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.

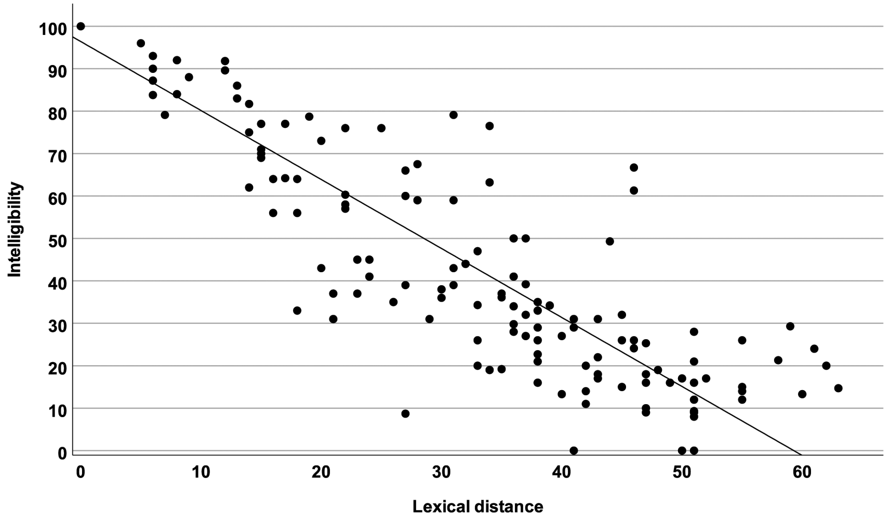

Figure 5.1: Scattergram showing the relationship between intelligibility scores and lexical distance scores in 8 studies. Reproduced from Simons (1979: 88).

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

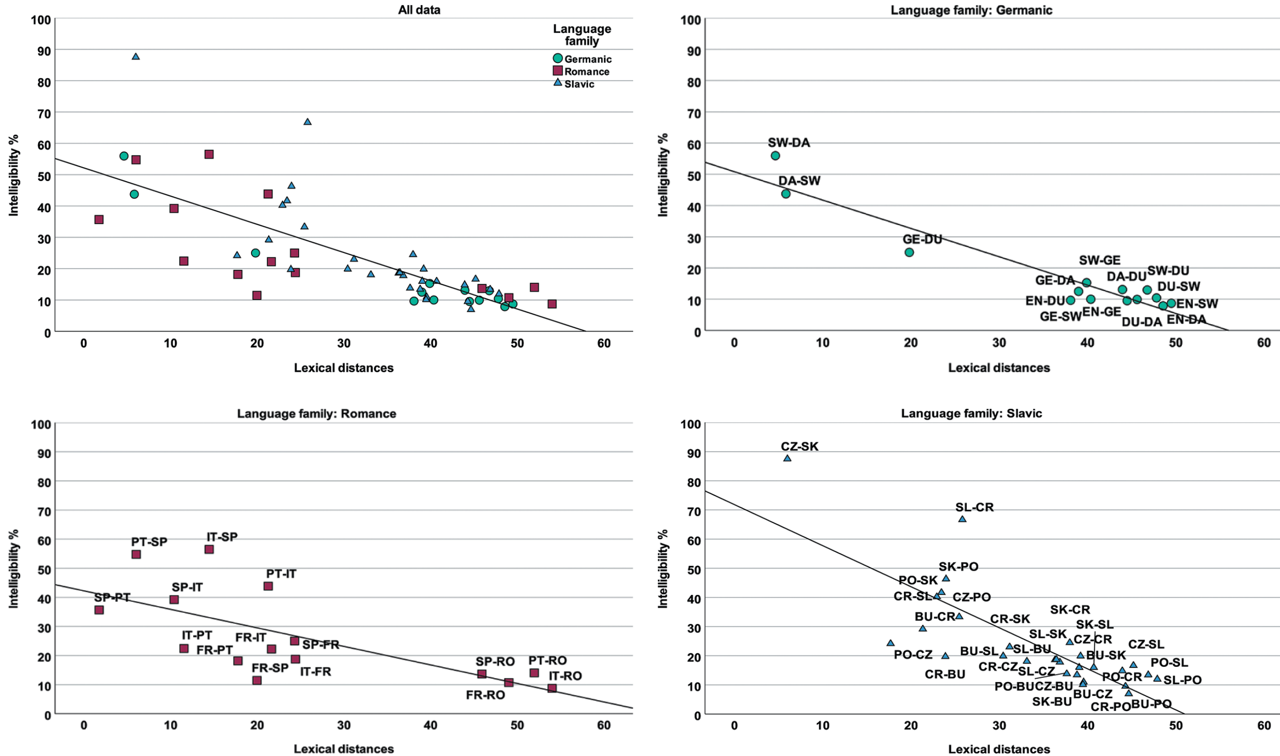

Figure 5.2: Correlations between inherent intelligibility scores and lexical distances for 57 European language pairs in the MICReLa project ($r = -.76$), and separately for Germanic ($r = -.95$), Romance ($r = -.69$) and Slavic ($r = -.80$). For explanations, see Figure 4.2.

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).

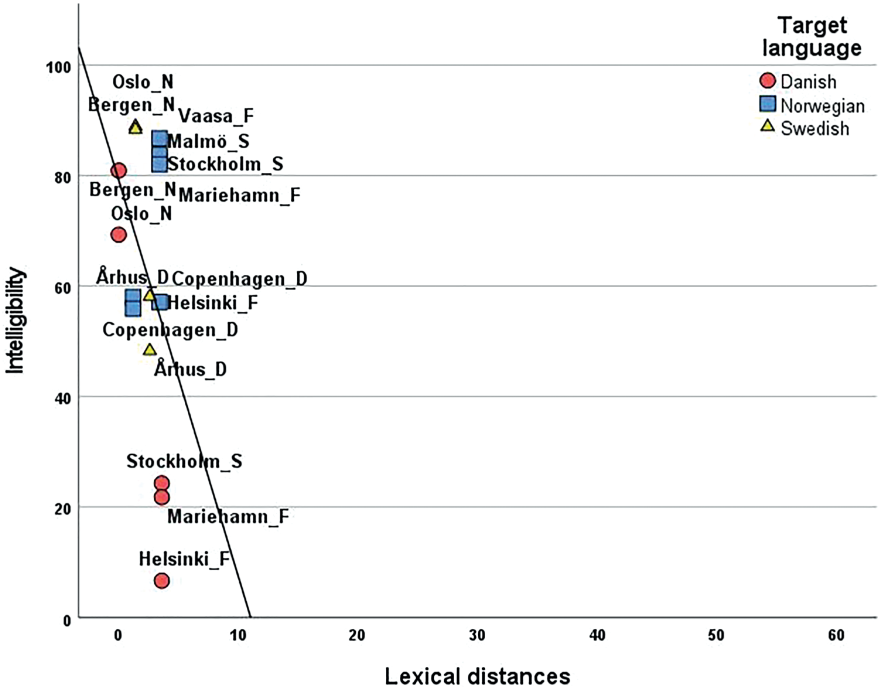

Figure 5.3: The correlation between intelligibility (percentages of correct answers in a cloze test) and lexical distances (percentage non-cognates), $r = -.42$, $p > .11$, from an investigation of Standard Danish, Standard Norwegian (represented by the Oslo variety) and Standard Swedish for listeners 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

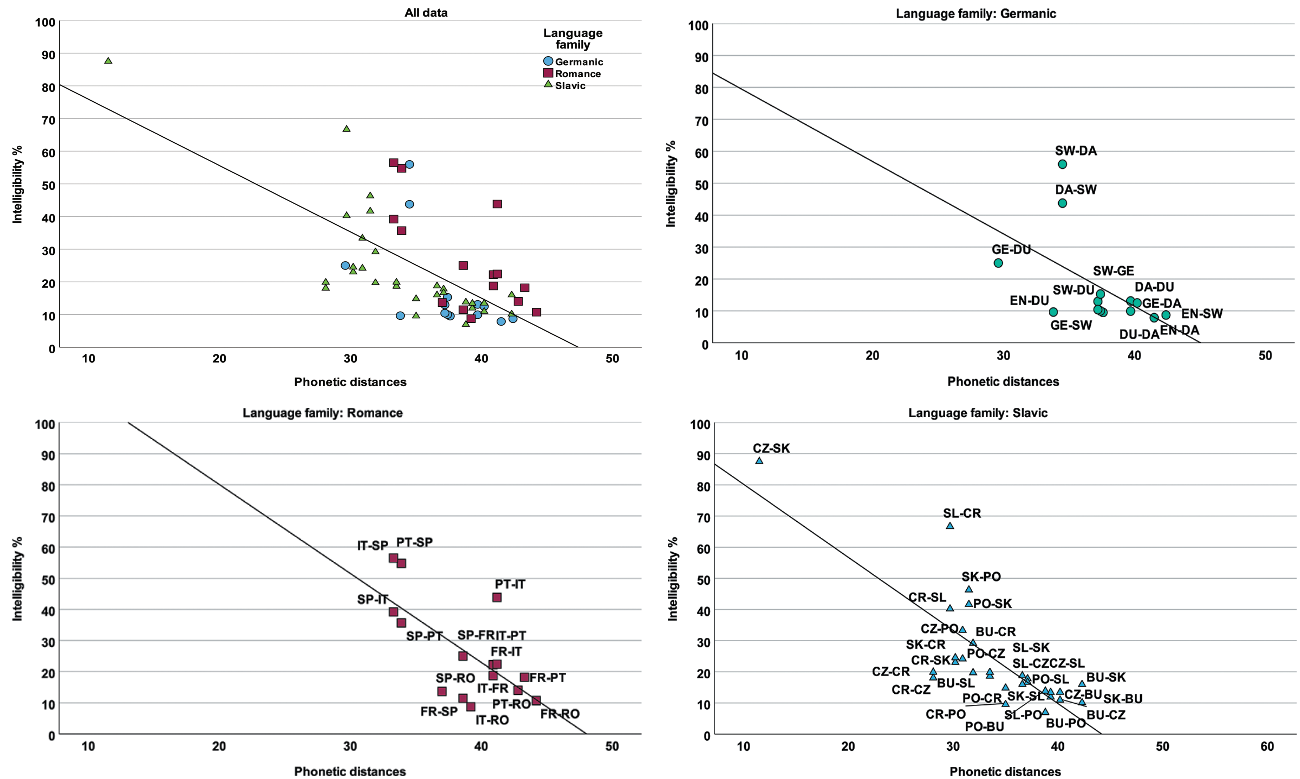

Figure 5.5: Correlations between inherent intelligibility scores and phonetic distances for 57 European language pairs in the MICReLa project ($r = -.66$), and separately for Germanic ($r = -.53$), Romance ($r = -.68$) and Slavic ($r = -.62$ without the Czech-Slovak outlier). For explanations, see Figure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

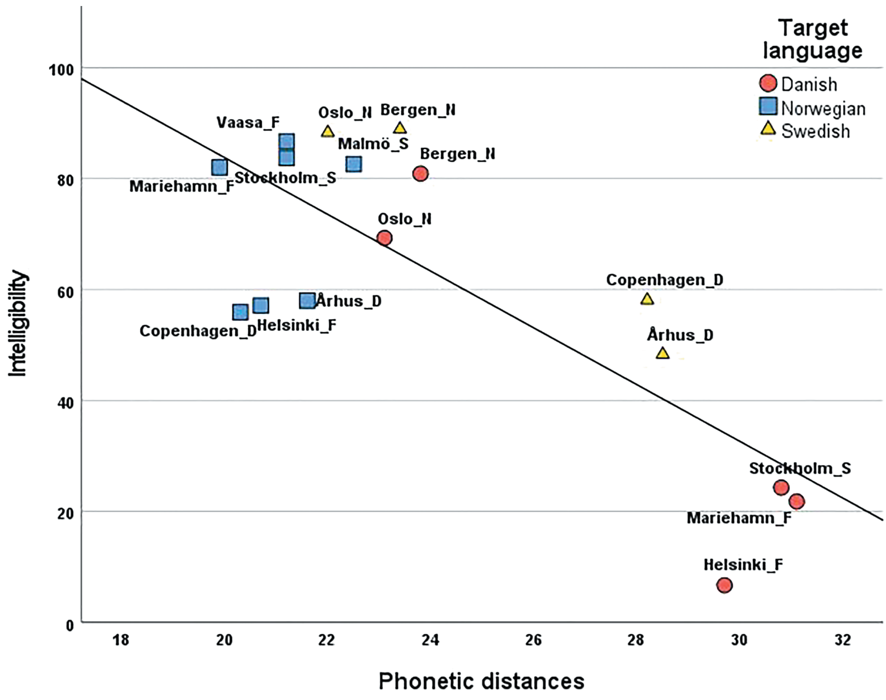

Figure 5.6: Correlation between intelligibility (percentage of correct responses in a cloze test) and phonetic distances (Levenshtein distance), $r = -.80$, $p < .01$, from an investigation of the intelligibility of Standard Danish, Standard Norwegian (represented by the Oslo variety) and Standard Swedish 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* (1^{st} person singular or imperative)*, werkt* (2^{nd} and 3^{rd} person singular)*, werken* (infinitive, $1^{st,}$ 2^{nd} and 3^{rd} 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.

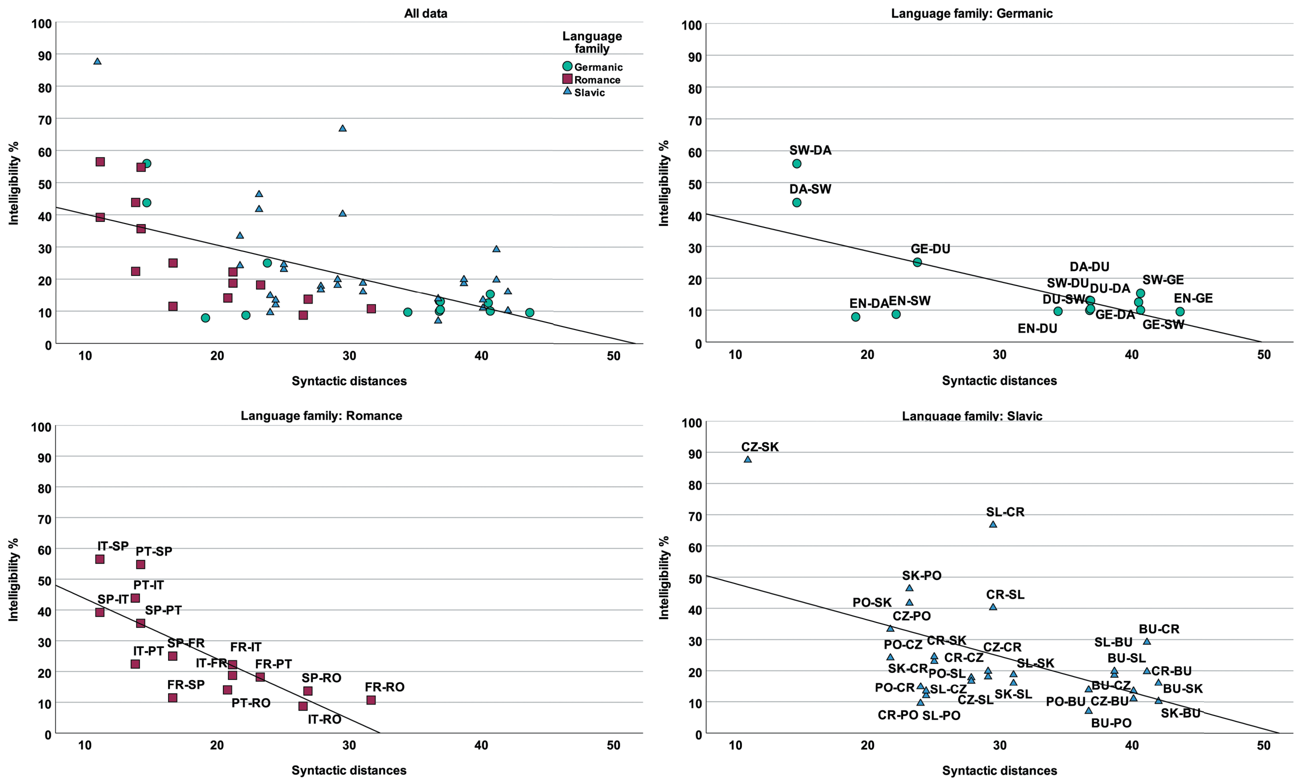

Figure 5.9: Correlation between inherent intelligibility scores and syntactic distances for 57 European language pairs in the MICReLa project ($r = -.56$), and separately for Germanic ($r = -.68$), Romance ($r = -.77$) and Slavic ($r = -.52$). For explanations, see Figure 4.2.

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.

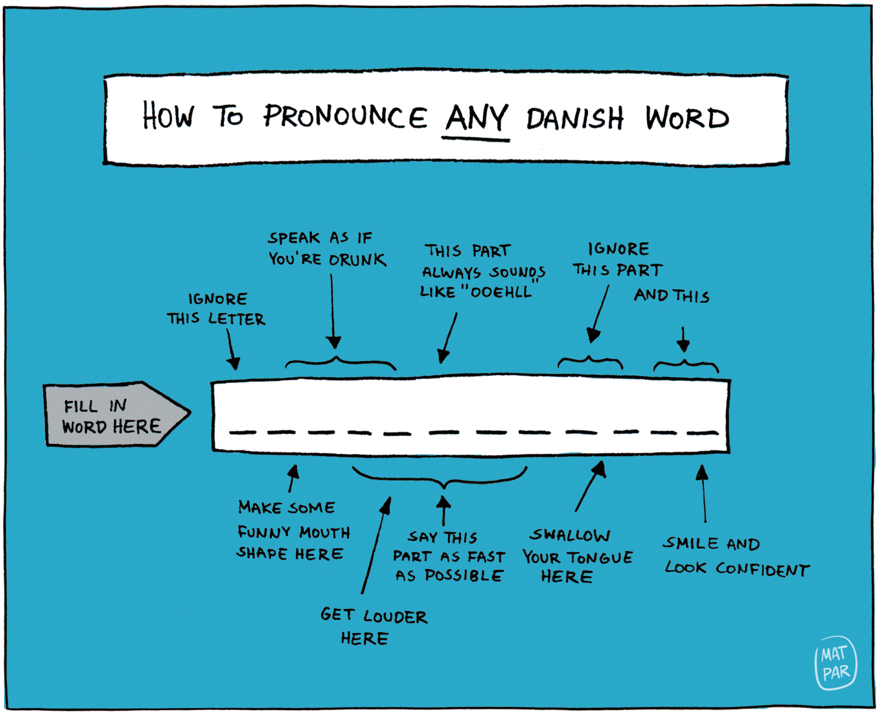

Figure 5.10: A caricature of the Danish language, showing the impression that many Scandinavians (including the Danes themselves) have of Danish as a language with a lot of reduction and unclear pronunciation. Made for the Copenhagen Language Center, copyright Matthias Parchettka.

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.

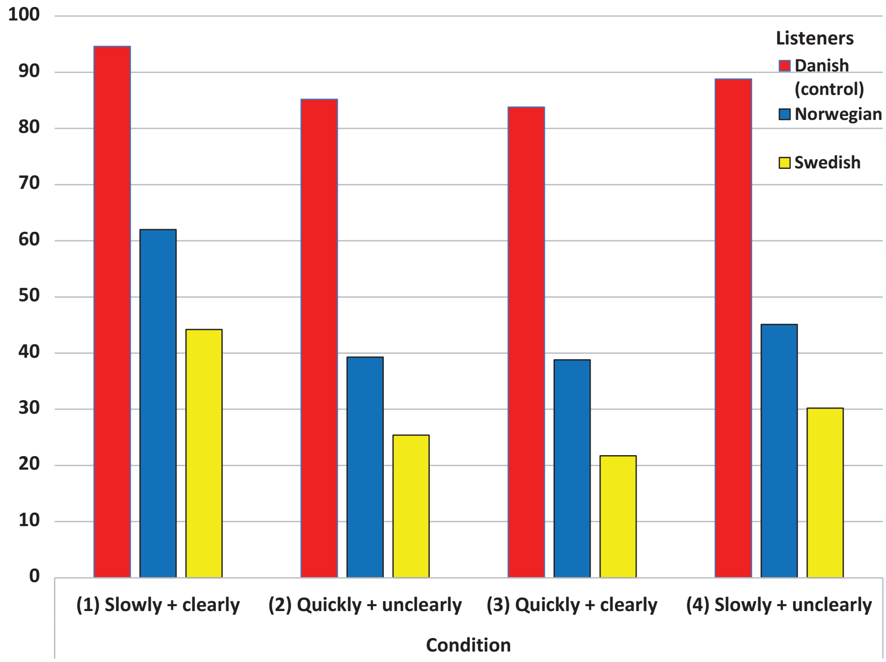

Figure 5.11: Mean intelligibility per listener group in four conditions: (1) slowly and clearly, (2) quickly and unclearly, (3) quickly and clearly, and (4) slowly and unclearly. Based on results from Schüppert, Hilton, and Gooskens (2016: 55).

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.

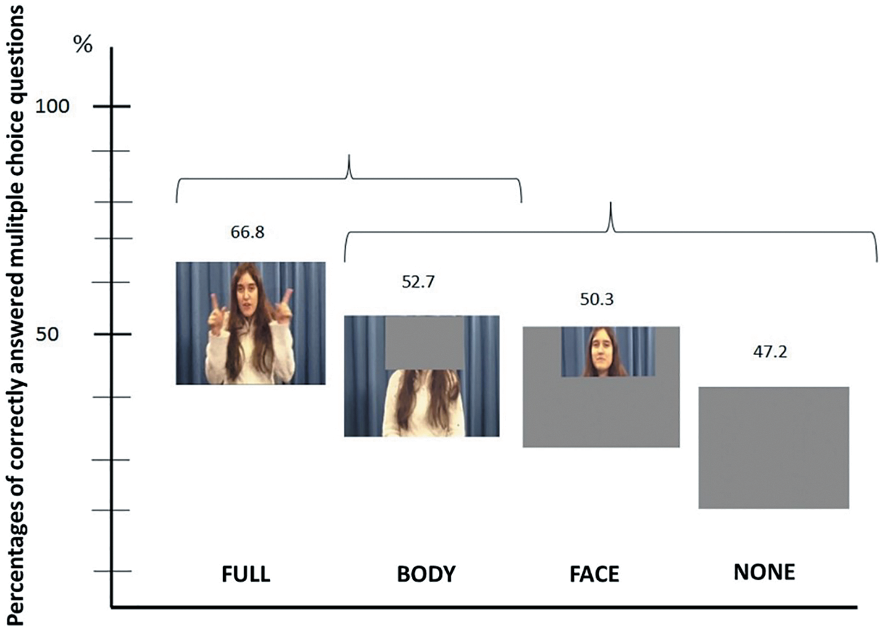

Figure 5.12: Percentage of correctly answered multiple-choice questions for four experimental conditions. Braces include conditions that do not differ from each other significantly ($p< .01$).

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.