# 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 /ts/ like in Dutch *tijd* /tɛit/ – German *Zeit* /tsait/ 'time', they can generalize this knowledge to other words such as Dutch *tekenen* /tekənə/ – German *zeichnen* /tsaixnən/ 'draw' and Dutch *twee* /tʋe:/ German *zwei* /tsʋai/ '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-

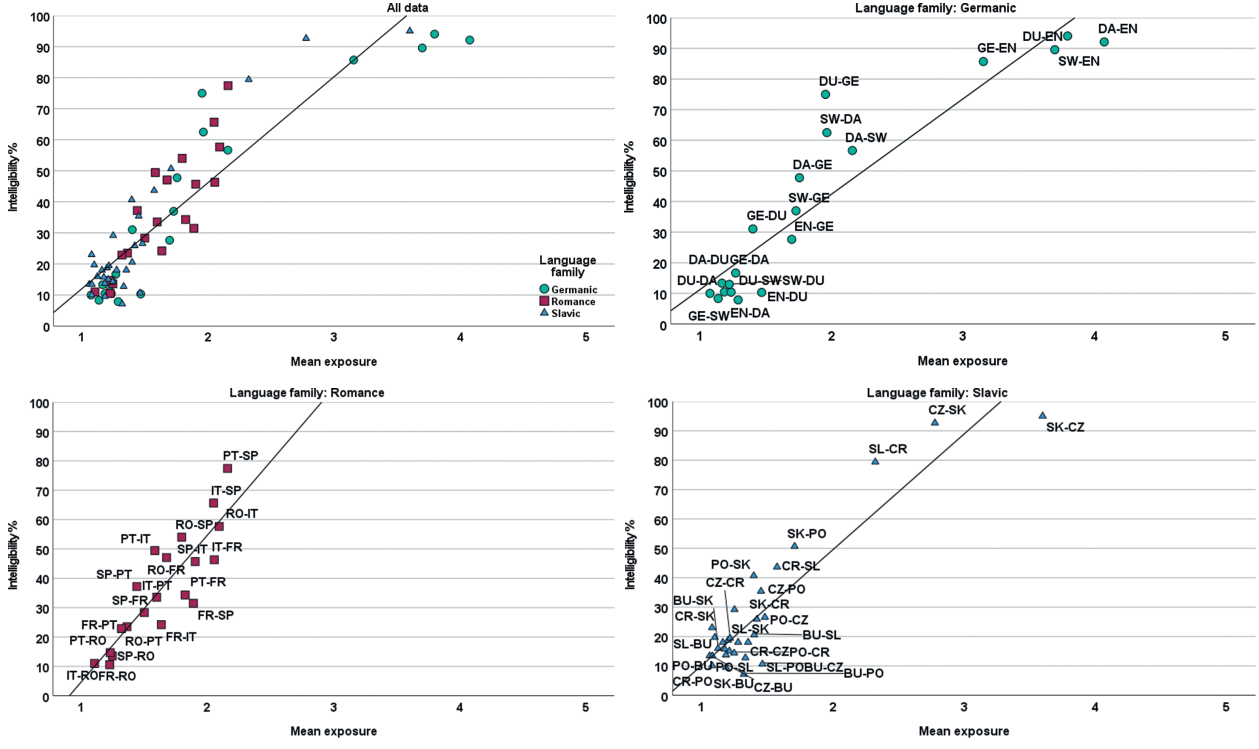


**Figure 4.2:** Correlation between intelligibility scores and mean exposure scores for all 70 European language pairs in the MICReLa project ($r$ = .90) and separately for Germanic ($r$ = .93), Romance ($r$ = .87), and Slavic ($r$ = .92). The abbreviations refer to the language varieties involved (see 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* [fɔˈtølj] 'armchair', which is a loanword from French *fauteuil* [foˈtœj], 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-

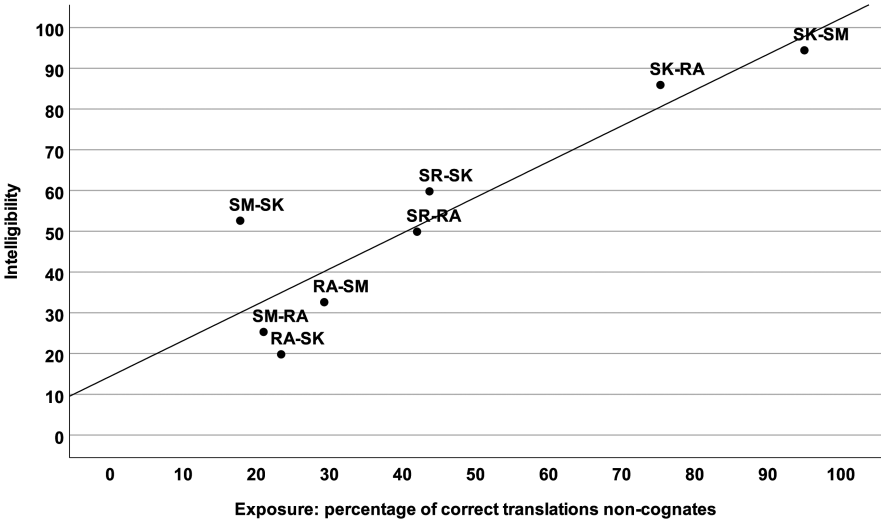


**Figure 4.3:** Exposure expressed as the percentages of correctly translated non-cognates and intelligibility measured in a word translation task ($r$ = .92). The abbreviations refer to the language varieties involved (RA = Raga, SK = Suru Kavian, SM = Suru Mwerani, SR = Rabwanga), see also the map 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 influ-

enced 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-

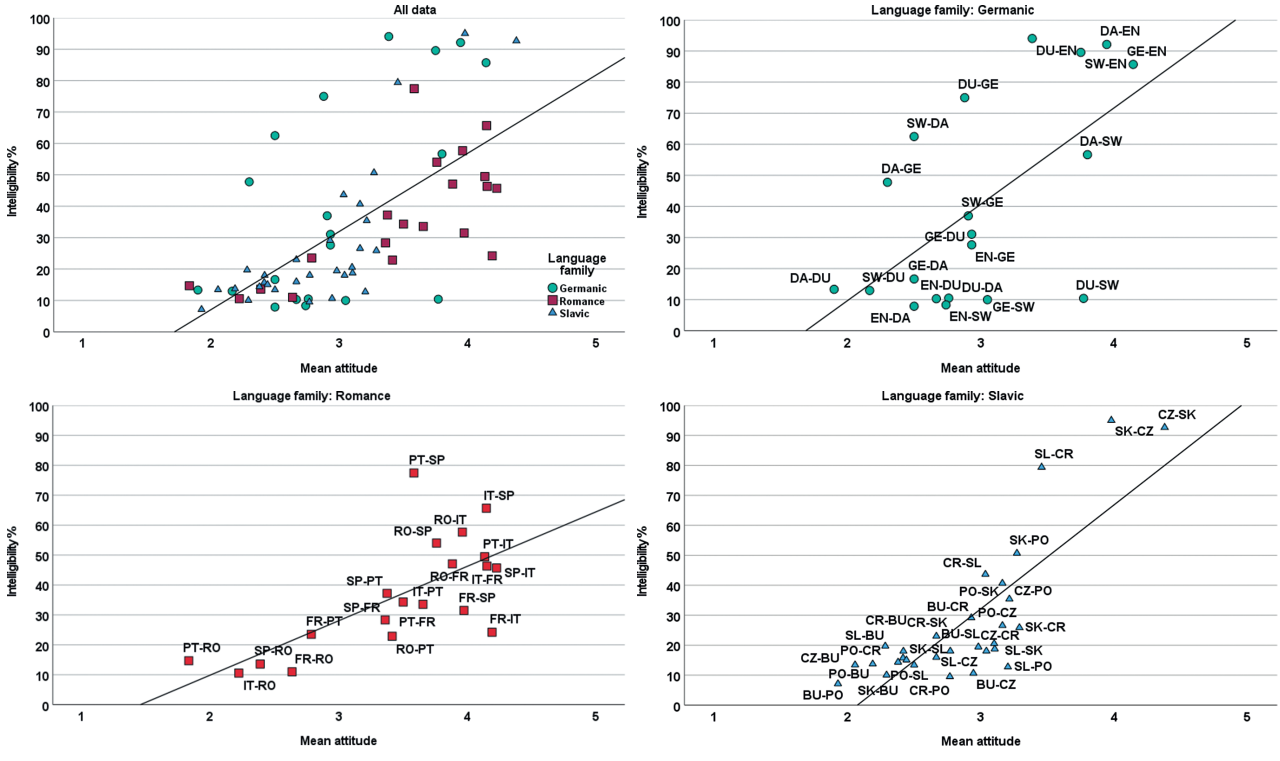


**Figure 4.5:** Correlation between intelligibility scores and mean attitude scores for all 70 European language pairs in the MICReLa project ($r$ = .65), and separately for Germanic ($r$ = .60), Romance ($r$ = .70), and Slavic ($r$ = .81). For explanations, see Figure 4.2.

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 [hat:] in Swedish. When listening to Swedish [hat:], 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] | [hat:] | 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 transla-

tions) 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.