How easy is it for speakers of Dutch to understand Frisian and Afrikaans, and why?

Renée van Bezooijen and Charlotte Gooskens
Radboud University Nijmegen / University of Groningen

1. Introduction

To what extent are speakers of related languages and language varieties able to communicate with each other in their own language? This question was first addressed in a series of studies of the mutual intelligibility of native Indian languages in the United States (e.g. Pierce 1952). Many other languages were to follow, such as Spanish and Portuguese (Jensen 1989), Slovak and Czech (Budovičová 1987), and Scandinavian languages (e.g. Maurud 1976, Börestam Uhlmann 1994). Results were generally explained in terms of language distance, language attitude and language contact: the smaller the distance, the more positive the attitude and the more frequent the contact, the more successful interlingual communication was assumed to be. However, the size and exact nature of the contributions of these three factors have hardly been tested experimentally.

In the present study the intelligibility of Frisian and Afrikaans for speakers of Dutch was investigated, both in written and spoken form. Just as in previous research, our study deals with related languages, all three belonging to the West-Germanic branch. However, in contrast to the languages examined in previous studies, contacts between these three languages are rare. For Dutch and Afrikaans this is self-evident, considering the large geographic distance between the Netherlands and South Africa. The situation for Frisian is more complicated. Frisian is the second officially recognized language in the Netherlands, next to Dutch. It is spoken in the northern province of Fryslân, by about half of the inhabitants. Frisian is mainly used in the countryside in the informal domains, i.e. within the family, with friends and with neighbors. With

Linguistics in the Netherlands 2005, 13–24.
ISSN 0929–7332 / E-ISSN 1569–9919 © Algemene Vereniging voor Taalwetenschap

non-Frisians, Frisians automatically switch to Dutch. Many tourists visiting Fryslân may never hear a Frisian-speaking Frisian. The chance that they will see it in its written form is even smaller.

Just like the majority of Dutchmen, the subjects participating in the present intelligibility study had no active and at the most (incidentally acquired) passive knowledge of Frisian and Afrikaans, so the factor of language contact could safely be assumed not to play a role. The role of language attitude was investigated in a previous study (Van Bezooijen and Gooskens, 2005); no correlation with subjects' reading performance was found. Therefore, the extent to which Frisian and Afrikaans are intelligible for speakers of Dutch seems to be completely determined by linguistic factors. These factors were first explored in Van Bezooijen and Gooskens (2005). However, that study was limited to the comprehensibility of written texts. The present study in addition examines the comprehensibility of spoken texts and words. Moreover, in contrast to the previous study, a reference condition is included, in the sense that the performance of Dutch-speaking subjects for Frisian and Afrikaans is compared with their performance for their own language.

2. Method

The intelligibility of written Afrikaans and Frisian texts was measured by means of a variant of the cloze test. The intelligibility of the two languages in their spoken form was determined at two levels; subjects had to answer a number of open questions pertaining to complete texts and they had to translate a list of isolated words. More information on the three tests is provided in Section 2.1. In Section 2.2 the measurement of the linguistic distances is explained.

2.1 Measuring intelligibility

Subjects

The subjects participating in the study were 67 pupils (27 boys and 40 girls) in their pre-final year of grammar school. The schools were located in various regions of the Netherlands (Coevorden, Zwolle, The Hague, and Meppel). The subjects' mean age was 16.3 years. They had no active and no, or not more than a very superficial, passive knowledge of Frisian or Afrikaans. All subjects spoke Dutch as their mother tongue.

Tasks

As a basis for the assessment of the intelligibility of written text we used two Dutch newspaper articles with an average level of difficulty.[1] One article ('the dating text') was about dating agencies and comprised 329 words; the other ('the feminist text') dealt with the image of women created by modern music stations and consisted of 317 words. Intelligibility was assessed by means of a variant of the cloze test. In both texts, five nouns, five adverbs, five adjectives, and five verbs were selected at random. These were placed in alphabetic order above the text and replaced by blanks in the text. Next, the two texts were translated into Frisian and Afrikaans and the same words were removed and placed above the texts. The subjects were given ten minutes to put the 20 words back in the right place in the texts. The percentage of words placed back correctly was taken as a measure of the intelligibility of the written texts.

To assess the intelligibility of a running spoken text, we replicated an experiment currently being run to test the mutual intelligibility of the Scandinavian languages.[2] In this experiment and in the present study, use was made of the same two texts. One ('the frog test') was about counting frogs to determine the quality of the environment (233 words) and the other ('the kangaroo text') about a run-away kangaroo in a big European city (277 words). The texts were translated from Danish into Dutch, Frisian, and Afrikaans and read aloud by native speakers of these three languages. For either text, there were five open questions. The subjects wrote down their answers while listening to the recordings. Three degrees of correctness were distinguished: completely correct, half correct (for example if a plural form was responded with instead of the singular form), and not correct. The points were summed, divided by 10 and multiplied by 100 to obtain the percentage of correct answers.

To test the intelligibility of spoken words under minimal conditions, i.e. without the help of a meaningful context, we constructed a list of 19 nouns referring to everyday objects such as 'chair', 'cow', 'foot', and 'coat'. These words were read onto tape by native speakers of the three languages and presented to the subjects to be translated into Dutch. Correctness was calculated in the same manner as for the spoken texts.

Design

Half of the subjects were tested for their understanding of written and spoken Frisian and the other half for their understanding of written and spoken Afrikaans. As a reference condition, all subjects were tested for their understanding of their mother tongue (Dutch). Half of the subjects read the feminist text in

Frisian (or Afrikaans) and the dating text in Dutch, and the other half the other way round. Likewise, half of the subjects heard the frog text in Frisian (or Afrikaans) and the kangaroo text in Dutch, and the other half the other way round. The isolated words were presented in different orders in the two languages. The Frisian (or Afrikaans) part of the test was always presented before the Dutch part. Within each part, first the spoken texts were presented, next the isolated spoken words, and finally the written texts. A short recording of a weather forecast preceded the Frisian (or Afrikaans) test, so that the subjects could get used to listening to the language.

2.2 Measuring linguistic distance

To calculate linguistic distances, we first aligned the Dutch texts with the Frisian and Afrikaans texts. In a few cases we adapted the word order to obtain a better alignment. The aligned word pairs formed the basis for four distance measures. The first three measures express the nature of the relationship between the Dutch and the Frisian/Afrikaans words. The fourth measure, the so-called Levenshtein distance, expresses the degree of phonetic or orthographic similarity between corresponding words. All distances were calculated both for the reading and the listening texts. Because we assumed that content words (nouns, adjectives, numerals, main verbs) are more important for intelligibility than function words (articles, conjunctions, prepositions, pronouns, auxiliaries, modals, particles, adverbs), distances were calculated separately for these two word categories. For the isolated words, not all measures were relevant (see 3.3). The four distance measures will now be explained in further detail.

Percentage of cognates
A large proportion of cognates, i.e. words in two languages with a common root, may be expected to facilitate comprehension. However, a direct relationship between words is not a necessary condition for mutual intelligibility. In some cases, the meaning of a word can be deduced via a cognate synonym. For example, the Dutch word *samenleving* ('society') in the original newspaper article was translated by Frisian *maatskippij*. These two words are non-cognates. Nevertheless, the Dutch reader can easily understand the Frisian word *maatskippij* because of the existence of the Dutch synonym *maatschappij*. The percentage of cognates, either related directly or via a synonym, constitutes the first linguistic distance measure.

Percentage of paradigm-related cognates
It is also possible to deduce the meaning of a word paradigmatically. For example, the Frisian translation of the Dutch word *zijn* ('are' plural present tense) is *binne*. These two words are not related, neither directly nor via a synonym. However, a Dutch reader may nevertheless understand the meaning of the Frisian word because it is related to the Dutch word *ben* ('am'), which belongs to the same paradigm as *zijn*. This measure can also operate through a synonym. The percentage of cognates related via a paradigm, either directly or via a synonym, constitutes the second distance measure. Most of the words in this category are function words.

Percentage of non-cognates
It should be impossible to deduce the meaning of a word in an unknown language if it bears no relationship with a word in the mother tongue.[3] The percentage of non-cognates should therefore be an important indicator of mutual intelligibility. It is the complement of the first two measures, but for the sake of completeness we present all three.

Levenshtein distance
In most cases, a non-linguistically trained person will be able to deduce the meaning of a cognate. However, some cognates are easier to recognize than others, depending on the present-day degree of similarity. Due to subsequent speech changes and/or spelling conventions, the relatedness between words might no longer be discernible. To estimate the transparency of word relatedness, we assessed the degree of similarity between word forms by means of the so-called Levenshtein distance. This measure can be calculated automatically by computer.

The distance between corresponding words was based upon the minimum number of letters or sounds that need to be inserted, deleted or substituted in order to transform the word in the one language into the corresponding word in the other language. All operations were given an equal weight of 1 point. Word length was compensated for by dividing the total sum of costs by the number of alignments of letters. We refer to Heeringa (2004) for a more extensive explanation. As an example we present the calculation of the distance between the written Dutch word *zoeken* ('search') and the Frisian word *sykje*:

alignments	1	2	3	4	5	6	7
Dutch	z	o	e	k	–	e	n
Frisian	s	y	–	k	j	e	–
costs	1	1	1		1		1

The sum of costs (1+1+1+1+1 = 5) is divided by the number of alignments (7). The result is a distance of 71%. The total distance between two languages is the mean distance over all word pairs. In the case of the spoken texts, the distances were calculated on the basis of phonetic transcriptions rather than letters. Moreover, a different procedure was applied. Only substitutions of a consonant by a vowel or a vowel by a consonant was assigned 1 point. Substitutions of a consonant by another consonant and of a vowel by another vowel was assigned 0.5 points.

3. Results

In Table 1 the results of the three intelligibility tests are shown. We see that the subjects, with 97.1%, 89.4% and 98.4% correct, perform almost maximally when taking the tests in their own language, Dutch. Apparently the tasks were not too difficult. It furthermore appears that it was easier for the subjects to understand written Afrikaans (66.4% correct) than written Frisian (31.9%). The difference is significant at the .01 level. In a previous study (Van Bezooijen and Gooskens 2005) with the same texts but — older — language students, basically the same result was obtained, although the scores for both languages were higher (81.8% and 50.3%, respectively). The intelligibility of the spoken words shows the same trend, namely a significantly higher percentage of correct answers for Afrikaans (64.0%) than for Frisian (45.6%). On the other hand, when intelligibility of spoken language is measured for longer texts by means of open questions the percentage correct is higher for Frisian (63.9%) than for Afrikaans (59.4%). However, the difference is not significant. In the next section we will look for explanations for the results in the linguistic distances between the languages.

Table 1. Percentage of correct answers for the three tests and the results of an independent sample *t*-test between Frisian and Afrikaans

				t-test		
	Dutch	Frisian	Afrikaans	*t*	*df*	*p*
written text	97.1	31.9	66.4	8.00	65	.000
spoken text	89.4	63.9	59.4	–1.30	65	.197
spoken words	98.4	45.6	64.0	8.31	65	.000

3.1 Intelligibility of written text

Assuming that intelligibility is inversely related to linguistic distance, we expect the linguistic distances between the Dutch and Frisian reading texts to be larger than those between the Dutch and Afrikaans reading texts. In Table 2 the results of the first three linguistic distance measures are presented. We see that the percentage of cognates (related directly or via a synonym) is almost identical for the content words (94.1% versus 94.6%). However, there is a marked difference in the function words. Whereas in Frisian, almost all function words (93.4%) are related directly to their Dutch counterparts, Afrikaans has relatively many function words (23.7%) that are related to Dutch via a paradigm.

Furthermore (not shown in the table), the nature of the cognates differs between Frisian and Afrikaans; the Afrikaans texts have a larger proportion of words that are related via a synonym than the Frisian texts. The connotative meaning of these synonyms sometimes deviates from that of the original Dutch words. For example, the Dutch word *opeens* 'suddenly' was translated by Afrikaans *skielik*. Dutch readers probably interpret this word correctly via the Dutch synonym *schielijk*. However, Dutch *schielijk* is less frequent and slightly archaic compared to *opeens*. Similarly, we are sure that a Dutch reader will correctly interpret Afrikaans *metgesel* 'partner' via Dutch *metgezel*, as the two words have the same referential meaning. However, Dutch *metgezel* is somewhat old-fashioned and does not fit the modern context of dating agencies.

So, both for the function words and the content words the relationship with the Dutch counterparts appears to be more direct for Frisian than for Afrikaans. However, we do not know to what extent this affects intelligibility.

What we do know is that a large proportion of non-cognates that cannot be interpreted by a synonym in the subjects' own language must have a negative effect on intelligibility. The percentage of non-cognates is higher for Frisian than for Afrikaans, both for function words (2.0% and 0.8%) and for content words (5.9% and 3.7%). This does not only hold for the number of tokens (24 and 14, respectively) but also for the number of types (20 and 10, respectively). These differences between Afrikaans and Frisian might seem small, but it should be borne in mind that one single unintelligible word can make a whole sentence or even a complete paragraph incomprehensible.

As mentioned before, it is not sufficient to look at the percentage of cognates as such, since the relationship between cognates is not necessarily transparent to a non-linguistically trained reader. For this reason we also measured the linguistic distance between the cognates. These results are shown in Table 3. We have limited ourselves to distances between the cognates that are

Table 2. Written texts. Percentage of Dutch-Frisian and Dutch-Afrikaans cognates, cognates via a paradigm, and non-cognates, for function words and content words separately and total

	Frisian			**Afrikaans**		
	function	content	total	function	content	total
cognates	93.4	94.1	93.8	75.5	94.6	84.1
cognates via a paradigm	4.6	–	2.5	23.7	1.7	13.8
non-cognates	2.0	5.9	3.7	0.8	3.7	2.1

Table 3. Written texts. Mean Levenshtein distance (%) between Dutch-Frisian and Dutch-Afrikaans cognates

	Frisian			**Afrikaans**		
	function	content	total	function	content	total
cognates	37.6	30.1	34.2	22.5	19.4	20.9

related directly, such as Dutch *vrouw* 'woman' and Frisian *frau*, or via a synonym, such as Dutch *dag* 'day' (synonym of *etmaal* in the original Dutch text) and Afrikaans *dag*. For the cognates via a paradigm we did not calculate the Levenshtein distance, because the relationships between the forms involved are phonologically spurious.

In Table 3 we can see that for the cognates that are related directly or via a synonym the linguistic distance is much smaller for Afrikaans (mean of 20.9%) than for Frisian (34.2%). So, in addition to the fact that there are more non-cognates in Frisian, the cognates are less similar to Dutch than the Afrikaans cognates. We argue that these two facts together are largely responsible for the lower intelligibility of the Frisian reading texts.

3.2 Intelligibility of spoken text

Whereas texts that are presented visually appear to be easier to understand in Afrikaans than in Frisian, no significant difference between the two languages was found for the texts that were presented auditorily (see Table 1). How does this relate to the linguistic distances? In Table 4 the percentages of cognates and non-cognates are presented for the spoken texts. Comparing this table with Table 2, we see an even higher percentage of non-cognates in the Frisian spoken texts (8.6% corresponding with 41 tokens and 19 types) than in the written texts (3.7%). For Afrikaans the percentages for the two types of texts are about the same, namely 2.4% (12 types and 12 tokens) and 2.1%, respectively. On the

Table 4. Spoken texts. Percentage of Dutch-Frisian and Dutch-Afrikaans cognates, cognates via a paradigm and non-cognates, for function words and content separately and total

	Frisian			**Afrikaans**		
	function	content	total	function	content	total
cognates	89.0	89.4	89.1	79.5	95.2	86.3
cognates via a paradigm	4.0	–	2.3	19.8	–	11.3
non-cognates	7.0	10.7	8.6	0.3	4.8	2.4

Table 5. Spoken texts. Mean Levenshtein distance (%) in Dutch-Frisian and Dutch-Afrikaans cognates

	Frisian			**Afrikaans**		
	function	content	total	function	content	total
cognates	21.8	27.0	23.9	17.8	19.4	18.6

basis of these results one would expect the listeners to have more problems with Frisian than with Afrikaans.

In Table 5 the Levenshtein distances between the cognates in the spoken texts are presented. The differences between Afrikaans and Frisian are rather similar to the differences found for the written texts (Table 3). Both for the function words and content words, separately and summed, the Levenshtein distance is larger for Frisian cognates than for Afrikaans. This means that the Frisian cognates are less transparent than the Afrikaans cognates, which should make the Frisian texts more difficult to understand.

As far as the spoken texts are concerned, we are thus not able to show a clear relationship between intelligibility and linguistic distances. The intelligibility results show Frisian and Afrikaans to be equally difficult, whereas the distances suggest that Frisian should present more problems. In retrospect we now suspect that the intelligibility results may not be valid. We will give three examples to illustrate our view.

First, the listeners were asked what animal the frog text was about. The Frisian word for 'frog' *kikkert* is very similar to its Dutch counterpart *kikker*, whereas the Afrikaans *paddas* is very different. Accordingly, the percentage of correct answers was 100.0 for Frisian and only 6.7 for Afrikaans.[4] Not understanding the central term *padda*, which occurred ten times in the text, may have had a negative effect on (some of) the other questions as well.

Second, the listeners were asked what the purpose was of the action described in the frog text. The correct answer was *het milieu bewaken* 'guarding

the environment'. The word *milieu* was literally present in the Frisian text, whereas in the Afrikaans text it had been translated by *omgewing*. This word exists in Dutch as well, but with the neutral meaning of 'surroundings', a concept which Dutch listeners will not easily associate with action. Consequently, the percentage correct for Frisian was 73.7 against 20.0 for Afrikaans.

Third, the listeners who listened to the kangaroo text were asked why the veterinarians were surprised. The correct answer, that the kangaroo was able 'to find food', is very similar in Dutch (*eten te vinden*) and in Frisian (*iten te finen*) while the Afrikaans expression is very different (*kos te kry*). As a result only 5.9% of the answers were correct for Afrikaans against 50.0% for Frisian.

We do not claim that the percentages correct for the three questions do not reflect the listeners' understanding. In these three cases the meaning of the relevant Frisian words is indeed more transparent than that of the Afrikaans words. However, we think that this is not representative of spoken Afrikaans and Frisian in general (see Table 5). We think that with only ten questions the results are too much influenced by chance. Moreover, we think that the test is not valid in the sense that it measures word intelligibility rather than text comprehension. All in all we think that the results of this part of the study should not be trusted. The reason why we opted for the test in the present study was that we hoped to be able to compare our results with the results of a large Scandinavian project on mutual intelligibility, where the same text was used.

3.3 Intelligibility of spoken words

The listeners had more problems understanding the 19 Frisian words than the Afrikaans words (mean percentages of 45.6 and 64.0, respectively). Is this reflected in larger linguistic distances? For Frisian, there were three non-cognates (Frisian *skonk* versus Dutch *been*, Frisian *hynder* versus Dutch *paard*, and Frisian *holle* versus Dutch *hoofd*). As expected, none of the listeners had these words right. For Afrikaans there were no non-cognates. This difference is partly responsible for the lower intelligibility of Frisian.

If we furthermore look at the Levenshtein distance — we limit ourselves here to the distances for the 16 words that were cognates to Dutch both in Frisian and Afrikaans — we find a slightly larger mean distance for Frisian (21.9%) than for Afrikaans (17.9%). It seems reasonable to assume that this difference may also have played a role in the lower intelligibility of Frisian. The correlation between the Levenshtein distance and the percentage of subjects that gave a correct translation for the 16 cognate words is .74 ($p = .01$) for the

16 cognates and .77 ($p = .01$) for all 19 words of both languages. So, generally speaking, the larger the phonetic distance of a Frisian or Afrikaans word from Dutch, the harder it is for speakers of Dutch to infer its meaning.

However, the phonetic distance is not the only determinant of intelligibility. If we look at the responses for Afrikaans in more detail, we see, for example, that Afrikaans *bed* deviated only from Dutch *bed* in that the initial consonant was tense, resembling Dutch /p/, and that the vowel was lengthened. This resulted in a relatively small Levenshtein distance of 25.0%. Nevertheless, hardly any subject gave the right response. Most listeners responded with *beet* 'bite'; some others responded with *pet* 'pet' or *pit* 'pit'. This example shows that the percentage correct depends on the existence of other words in the language of the listener that are phonetically similar (neighborhood density). As soon as a pronunciation difference — however small — leads to another meaningful word, this word will be responded with, like in the case of Afrikaans *bed*. Such cases will lower the correlation coefficient. Of course, misidentifications like this have a larger chance of occurring as there is less pressure from the context on the semantic content of the target words.

4. Conclusion

We conclude that speakers of Dutch have fewer problems understanding Afrikaans than Frisian. This conclusion is based on a cloze test, which is generally seen as a reliable and valid measure for reading comprehension, for two written newspaper articles of average difficulty and on the translation of 19 spoken words in isolation. A third test, which consisted of five open questions for each of two spoken texts, yielded no significant difference between Frisian and Afrikaans. However, we have argued that this test may not be valid. The greater accessibility of Afrikaans appears to be due both to a smaller proportion of non-cognates and a smaller phonetic distance of the cognates.

Acknowledgements

We wish to thank Vincent van Heuven for his comments on earlier versions of this paper, Bart Alewijnse for calculating the Levenshtein distances, and Sulette Bruwer for running the experiment. We thank the teachers and pupils from the following schools for participating in the experiment: De Nieuwe Veste in Coevoerden, Celeanum Gymnasium in Zwolle, Maerlant Lyceum in The Hague, and Dingstede in Meppel.

Notes

1. We determined the difficulty of the texts with the so-called LIX-index (Björnsson 1968). This is the mean number of words per sentence plus the percentage of words which are longer than seven letters. Texts with a LIX-value of 35 to 44 have a mean level of difficulty. The mean LIX-value for the two Dutch texts was 42.

2. See http://www.nordkontakt.nu.

3. In some cases knowledge of a third language such as English or German might help to understand a non-cognate. We leave such cases out of consideration, since we do not know enough about the language knowledge of the subjects and their ability to use this knowledge to improve their understanding of Frisian or Afrikaans.

4. Dutch has the word *pad* 'toad', which is a direct cognate of Afrikaans *padda*. In the text the word *padda* was always presented in the plural form *paddas*. Apparently this word has too little similarity with Dutch *pad* or *padden* (plural), for only one of the listeners has given *pad* as a response.

References

Bezooijen, R. van and Gooskens, C. (2005) 'Intertalig tekstbegrip. De begrijpelijkheid van Friese en Afrikaanse teksten voor Nederlandse lezers [Interlingual reading comprehension. Intelligibility of Frisian and Afrikaans texts for Dutch readers].' *Nederlandse Taalkunde* 10, 129–152.

Björnsson, C.H. (1968). *Läsbarhet* [Readability]. Liber: Stockholm.

Börestam Uhlmann, U. (1994) *Skandinaver samtaler. Språkliga och interaktionella strategier i samtal mellem danskar, norrmän och svenskar* [Scandinavians converse. Linguistic and interactive strategies during conversations between Danes, Norwegians and Swedes]. Uppsala University, Uppsala.

Budovičová, V. (1987) 'Literary languages in contact' (A sociolinguistic approach to the relation between Slovak & Czech today). In J. Chloupek, J. Nekvapil et al., eds., *Reader in Czech sociolinguistics*. John Benjamins, Amsterdam / Philadelphia,156–175.

Heeringa, W. (2004) *Measuring dialect pronunciation differences using Levenshtein distances.* Doctoral dissertation, University of Groningen.

Jensen, J.B. (1989) 'On the mutual intelligibility of Spanish and Portuguese.' *Hispania* 72, 849–852.

Maurud, Ø. (1976) *Nabospråksforståelse i Skandinavia. En undersøkelse om gjensidig forståelse av tale- og skriftspråk i Danmark, Norge og Sverige* [Understanding the languages of the neighbouring countries in Scandinavia. An investigation of mutual intelligibility of written and spoken language in Denmark, Norway and Sweden]. Nordiska rådet, Stockholm.

Pierce, J. E. (1952) 'Dialect distance testing in Algonquian.' *International Journal of American Linguistics* 18, 203–210.