, NY: Cambridge University Press.
Kruskal, Joseph. 1999. An overview of sequence comparison. In David A. Sankoff & Joseph Kruskal (eds.), *Time warps, string edits and macromolecules: The theory and practice of sequence comparison*, 1–44. Stanford, CA: CSLI Publications.
Kudera, Jacek. 2022. *Slavic receptive multilingualism: Intercomprehension of speech*. Saarbrücken: Saarland University Press dissertation.
Kumove, Michael. 2020. Diversity, semi-communication and cross-country trust: A quantitative analysis. *Social Science Research* 86. 1–18.
Kurpaska, Maria. 2010. *Chinese language(s): A look through the prism of the great dictionary of modern Chinese dialects*. Berlin: De Gruyter Mouton.
Kürschner, Sebastian, Charlotte Gooskens & Renée van Bezooijen. 2008. Linguistic determinants of the intelligibility of Swedish words among Danes. *International Journal of Humanities and Arts Computing* 2(1–2). 83–100.
Kyjánek, Lukáš & Jiří Haviger. 2019. The Measurement of mutual intelligibility between West-Slavic languages. *Journal of Quantitative Linguistics* 26(3). 205–230.
Ladefoged, Peter. 1968. *The measurement of cross-language communication*. Washington, DC: ERIC Clearinghouse.
Ladefoged, Peter. 1970. The measurement of phonetic similarity. *Statistical Methods in Linguistics* 6. 23–32.

Lambelet, Amelia & Pierre-Yves Mauron. 2017. Receptive multilingualism at school: An uneven playing ground? *International Journal of Bilingual Education and Bilingualism* 20(7). 854–867.

Lambert, Wallace E., Richard C. Hodgson, Robert D. Gardner & Samuel Fillenbaum. 1960. Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology* 60(1). 44–51.

Lecumberri, Maria L. G., Martin Cooke & Anne Cutler. 2010. Non-native speech perception in adverse conditions: A review. *Speech Communication* 52(11–12). 864–886.

Leemann, Adrian, Curdin Derungs & Stephan Elspaß. 2019. Analyzing linguistic variation and change using gamification web apps: The case of German-speaking Europe. *PLoS ONE* 14(12): e0225399.

Leinonen, Therese, Çağrı Çöltekin & John Nerbonne. 2016. Using Gabmap. *Lingua* 178. 71–83.

Lepage, Andrée & Maria Grazia Busà. 2013. Intelligibility of English L2: The effects of incorrect word stress placement and incorrect vowel reduction in the speech of French and Italian learners of English. In *Proceedings of the international symposium on the acquisition of second language speech*, 387–400. Montreal, Quebec: Concordia Working Papers in Applied Linguistics.

Lev-Ari, Shiri & Boaz Keysar. 2010. Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology* 46(6). 1093–1096.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation* (ACL-MIT Press series in natural-language processing). Cambridge, MA: MIT Press.

Levenshtein, Vladimir I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10(8). 707–710.

Lewis, M. P., Gary Simons & Charles Fennig (eds.). 2013. *Ethnologue: Languages of the World, 17th ed*. Dallas, TX: SIL International.

Lieberman, Philip. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6(3). 172–187.

Lightbown, Patsy & Nina Spada. 2006. *How languages are learned*. Oxford: Oxford University Press.

Lilley, Jason, Amanda Stent & Ilija Zeljkovic. 2012. A random, semantically appropriate sentence generator for speaker verification. In *Proceedings of INTERSPEECH 2012*, 739–742.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H & H theory. In William J. Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Springer.

Lindemann, Stephanie & Nicholas Subtirelu. 2013. Reliably biased: The role of listener expectation in the perception of second language speech. *Language Learning* 63(3). 567–594.

Loureiro-Rodríguez, Verónica & Elif Fidan Acar. 2022. The matched-guise technique. In Ruth Kirchner & Lena Zipp (eds.), *Research methods in language attitudes*, 185–202. Cambridge: Cambridge University Press.

Luce, Paul A. & David B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*. 19. 1–36.

Lüdi, Georges. 2013. Receptive multilingualism as a strategy for sharing mutual linguistic resources in the workplace in a Swiss context. *International Journal of Multilingualism* 10(2). 140–158.

Lundin, Katarina & Robert Z. Christensen. 2001. *Grannspraksförståelse i Öresundsregionen år 2000. Gymnasisters läsförståelse* [Mutual intelligibility in the Oresund region year 2000: The reading comprehension of secondary school students]. Lund: University of Lund.

Lyons, John. 1977. *Semantics, Vol. 2.* Cambridge: Cambridge University Press.

Markham, Duncan & Valerie Hazan. 2004. The effect of talker- and listener-related factors on intelligibility for a real-word, open-set perception test. *Journal of Speech, Language, and Hearing Research* 47(4). 725–737.

Marslen-Wilson William, D. 1978. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10(1). 29–63.

Martin, Frederick N., Craig A. Champlin & Desiree D. Perez. 2000. The question of phonetic balance in word recognition testing. *Journal of the American Academy of Audiology* 11(9). 489–493.
Marx, Nicole. 2012. Reading across the Germanic languages: Is equal access just wishful thinking? *International Journal of Bilingualism* 16(4). 467–483.
Maurud, Øivind. 1976. Reciprocal comprehension of neighbour languages in Scandinavia. *Scandinavian Journal of Educational research* 20(1). 49–72.
McDonough, Kim, Pavel Trofimovich, Oguzhan Tekin, & Masatoshi Sato. 2022. Exploring linguistic stereotyping of international students at a Canadian university. *Journal of Multilingual and Multicultural Development*. 1–16.
Mckaughan, Howard. 1964. A study of divergence in four New Guinea languages. *American Anthropologist* 66(4). 98–120.
Melo-Pfeifer, Silvia. 2014. Intercomprehension between Romance languages and the role of English: A study of multilingual chat rooms. *International Journal of Multilingualism* 11(1). 120–137.
Mennen, Ineke, Felix Schaeffler & Gerard Docherty. 2012. Cross-language differences in fundamental frequency range: A comparison of English and German. *Journal of the Acoustical Society of America* 131(3). 2249–2260.
Mercer, Neil. 2000. *Words and minds: How we use language to think together*. London: Routledge.
Mieszkowska, Karolina & Agnieszka Otwinowska. 2015. Is A2 in German better than B2 in French when reading Danish? The role of prior language knowledge when faced with an unknown language. In Gessica De Angelis, Ulrike Jessner & Marijana Kresic (eds.), *Crosslinguistic influence and crosslinguistic interaction in multilingual language learning*, 213–234. London: Bloomsbury Academic.
Milliken, Margaret E. 1988. *Phonological divergence and intelligibility: A case study of English and Scots*. Ann Arbor, MI: Cornell University dissertation.
Milliken, Margaret E. & Stuart R. Milliken. 1996. System relationships in assessing dialect intelligibility. *Notes on Linguistics* 72. 15–31.
Moberg, Jens, Charlotte Gooskens, John Nerbonne & Nathan Vaillette. 2007. Conditional entropy measures intelligibility among related languages. In Peter Dirix, Ineke Schuurman, Vincent Vandeghinste & Frank Van Eynde (eds.), *Computational linguistics in the Netherlands 2006: Selected papers from the 17th CLIN meeting*, 51–66. Utrecht: LOT.
Möller, Robert. 2011. Wann sind Kognaten erkennbar? Ähnlichkeit und synchrone Transparenz von Kognatenbeziehungen in der germanischen Interkomprehension. *Linguistik Online* 46(2). 79–101.
Möller, Robert & Ludger Zeevaert. 2015. Investigating word recognition in intercomprehension: Methods and findings. *Linguistics* 53(2). 313–352.
Mulken, Margo van & Berna Hendriks. 2015. Your language or mine? or English as a lingua franca? Comparing effectiveness in English as a lingua franca and L1-L2 interactions: Implications for corporate language policies. *Journal of Multilingual and Multicultural Development* 36(4). 404–422.
Mulken, Margo van & Berna Hendriks. 2017. Effective communication modes in multilingual encounters: Comparing alternatives in computer mediated communication (CMC). *Multilingua* 36 (1). 89–111.
Munro, Murray J. & Tracey M. Derwing. 1995a. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning* 45(1). 73–97.
Munro, Murray J. & Tracey M. Derwing. 1995b. Processing time, accent and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38(3). 289–306.
Munro, Murray J., Tracey M. Derwing & Susan L. Morton. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28(1). 111–131.

Myers, Isabel Briggs, Mary H. McCaulley, Naomi L. Quenk & Allen L. Hammer. 2018. MBTI manual for the global step I and step II assessments. Sunnyvale, CA: The Myers-Briggs Company.

Nábělková, Mira. 2007. Closely-related languages in contact: Czech, Slovak, “Czechoslovak”. *International Journal of the Sociology of Language* 183. 53–73.

Nahhas, Ramzi W. 2007. *The steps of recorded text testing: A practical guide*. https://www.sil.org/. (accessed 4 August 2024).

Nation, Paul & Rob Waring. 1997. Vocabulary size, text coverage, and word lists. In Norbert Schmitt & Michael McCarthy (eds.), *Vocabulary: Description, acquisition and pedagogy*, 6–19. Cambridge: Cambridge University Press.

Nerbonne, John & Wilbert Heeringa. 2010. Measuring dialect differences. In Jürgen E. Schmidt & Peter Auer (eds.), *Language and space: Theories and methods*, 550–567. Berlin: Mouton De Gruyter.

Nerbonne, John & Wybo Wiersma. 2006. A measure of aggregate syntactic distance. In John Nerbonne & Erhard Hinrichs (eds.), *Proceedings of the workshop on linguistic distances*, 82–90. Sydney, Australia: Association for Computational Linguistics.

Nooteboom, Sieb G. & Gert J. N. Doodeman. 1984. Speech quality and the gating paradigm. In Van den Broecke, Marcel P. R. & Antonie Cohen (eds.), *Proceedings of the tenth international congress of phonetic sciences*, 1–6 August, Utrecht, The Netherlands, 481–485. Dordrecht: Foris.

Nooteboom, Sieb G. & Maarten J. Van der Vlugt. 1988. A search for a word-beginning superiority effect. *Journal of the Acoustical Society of America* 84(6). 2018–2032.

Osgood, Charles E., George J. Suci & Percy H. Tannenbaum. 1957. *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Otwinowska-Kasztelanic, Agnieszka. 2011. Awareness and affordances. Multilinguals versus bilinguals and their perceptions of cognates. In G. De Angelis & Jean-Marc Dewaele (eds.), *New trends in crosslinguistic influence and multilingualism research*, 1–18. Bristol: Multilingual Matters.

Özek, Fatih, Bilgit Sağlam & Charlotte Gooskens. 2021. Mutual intelligibility of a Kurmanji and a Zazaki dialect spoken in the province of Elazığ, Turkey. *Applied Linguistic Review*, 14(5) 1411–1449.

Pedersen, Inge L. 2003. Traditional dialects of Danish and the de-dialectalization 1900–2000. *International Journal of the Sociology of Language* 159. 9–28.

Pierce, Joe E. 1952. Dialect distance testing in Algonquian. *International Journal of American Linguistics* 18. 208–218.

Pinto, Manuela & Shalom Zuckerman. 2019. Coloring Book: A new method for testing language comprehension. *Behavior Research Methods* 51. 2609–2628.

Pinto, Manuela & Shalom Zuckerman. 2020. KleurenSchat. Amsterdam: Boom uitgevers.

Posner, Rebecca. 1996. *The Romance languages*. Cambridge: Cambridge University Press.

Prince, Alan & Paul Smolensky. 2004. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.

Rastatter, Michael P. & Alan J. Gallaher. 1982. Reaction-times of normal subjects to monaurally presented verbal and tonal stimuli. *Neuropsychologia* 20(4). 465–473.

Raven. John, C. 1965. *Advanced progressive matrices. Sets I and II*. London: K. Lewis & Co.

Rehbein, Jochen & Olena Romaniuk. 2014. How to check understanding across languages. An introduction into the pragmatic index of language distance (PILaD) usable to measure mutual understanding in receptive multilingualism, illustrated by conversations in Russian, Ukrainian and Polish. *Applied Linguistics Review* 5(1). 131–171.

Rehbein, Jochen, Jan D. ten Thije & Anna Verschik. 2012. Lingua receptiva (LaRa)–remarks on the quintessence of receptive multilingualism. *International Journal of Bilingualism* 16(3). 248–264.

Reiterer, Susanne M., Vita Kogan, Annemarie Seither-Preisler & Gašper Pesek. 2020. Foreign language learning motivation: Phonetic chill or Latin lover effect? Does sound structure or social stereotyping drive FLL? In Kara D. Federmeier & Hsu-Wen Huang (eds.), *Psychology of learning and motivation: Adult and second language learning*, 165–205. Cambridge, MA: Academic Press.

Ribbert, Anne & Jan D. ten Thije. 2007. Receptive multilingualism in Dutch-German intercultural team cooperation. In Jan D. ten Thije & Ludger Zeevaert (eds.), *Receptive multilingualism: Linguistic analyses, language policies and didactic concepts*, 73–101. Amsterdam: John Benjamins.

Rijt, Luuk P. H. van de, Anja Roye, Emmanuel A. M. Mylanus, A. John van Opstal & Marc M.van Wanrooij. 2019. The principle of inverse effectiveness in audiovisual speech perception. *Frontiers in Human Neuroscience*, 13: 335.

Ring, James A. 1981. Ewe as a second language: A sociolinguistic survey of Ghana's Central Volta region. *Research Review* 12(2–3).

Ringbom, Håkan. 2007. *Cross-linguistic similarity in foreign language learning*. Bristol: Multilingual Matters.

Robb, Michael P., Margaret A. Maclagan & Yang Chen. 2004. Speaking rates of American and New Zealand varieties of English. *Clinical linguistics & phonetics* 18(1). 1–15.

Rooy, Raf. van 2020. *Language or dialect? The history of a conceptual pair*. Oxford: Oxford University Press.

Rosseel, Laura. 2022. The implicit association test paradigm. In Ruth Kirchner & Lena Zipp (eds.), *Research methods in language attitude*, 250–268. Cambridge: Cambridge University Press.

Rosseel, Laura, Dirk Speelman & Dirk Geeraerts. 2019. Measuring language attitudes in context: Exploring the potential of the personalized implicit association test. *Language in Society* 48(3). 429–461.

Røyneland, Unn. 2009. Dialects in Norway: Catching up with the rest of Europe? *International Journal of the Sociology of Language* 2009(196–197). 7–30.

Salehi, Mohammad, Aydin Neysani & Maria Popescu. 2017. Receptive intelligibility of Turkish to Iranian-Azerbaijani speakers. *Cogent Education* 4(1).

Santiago, Fabian & Paolo Mairano. 2022. Spaniards articulate faster than Mexicans. Temporal patterns in two varieties of Spanish. *Spanish in Context* 19(2). 244–264.

Scharpff, Peter J. & Vincent J. van Heuven. 1988. Effects of pause insertion on the intelligibility of low quality speech. In W. A. Ainsworth & John N. Holmes (eds.), *Proceedings of the 7th FASE/speech-88 symposium*, 261–268. Edinburgh: The Institute of Acoustics.

Schmidt, Karl-Heinz & Peter Metzler. 1992. *Wortschatztest (WST)*. Weinheim: Beltz.

Schneider, Cindy & Charlotte Gooskens. 2018. A follow-up analysis of listener (mis)comprehension across language varieties in Pentecost, Vanuatu. *Oceanic Linguistics* 57(1). 144–176.

Schüppert, Anja. 2011. *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish*. Groningen: University of Groningen dissertation.

Schüppert, Anja & Charlotte Gooskens. 2011. Investigating the role of language attitudes for perception abilities using reaction times. *Dialectologia* SI II. 119–140.

Schüppert, Anja & Charlotte Gooskens. 2012. The role of extra-linguistic factors in receptive bilingualism: Evidence from Danish and Swedish pre-schoolers. *International Journal of Bilingualism* 16(3). 332–347.

Schüppert, Anja, Wilbert Heeringa, Jelena Golubović & Charlotte Gooskens. 2017. Write as you speak? A cross-linguistic investigation of orthographic transparency in 16 Germanic, Romance and Slavic languages. In Martijn Wieling, Martin Kroon, Gertjan van Noord & Gosse Bouma (eds.), *From semantics to dialectometry. Festschrift in honor of John Nerbonne*, 303–313. Milton Keynes: College Publications.

Schüppert, Anja, Nanna H. Hilton & Charlotte Gooskens. 2016. Why is Danish so difficult to understand for fellow Scandinavians? *Speech Communication* 79. 47–60.

Schüppert, Anja, Nanna H. Hilton, Charlotte Gooskens & Vincent J. van Heuven. 2012. Syllable deletion in contemporary Danish. In Jan Heegård & Peter J. Henrichsen (eds.), *Speech in action. Proceedings of the 1th SJUSK conference on contemporary speech habits. Copenhagen studies in language*, *42*, 73–101. Frederiksberg: Samfundslitteratur.

Schüppert, Anja, Johannes C. Ziegler, Holger Juul & Charlotte Gooskens. 2022. Online activation of L1 Danish orthography enhances spoken word recognition of Swedish. *Nordic Journal of Linguistics* 45(1). 80–98.

Selinker, L. & B. Baumgartner-Cohen. 1995. Multiple language acquisition: "Damn it, why can't I keep these two languages apart". *Language, Culture and Curriculum*, 8(2). 115–121.

Shen, Yin-Chen & Elizabeth A. Franz. 2005. Hemispheric competition in left-handers on bimanual reaction time tasks. *Journal of Motor Behavior* 37(1). 3–9.

Sherkina-Lieber, Marina. 2020. A classification of receptive bilinguals: Why we need to distinguish them, and what they have in common. *Linguistic Approaches to Bilingualism* 10(3). 412–440.

Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press.

Simons, Gary F. 1979. *Language variation and limits to communication. Technical report no. 3.* Ithaca, NY: Department of Modern Languages and Linguistics, Cornell University.

Sindoni, Maria G. 2013. *Spoken and written discourse in online interactions: A multimodal approach*. New York, NY: Routledge.

Skjold Frøshaug, Andrea & Truls Stende. 2021. *Does the Nordic language community exist?* Copenhagen: Nordisk Ministerråd.

Sletten, Iben S., Arne Torp, Kaisa Häkkinen, Mikael Svonni. 2004. *Nordens sprog med rødder og fødder*. [The languages of the North with roots and feet]. Copenhagen: Nordisk Ministerråd.

Slik, Frans W. P. van der 2010. Acquisition of Dutch as a second language: The explanative power of cognate and genetic linguistic distance measures for 11 West-European first languages. *Studies in second language acquisition* 32. 401–432.

Smidfelt, Linda. 2018. An intercomprehension study of multilingual Swedish L1 speakers reading and decoding words in text in Italian, an unknown language. *Lingua* 204. 62–77.

Smith, Larry E. & Cecil L. Nelson. 1985. International intelligibility of English: Directions and resources. *World Englishes* 4(3). 333–342.

Smith, Larry E. & Khalilullah Rafiqzad. 1979. English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly* 13(3). 371–380.

Smorenburg, Laura & Aoju Chen. 2020. The effect of female voice on verbal processing. *Speech Communication* 122. 11–18.

So, Wing Chee. 2010. Cross-cultural transfer in gesture frequency in Chinese-English bilinguals. *Language and Cognitive Processes* 25(10). 1335–1353.

Sokal, Robert R. & F. James Rohlf. 1962. The comparison of dendrograms by objective methods. *Taxon* 11(2). 33–40.

Stenger, Irina, Tania Avgustinova & Roland Marti. 2017. Levenshtein distance and word adaptation surprisal as methods of measuring mutual intelligibility in reading comprehension of Slavic languages. *Computational Linguistics and Intellectual Technologies: International Conference "Dialogue 2017" Proceedings*. Issue 16(23), vol. 1. 304–317.

Stichting Lezen en Schrijven. 2021. Laaggeletterdheid in Nederland [Low literacy in the Netherlands]. *Kennisblad Laaggeletterdheid in Nederland*. https://www.lezenenschrijven.nl/informatie-over-laaggeletterdheid-nederland (accessed 7 April 2024).

Sunderland, Jane. 2018. Research questions in linguistics. In Lia Litosseliti (ed.), *Research methods in linguistics*, 13–34. London: Bloomsbury Academic.

Sussex, Roland & Paul Cubberley. 2006. *The Slavic languages*. Cambridge: Cambridge University Press.

Swadesh, Morris. 1971. *The origin and diversification of language: Edited post mortem by Joel Sherzer*. Chicago, IL: Aldine.

Swain, Merrill. 2000. The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In James P. Lantolf (ed.), *Sociocultural theory and second language learning*, 97–114. Oxford: Oxford University Press.

Swarte, Femke. 2016. *Predicting the mutual intelligibility of Germanic languages from linguistic and extra-linguistic factors*. Groningen: University of Groningen dissertation.

Swarte, Femke, Anja Schüppert & Charlotte Gooskens. 2015. Does German help speakers of Dutch to understand written and spoken Danish words? – The role of second language knowledge in decoding an unknown but related language. In Gessica De Angelis, Ulrike Jessner & Marijana Kresic (eds.), *Crosslinguistic influence and crosslinguistic interaction in multilingual language learning*, 173–197. London: Bloomsbury Publishing.

Swerts, Marc & René Collier. 1992. On the controlled elicitation of spontaneous speech. *Speech Communication* 11(4–5). 463–468.

Syrdal, Ann K., H. Timothy Bunnell, Susan R. Hertz, Taniya Mishra, Murray Spiegel, Corine Bickley, Deborah Rekart & Matthew J. Makashay. 2012. Text-to-speech intelligibility across speech rates. *Proceedings of the 13th annual conference of the international speech communication association (Interspeech 2012)*, 623–626.

Szubko-Sitarek, Weronika. 2011. Cognate facilitation effects in trilingual word recognition. *Studies in second language learning and teaching* 1(2). 189–208.

Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages: A comparative handbook*, 55–75. Berlin, New York: De Gruyter Mouton.

Tamburelli, Marco. 2014. Uncovering the 'hidden' multilingualism of Europe: An Italian case study. *Journal of Multilingual and Multicultural Development* 35(3). 252–270.

Tamburelli, Marco & Mauro Tosco. 2021. *Contested languages: The hidden multilingualism of Europe*. Amsterdam: John Benjamins.

Tang, Chaoju & Vincent J. van Heuven. 2008. Mutual intelligibility of Chinese dialects tested functionally. In Marjo v. Koppen & Bert Botma (eds.), *Linguistics in the Netherlands*, 145–156. Amsterdam: John Benjamins.

Tang, Chaoju & Vincent J. van Heuven. 2009. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua* 119(5). 709–732.

Tang, Chaoju & Vincent J. van Heuven. 2015. Predicting mutual intelligibility of Chinese dialects from multiple objective linguistic distance measures. *Linguistics* 53(2). 285–312.

Teleman, Ulf. 1987. Om grannspråksförståelse. Hinder och möjligheter [On understanding neighboring languages. Problems and possibilities]. In Else Bojsen (ed.), *Språk i Norden*, 70–82. Oslo: Novus.

Thije, Jan D. ten 2018. Lecture 12: Receptive multilingualism. In David Singleton & Larissa Aronin (eds.), *Twelve lectures on multilingualism*, 329–364. Bristol: Blue Ridge Summit: Multilingual Matters.

Thorvaldsson, A. 2011. *Frozen assets: How I lived Iceland's boom and bust*. Chichester: John Wiley & Sons.

Tielen, Mirjam. 1992. *Male and female speech: An experimental study of sex-related voice and pronunciation characteristics*. Amsterdam: University of Amsterdam dissertation.

Torp, Arne. 1998. *Nordiske språk i nordisk og germansk perspektiv* [Nordic languages in Nordic and Germanic perspective]. Oslo: Novus.

Trecca, Fabio, Kristian Tylén, Anders Højen Morten H. & Christiansen. 2021. Danish as a window onto language processing and learning. *Language Learning* 71(3). 799–833.

Tripathi, Kartik. 2016. On ausbausprachen and abstandsprachen: How do you define a language? *Unravel Magazine*. https://unravellingmag.com/articles/ausbausprachen-abstandsprachen-define-language/ (accessed 7 April 2024).

Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.

Trudgill, Peter. 2000. *Sociolinguistics: An introduction to language and society*. London: Penguin.

Vanhove, Jan. 2014. *Receptive multilingualism across the lifespan: Cognitive and linguistic factors in cognate guessing*. Freiburg: Universität Freiburg in der Schweiz dissertation.

Vanhove, Jan. 2016. The early learning of interlingual correspondence rules in receptive multilingualism. *International Journal of Bilingualism* 20(5). 580–593.

Vanhove, Jan & Raphael Berthele. 2015. Item-related determinants of cognate guessing in multilinguals. In Gessica De Angelis, Ulrike Jessner & Marijan Kresić (eds.), *Crosslinguistic influence and crosslinguistic interaction in multilingual language learning*, 95–118. London: Bloomsbury.

Verhoeven, Jo, Guy de Pauw & Hanne Kloots. 2004. Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech* 47(3). 297–308.

Verschik, Anna. 2012. Practising receptive multilingualism: Estonian–Finnish communication in Tallinn. *International Journal of Bilingualism* 16(3). 265–286.

Vetter, Eva. 2012. Exploiting receptive multilingualism in institutional language learning: The case of Italian in the Austrian secondary school system. *International Journal of Bilingualism* 16(3). 348–365.

Voegelin, Charles F. & Zellig S. Harris. 1951. Methods for determining intelligibility among dialects of natural languages. *Proceedings of the American Philosophical Society* 95(3). 322–329.

Walker, Abby, Katie Drager & Jennifer Hay. 2022. The use of priming in language attitudes research. In Ruth Kircher & Lena Zipp (eds.), *Research methods in language attitudes*, 313–329. Cambridge: Cambridge University Press.

Wang, Hongyan. 2007. *English as a lingua franca: Mutual intelligibility of Chinese, Dutch and American speakers of English*. Utrecht: University of Utrecht dissertation.

Wang, Hongyan, Ligang Zhu, Xiaotong Li & Vincent J. van Heuven. 2011. Relative importance of tone and segments for the intelligibility of Mandarin and Cantonese. In Wai S. Lee & Eric Zee (eds.), *Proceedings of the 17th international congress of phonetic sciences*, 2090–2093. Hong Kong: City University of Hong Kong.

Wardhaugh, Ronald & Janet M. Fuller. 2021. *An introduction to sociolinguistics*. Hoboken, NJ: John Wiley & Sons.

Weller, Marion, Fabienne Cap, Stefan Müller, Sabine Schulte im Walde & Alexander Fraser. 2014. *Distinguishing degrees of compositionality in compound splitting for statistical machine translation*. Dublin: Dublin City University.

Wendt, Dorothea, Birger Kollmeier & Thomas Brand. 2015. How hearing impairment affects sentence comprehension: Using eye fixations to investigate the duration of speech processing. *Trends in Hearing* 19. 1–18.

Wenzel, Veronika. 2018. L2 status factor en netwerken binnen de Germaanse taalfamilie [L2 status factor and networks within the Germanic language family]. *Internationale Neerlandistiek* 56(3). 243–263.

Wesche, Marjorie B. 1994. Input and interaction in second language acquisition. In Clare Gallaway & Brian J. Richards (eds.), *Input and interaction in language acquisition*, 219–250. Cambridge: Cambridge University Press.

Wichmann, Søren. 2019. How to distinguish languages and dialects. *Computational Linguistics* 45(4). 823–831.

Wieling, Martijn, Eliza Margaretha & John Nerbonne. 2012. Inducing a measure of phonetic similarity from pronunciation variation. *Journal of Phonetics* 40(2). 307–314.

Wilde, Vanessa de, Marc Brysbaert & June Eyckmans. 2020. Learning English through out-of-school exposure: How do word-related variables and proficiency influence receptive vocabulary learning? *Language Learning* 70(2). 349–381.

Wolff, Hans. 1959. Intelligibility and inter-ethnic attitudes. *Anthropological Linguistics* 1(3). 34–41.

Yang, Cathryn. 2012. Classifying Lalo languages: Subgrouping, phonetic distance, and intelligibility. *Linguistics of the Tibeto-Burman Area* 35(2). 113–137.

Yang, Cathryn & Andy Castro. 2008. Representing tone in Levenshtein distance. *International Journal of Humanities and Arts Computing* 2(1–2). 205–219.

Yang, Changyong, William O'Grady, Sejung Yang, Nanna H. Hilton, Sang-Gu Kang & So-Young Kim. 2020. Revising the language map of Korea. In D. B. Stanley & Roland Kehrein (eds.), *Handbook of the changing world language map*, 1–15. Cham: Springer Cham.

Yenkimaleki, Mahmood. 2017. *Effect of prosody awareness training on the quality of consecutive interpreting between English and Farsi*. Leiden: Leiden University dissertation.

Yoder, Zachariah. 2017. The reliability of recorded text test scores: Widespread inconsistent intelligibility testing in minority languages. *Journal of Multilingual and Multicultural Development* 38(9). 843–855.

Zahn, Christopher J. & Robert Hopper. 1985. Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology* 4(2). 113–123.

Zeevaert, Ludger. 2004. *Interskandinavische Kommunikation. Strategien zur Etablierung von Verständigung zwischen Skandinaviern im Diskurs*. Hamburg: Dr. Kovač.

Zeevaert, Ludger. 2007. Receptive multilingualism and inter-Scandinavian semicommunication. In Ludger Zeevaert & Jan D. Ten Thije (eds.), *Receptive multilingualism. Linguistic analyses, language policies and didactic concepts*, 103–135. Amsterdam: John Benjamins.

Zekveld, Adriana A., Sophia E. Kramer & Joost M. Festen. 2010. Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing* 31(4). 480–90.

Zipf, George K. 1972. *Human behavior and the principle of least effort: An introduction to human ecology*. New York, NY: Hafner.

Zissman, Amy & Edith Neimark. 1990. The influence of familiarity on evaluations of liking and goodness of several types of music. *The Psychological Record* 40. 481–490.

Zuckerman, Miron & Robert E. Driver. 1989. What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior* 13. 67–82.

Index

https://doi.org/10.1515/9783111134697-012