

Mutual comprehensibility of written Afrikaans and Dutch: symmetrical or asymmetrical?

Charlotte Gooskens: Dept. of Scandinavian Studies, University of Groningen, P.O.Box 716,
9700 AS Groningen, the Netherlands, c.s.gooskens@rug.nl.

Renée van Bezooijen: Dept. of Linguistics, Radboud University Nijmegen, P.O.Box 9103,
6500 HD Nijmegen, the Netherlands, r.v.bezooijen@let.ru.nl

Correspondence should be sent to Charlotte Gooskens

Mutual comprehensibility of written Afrikaans and Dutch: symmetrical or asymmetrical?**Charlotte Gooskens and Renée van Bezooijen****Abstract**

The two West-Germanic languages Dutch and Afrikaans are so closely related that they can be expected to be mutually intelligible to a large extent. The present investigation focuses on written language. Comprehension was established by means of cloze tests on the basis of two newspaper articles. Results suggest that it is easier for Dutch subjects to understand written Afrikaans than it is for South African subjects to understand written Dutch. In order to explain the results, attitudes as well as several types of linguistic distances were assessed. The relations between attitude scales and intelligibility scores were few and weak. Asymmetries in the linguistic relationships between the two languages are probably more important, especially the asymmetries in the number of non-cognates and the opacity of the relatedness of cognates. These asymmetries are caused by historical developments in Dutch and Afrikaans, with respect to the lexicon, the grammar, and the spelling.

Mutual comprehensibility of written Afrikaans and Dutch: symmetrical or asymmetrical?

Charlotte Gooskens and Renée van Bezooijen

1 Introduction

1.1 Background

When speakers of related languages communicate, there are three options: one speaker switches to the language of the other, both speakers adopt a third language, or both speakers stick to their own language. The third type of interaction, which is referred to with terms such as ‘semi-communication’ (Haugen 1966) or ‘receptive multilingualism’ (Braunmüller and Zeevaert 2001), has many advantages, in any case on the production side. People usually find it easier to express themselves in their mother tongue than in a later acquired second or foreign language. What, however, are the effects on the reception side? How much effort is it for listeners to understand what is being said or for readers to understand what is being written, specifically in those cases where they are being confronted with the related language for the first time? 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). Studies of many other languages were to follow, such as Spanish and Portuguese (Jensen 1989), Slovak and Czech (Budovičová 1987), and the Scandinavian languages (e.g. Maurud 1976, Bø 1978, Börestam 1987, Zeevaert 2004).

Thus far, most attention has been paid to the comprehensibility of spoken language, but there have also been some studies on written language. Van Bezooijen and Gooskens (2005) examined the comprehensibility of written Afrikaans for 20 Dutch-speaking language students, using a cloze test (see Section 2.1.2) for texts of an average level of difficulty. The mean percentage of items that was placed back correctly in the text was high, namely 81.8%. It is unclear whether this was the maximum achievable, as the subjects’ performance was not tested for their own language. However, the same texts were presented in Frisian, and this yielded a much lower percentage correct of 50.3. The difference with Afrikaans was significant at the level $p=0.01$. No correlation was found between comprehensibility scores and ratings on attitudinal scales related to the language, the speakers of the language and the countries where the languages are spoken. On the other hand, a number of linguistic distance measures did show a relationship with comprehension. The Frisian texts had more words that were historically unrelated to their Dutch equivalents than Afrikaans. Also, the Frisian words that were related to Dutch were less transparently recognizable and their orthographic forms more deviant

Just like Van Bezooijen and Gooskens (2005), the present study focuses on the comprehensibility of written Afrikaans for Dutch subjects. However, this time the symmetry of the communicative performance is examined. In other words, we are interested in knowing whether South African subjects understand as much of Dutch as Dutch subjects understand of Afrikaans. Moreover, the study has been extended in that this time we include a reference condition to be better able to interpret the results. So, the South African and Dutch subjects did a cloze test both for their own language and the other language. Finally, instead of language students, who might have a special talent for understanding unfamiliar languages, we have opted for secondary school pupils. With our study we seek answers to the following questions:

- (1) How well do Dutch subjects with no knowledge of Afrikaans understand written Afrikaans?
- (2) How well do South African subjects with no knowledge of Dutch understand written Dutch?
- (3) Can a possible asymmetry in comprehensibility be explained by asymmetrical attitudes?
- (4) Can a possible asymmetry in comprehensibility be explained by asymmetrical linguistic relationships?

Before describing what method we used to answer these questions and what the results were, we will give some background information on the two languages involved, i.e. Dutch and Afrikaans.

1.2 Dutch and Afrikaans

Together with High-German and Low-German, Afrikaans and Dutch constitute the continental branch of the West-Germanic languages. Dutch is older than Afrikaans. The oldest text which is generally seen as being written in Dutch dates from the first half of the twelfth century (Van der Wal 1992). The origin of Afrikaans can be traced back to 1652, when a small group of not more than 200 Dutch Calvinists settled on The Cape to secure the fresh water supply for Dutch ships on their way to the East-Indies. The central persons came from the western part of the Netherlands, and some characteristics of their speech are still present in modern Afrikaans. However, the Dutch as spoken by the early colonists changed rather quickly due to the fact that it mostly came to be used by speakers with an insufficient command of Dutch (see Ponelis 1997). At the end of the seventeenth century, the linguistic diversity at The Cape was enormous. There were speakers of, among others, High-German, Low-German, French, Malay, Portuguese, and Khoikoi. Although the proportion of native Dutchmen was low and their social structure loose, Dutch nevertheless won the competition with the other languages due to the political and economic dominance of the Dutch East India Company. Creolisation processes led to severe reduction of Dutch morphology, especially in the verb system. There was a loss of congruence, the imperfect tense, the infinitive, and the distinction between strong and weak verbs. The nominal system was reduced by the loss of gender distinctions and the pronominal system by the loss of unstressed forms. The incorporation of elements and structures from other languages, such as Malay, Khoikoi and English (see Ponelis 1997), led to a further divergence between Afrikaans and Dutch. In 1921 Afrikaans was assigned the official status of an independent language.

Present-day South Africa hosts a large number of languages. In addition to Afrikaans there are ten other languages with an official status.¹ According to the last census of 1996, Afrikaans is now spoken by almost six million people, both white and black. This amounts to 14% of the South African population. Most speakers of Afrikaans live in the Western Cape (39%), Gauteng (21%), the Northern Cape (10%) and the Eastern Cape (10%). Afrikaans is the native language of people of all social classes. Most live in urban areas, but a minority of 20% live in the countryside. Before 1994 Afrikaans was used in the official domains more often than English, but nowadays it is the other way round. Due to the increasing prestige of English and the recognition of various Bantu languages, the number of speakers of Afrikaans is decreasing. Afrikaans is a standardized language which is taught at all educational levels. There are five universities where Afrikaans is used, at least partly, as the medium of instruction.

Modern Standard Dutch is the mother tongue of 16 million people living in the Netherlands. The second official language of the Netherlands, Frisian, is spoken by 350.000 people in the province

of Friesland in the north. Dutch is also the mother tongue of 6 million people from the northern part of Belgium (Flanders). The remaining 4 million Belgians (the Walloons) speak French. The large majority of inhabitants have a fair command of the standard language, though dialect differences are still considerable, especially in peripheral areas in the south, the north, and the east of the Netherlands and in the whole of Flanders. Outside of Europe, Dutch is the official language (but not necessarily the mother tongue) of the inhabitants of Surinam and the Dutch Antilles. Indonesia has been independent since 1948, but there are still quite a few older people who speak the language well.

2 Comprehensibility

2.1 Method

2.1.1 Subjects

The Dutch subjects participating in the study were 32 pupils (11 boys and 21 girls). About half originated from Zwolle and surroundings in the northeastern part of the country; the other half came from The Hague in the west. In the Netherlands, children have to choose at an early age between various types of secondary education. All subjects participating in the present study were in their pre-final year of Dutch VWO, i.e. pre-university education. Their mean age was 16.0 years. They reported no active and at most a very superficial, passive knowledge of Afrikaans. All subjects spoke Dutch as their mother tongue.

The South African subjects were 33 pupils (20 boys and 13 girls) attending an Afrikaans-speaking school. They originated from Hennenman and surroundings in the province of Vrystaat. The South African educational system differs from the Dutch system in that all children attend the same type of secondary school. However, courses are offered at two different levels for each school subject. All subjects participating in the present study took at least four school subjects at the higher level, which allows access to the university. Their mean age was 16.9 years. They reported no active and no, or not more than a very superficial, passive knowledge of Dutch. All subjects spoke Afrikaans as their mother tongue.

2.1.2 Task

Text comprehension was assessed on the basis of two Dutch newspaper articles with an average level of difficulty.² 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.³ 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 Afrikaans and the same words were removed and placed above the texts.⁴ 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.

2.1.3 Design

The experiment consisted of three blocks. In the first block the subjects filled in a questionnaire about their personal background. In the second block they rated a number of attitude scales (see Section 3) and completed the cloze test for the language they were not familiar with (L2), i.e. Dutch for the South African subjects and Afrikaans for the Dutch subjects. The third block consisted of the

same tasks for the subjects' own language (L1). Half of the subjects from each country was given the feminist text in Afrikaans and the dating text in Dutch, and the other half of the subjects the other way round.⁵

2.2 Results

To determine how well Dutch subjects understand written Afrikaans and how well South African subjects understand written Dutch (research questions 1 and 2) we counted how many items were placed back correctly in the text. To obtain a greater stability we added up the data for the two texts. The results are given in Table 1. It can be seen that the Dutch subjects performed considerably better for the Afrikaans versions of the texts (66.5% correct) than the South African subjects pupils for the Dutch versions of the text (14.3% correct). To make sure that the task as such was feasible, the subjects had also carried out the cloze test for their own language. The Dutch subjects performed as we had hoped, they hardly made any mistakes (97.7% correct). However, the score for the South African subjects was unexpectedly low (64.1% correct).

A low level of proficiency in Afrikaans cannot have caused the mediocre performance of the South African subjects for their own language as such. They were native speakers of Afrikaans, they lived in a predominantly Afrikaans-speaking community, they went to an Afrikaans-speaking school, and the school itself ranks among the 25 best schools in the province. Moreover, the fact that seven pupils had the maximum score of 20 points (four for the dating text and three for the feminist text) proves that some subjects were indeed able to carry out the test without making any mistake. The difference between the South African and the Dutch subjects cannot be attributed to a difference in age or the level of education, since they had been carefully matched in this respect. According to their teacher the children 'is goed opgevoed en hul ontwikkeling is op standaard' (are well educated and their development is normal). Was the subject matter perhaps less interesting and less familiar to the South African pupils than to the Dutch pupils? Or could it be that the South African pupils were less used to making (cloze) tests than the Dutch pupils?

Whatever the reason may be, the difference between the two subject groups in their performance for L1 is unfortunate, because it hampers the comparison of the test results for L2. Nevertheless, there are two indications that the Dutch texts were more difficult for the South African subjects than the Afrikaans texts were for the Dutch subjects. In the first place, the within-group difference between the scores for L2 and L1 is larger for the South African subjects (64.1% - 14.3% = 49.8%) than for the Dutch subjects (97.7% - 66.5% = 31.2%). The difference, tested by means of a *t*-test for independent samples, is significant at the 1% level. If the subjects' interest in the texts or their experience with (cloze) tests had been the only factors playing a role, one would have expected a similarly large effect on the scores for both L1 and L2. In the second place, there were four Dutch subjects who did very well (at least 18 out of the 20 items correct) for both L1 and L2, whereas there were no such subjects among the South African pupils. None of the seven South African subjects who had all 20 items correct in L1 had more than five items correct in L2. Again, this result cannot be explained by general factors such as lack of interest or lack of experience.⁶ We therefore feel justified in concluding that the comprehension of written Afrikaans and Dutch is asymmetrical: Dutch is more difficult for South African readers than written Afrikaans is for Dutch readers.

3.2 Results

The mean attitude ratings given by the two groups of subjects are visualized in Fig. 1. In this figure as well as in Table 2 the direction of the scales has been reversed to facilitate the interpretation. In contrast to the original form, higher values now denote more positive ratings. It can be seen in Fig. 1 that in all cases the Dutch subjects are more positive towards Afrikaans than the South African subjects towards Dutch. All differences, tested by means of *t*-tests for independent samples, are significant at the 1% level. So, generally speaking one would tend to answer the third research question *Can the asymmetry in comprehensibility be explained by asymmetrical attitudes?* in the affirmative.

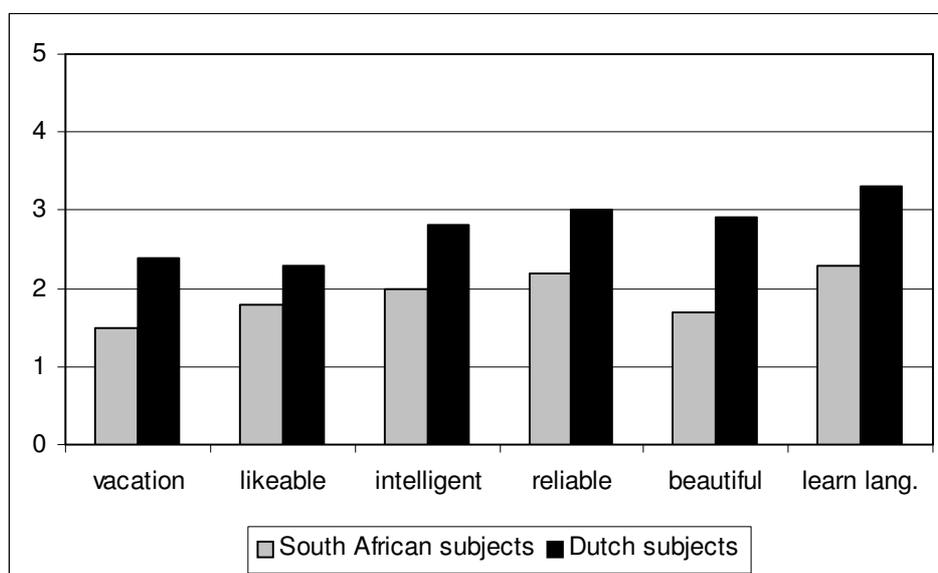


Fig. 1 Mean attitude ratings for Afrikaans as given by the Dutch subjects and for Dutch as given by the South African subjects. Higher values indicate more positive attitudes.

However, more conclusive is the direct relationship between the attitudinal ratings and the comprehension ratings for the individual subjects. One would like to know whether the same subjects who expressed a relatively positive attitude towards L2 were also the more successful ones in understanding L2. The correlation results are given in Table 2, both for the six individual scales and for all scales together.⁷ Again, we find some evidence that there is indeed a relation between attitudes and reading performance in the expected direction. For both groups of subjects there is a positive and significant correlation between overall attitude scores and test results. However, the percentages explained variance are low, not more than 19.4% for the Dutch subjects and 20.2% for the South African subjects. Moreover, there are no significant correlations for the separate scales rated by the South African subjects and only two significant correlations for the separate scales rated by the Dutch subjects. The Dutch pupils who find South African people likeable and intelligent tend to do better on the cloze test than their classmates who have a less positive opinion with respect to these two personality traits. Again, the percentages explained variance are low, namely 26.0% and 16.0%, respectively.

The correlations in Table 2 are based upon the absolute attitude ratings for L2 as presented in Fig. 1. We also computed relative attitude ratings by subtracting the ratings for L1 from the ratings

for L2. This shows for example how much more beautiful the Dutch subjects find Afrikaans than their own language. These relative attitude ratings were also correlated with the comprehension ratings. Now there were even fewer significant coefficients. We conclude that attitudes may have played at the most a weak role in the subjects' understanding of L2. In the next section, we will investigate the relationship between intelligibility and linguistic relationships.

Table 2 Correlation between attitude scores and comprehension scores for the two groups of subjects

Scale	Subjects			
	Dutch		South African	
	<i>r</i>	sig.	<i>r</i>	sig.
Would you like to spend your vacation in the Netherlands/South Africa?	.22	.24	.22	.22
Dutch/South-African people are:				
Likeable – not likeable	.51	.00	.07	.73
intelligent – not intelligent	.40	.03	.06	.76
reliable – not reliable	.30	.10	.01	.97
What do you think of the Dutch/Afrikaans language?	.16	.40	.11	.54
Would you like to learn Dutch/Afrikaans?	.27	.13	.17	.35
Total	.44	.01	.45	.05

4 Linguistic relationships

In addition to attitudes, comprehensibility is likely to be influenced by the linguistic similarity between the languages involved. To answer our fourth research question *Can the asymmetry in comprehensibility be explained by asymmetrical linguistic relationships?* we computed seven linguistic distance measures.

4.1 Method

As a first step in the linguistic analysis, the Afrikaans texts were aligned with the Dutch texts and the Dutch texts with the Afrikaans texts. To predict intelligibility, the linguistic relationships between the texts had to be calculated in both directions. An Afrikaans word in the text may have an equivalent synonym in Dutch, so that its meaning can be easily deduced by the Dutch reader. However, the opposite is not necessarily the case. The Dutch word may not have an equivalent synonym in Afrikaans (see measure A for further explanation). So, from a communicative point of view the linguistic relationships between the texts may be asymmetrical.

As an example of an aligned word string, we take the title of the dating text in the two languages: Du. *niet lang en blond maar een kleine kale sukkel* vs. Afr. *nie lank en blond nie maar 'n klein kaalkop lamsak* 'Not long and blond but a small bald duffer'.

Dutch	Afrikaans
niet	nie nie
lang	lank
en	en
blond	blond
maar	maar
een	'n
kleine	klein
kale	kaalkop
sukkel	lamsak

The example shows that we sometimes adjusted the word order to obtain a better matching of corresponding words: the two separate elements of Afrikaans double negative *nie nie* (*nie lank en blond nie*) were joined to match the single Dutch negative *niet*. This was done to avoid having word pairs with an empty element. Another adjustment pertained to the imperfect tense in Dutch (e.g. *vormden* 'formed'), which was replaced by the perfect tense (*hebben gevormd* 'have formed') to match the Afrikaans verb form (*het gevorm*). Due to these adjustments the distances between the texts were reduced to some degree. The alternative, even less satisfying, solution would have been to have incomplete word pairs and leaving part of the texts out of consideration.

The first three measures A, B and C bear on the historical relationship between the words. Measure D consists of the Levenshtein distance, a computer-based algorithm that we used to determine the orthographic similarity between the word forms. Measures E, F, and G were developed to express the communicative effects of divergent word forms. The measures will be explained in further detail below.

Measure A. Percentage of related words

A large proportion of related words, i.e. words in two languages with a common root, can be expected to facilitate mutual comprehension. Therefore, the percentage of related words in the texts is an obvious candidate to predict comprehensibility. The relationship between words does not necessarily have to be *direct*. In some cases, the meaning of a word can be deduced indirectly, via a cognate *synonym*. For example, the word *vaak* ('often') in the Dutch text corresponded with *dikwels* in the Afrikaans text. These two words are non-cognates. Nevertheless, the Dutch reader can easily understand Afrikaans *dikwels* because the word *dikwyls* exists in Dutch as a synonym of Dutch *vaak*. Note that the opposite does not hold true, the South African reader cannot understand Dutch *vaak*, as this word does not exist in Afrikaans as a synonym of *dikwels*. This shows that mutual comprehensibility may be asymmetrical and that the linguistic relationships between the two languages may depend on whether they are considered from the viewpoint of the Dutch subjects or from the viewpoint of the South African subjects. The percentage of related words, either directly or via a synonym, in the text constitutes the first linguistic distance measure.⁸ Because content words (nouns, adjectives, numerals, main verbs) are more important for intelligibility than function words (articles, conjunctions, prepositions, pronouns, auxiliaries, modals, particles, adverbs), the measure was calculated separately for these two word categories.

Measure B. Percentage of paradigm-related words

If it is not possible to deduce the meaning of a word because the words are related directly, it may be possible to deduce the meaning paradigmatically, i.e. via the grammar. For example, the Afrikaans translation of the Dutch word *zijn* ('are', third person *plural* present tense of the verb 'to be') is *is*. Strictly speaking, *zijn* and *is* are not related, neither directly nor via a synonym. A Dutch reader may nevertheless understand the meaning of Afrikaans *is*, because the form exists in Dutch as well with a slightly different grammatical meaning. Dutch *is* is the third person *singular* present tense of the verb 'to be'. So, in Dutch *is* and *zijn* belong to the same paradigm. Paradigm-related cognates can also operate via a synonym. The percentage of words in the text that are 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.

Measure C. Percentage of non-related words

If a word in an unfamiliar language bears no relationship with any word in the mother tongue, its meaning cannot be retrieved. Examples in the texts are Afr. *boude* en Du. *billen* 'buttocks'. In our study, we left a possible relationship with a third language out of consideration. In the first place we did not measure subjects' command of other languages, and in the second place we do not know to what extent subjects made use of their knowledge of languages in the task. The percentage of 'true' non-cognates (i.e. those non-related words that have no related synonym in L1) should be an important indicator of intelligibility. It is the complement of the first two measures. Measure A comprehends the paired words in L1 and L2 that are related directly, measure B comprehends the paired words in L1 and L2 of which the forms are not related directly but of which the meaning can nevertheless be retrieved by the reader because the form in L2 is related to a form in L1 in the same paradigm, and measure C comprehends the paired words in L1 and L2 which are not related directly nor via a paradigm.

Measure D. Levenshtein distance⁹

In many cases a non-linguistically trained person will be able to recognize a related word. However, some cognates will be easier to recognize than others. Due to speech changes and/or spelling conventions, the word forms may have diverged to such an extent that the original relatedness is no longer discernible. We assessed the degree of similarity between word forms by means of the so-called Levenshtein distance. This is an objective measure that can be calculated automatically by computer. The Levenshtein distances are based on the aligned texts as described above. However, whenever a non-related word was interpretable indirectly via a cognate synonym (see measure A), the distance with the synonym was computed rather than with the original word.

In the present study, the Levenshtein distance between corresponding words is based upon the minimum number of letters 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. The fewer operations are needed, the greater the similarity. All operations were given an equal weight.¹⁰ 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 of the procedure and an overview of different variants of the measure. As an example we present the calculation of the distance between the written Dutch word *dochtertje*, the diminutive of 'daughter', and the corresponding Afrikaans word *dogtertjie*.

alignments	1	2	3	4	5	6	7	8	9	10	11
Afrikaans	d	o	g	-	t	e	r	t	j	i	e
Dutch	d	o	c	h	t	e	r	t	j	-	e
costs	0	0	1	1	0	0	0	0	0	1	0

It can be seen that the transformation involved one substitution of a consonant by another consonant (*g* by *c*), one insertion (*h*), and one deletion (*i*). The sum of costs ($1 + 1 + 1 = 3$) is divided by the number of alignments (11). The result is a distance of 27.3%. The transformation from Afr. *vir* to Du. *voor* ‘for’ involves one substitution and one insertion and only four alignments, resulting in a rather large distance (50.0%). An example of a small distance can be found for the word for ‘visitors’ Afr. *besoekers* vs. Du. *bezoekers*. This transformation involves only a substitution of *z* by *s* while the number of alignments is large (9). This results in a distance of 11.1%. The total Levenshtein distance between texts is the mean distance over all word pairs.

Measure E. Lexical opacity

The Levenshtein distance as described in the preceding paragraphs only assesses whether letters are identical or different. It does not take into account the communicative consequences of the differences. For example, differences at the beginning of a word have been found to be more detrimental to word recognition than differences at the end of a word (Broerse and Zwaan 1966). At present it is not yet possible to compute the effect of this and other factors, such as the relative contribution of vowel and consonants, automatically, even less so in the context of related languages. We therefore estimated the recognizability of the relationship between pairs of words ourselves, basing our judgments on our own intuitions as language users. We distinguished the following four degrees of lexical opacity.

- Lexical relatedness completely transparent (0 point)

When two cognates have the same form in present-day Afrikaans and Dutch, there is obviously no recognition problem. An example is Du. and Afr. *uitbuiting* ‘exploitation’. Identical word forms were given a rating of 0 points.

- Lexical relatedness reasonably transparent (1 point)

One point was assigned when differences between cognates were so few and so obvious that we assumed the reader to have few problems seeing the relationship. Example: Du. *zeventig* and Afr. *sewentig* ‘seventy’.

- Lexical relatedness rather opaque (2 point)

Two points were assigned to cognates that differed so much that we assumed that it would be hard for non-linguistically trained readers to see the relationship. In almost all cases differences in several letters are involved. Example: Afr. *seuntjie* and Du. *zoontje* ‘son’.

- Lexical relatedness completely opaque (3 point)

In some cases two cognates have so little in common synchronically that it must be (virtually) impossible for a non-linguistically trained reader to see that they are related. Example: Afr. *hê* and Du. *hebben* ‘have’.

The lexical opacity of the Afrikaans words for Dutch readers was rated by the second author, who is Dutch. Her ratings were compared with the independent ratings of a second Dutch linguist.¹¹ The percentage of identical scores was 90.1% and the correlation coefficient .92. On the basis of these outcomes we decided that the ratings given by the author were reliable enough to be kept unchanged. To rate the lexical opacity of the Dutch words for South African readers the second author worked together with a native speaker of Afrikaans.¹² Measure E, and also measures F and G, was computed only for related words (directly, via a synonym or via a paradigm) and not for non-related words. The latter are not interesting, as they are by definition non-transparent. At the text level, measure E was computed by adding the scores for all word pairs and dividing it by the total number of word pairs. The ratings can vary between 0 and 3, but they have been rescaled to obtain percentages.

Measure F. Grammatical opacity

In addition to a correct interpretation of the lexical meaning, a correct interpretation of the grammatical meaning of a word is also necessary for a good understanding of a text. We distinguish two aspects of grammatical meaning: (a) Is it clear what grammatical word class is involved? (b) Is it clear what tense, number, gender, person, etc. is involved? An example of type (a) is Afr. *die* as in *die feministe* ‘the feminist’. In this context speakers of Dutch will be inclined to interpret *die* as a demonstrative pronoun, whereas in fact it is a definite article in Afrikaans. An example of type (b) is Afr. *is*, which will be interpreted by speakers of Dutch as the third person singular present tense of the verb *zijn* ‘to be’ (see measure B). In Afrikaans, however, the form may refer both to the singular and to the plural. Grammatical meaning is either transparent (0 point) or opaque (1 point). The two types (a) and (b) were rated separately. The ratings for the grammatical opacity of a word may therefore vary between 0 and 2. All points were added up and divided by the total number of word pairs. The resulting values were rescaled to percentages.

Measure G. Total opacity

To determine the total opacity of a word we have added up measures E and F. As measure E goes from 0 to 3 and measure F from 0 to 2, lexical opacity is weighted more heavily than grammatical opacity. We think this adequately represents the relative contribution of the two measures to text comprehension. The values from 0 to 5 have been rescaled to percentages.

4.2 Results

The results of the first three linguistic measures are presented in Table 3. In all cases tokens are presented rather than types. It can first of all be observed (measure C) that the South African subjects have to deal with a considerably higher percentage of non-related words (17.3%) when reading the texts in L2 than the Dutch subjects (2.1%). True non-cognates, that have no related synonym in L1, are of the utmost importance, since they are by definition unintelligible and a single unintelligible word can make a whole sentence or an even larger part of the text incomprehensible. In the original texts, the percentage of non-cognates was of course symmetrical. The asymmetry in the percentages

of non-related words in Table 3 is caused by two factors, which in turn explain several other asymmetries in the table.

The lower percentage of non-related content words from the perspective of the Dutch reader (3.7% of all content words) relative to that of the South African reader (16.9% of all content words) is due to the fact that there were more non-related content words in the Afrikaans text that could be replaced by a related synonym in Dutch (81.3% of all non-related content words in the original text) than there were non-related content words in the Dutch texts that could be replaced by a related synonym in the Afrikaans (27.9%). This is to a large extent due to the moment at which Afrikaans originated. In Dutch there are many doublets of which one element is a relatively recent borrowing (from French, English or German) and the other indigenous. Afrikaans often only has the older form, for example (first attestation in Dutch in parentheses) Du. *café* (1897 < French) / *kroeg* (1588) versus Afr. *kroeg* 'pub'. That the age of words often played a role also appears from other examples in the texts, such as Du. *lol* (1897) / *plezier* (1574) versus Afr. *plesier*. In Afrikaans, there have been more restrictions on borrowing:

[B]ecause of the perceived threat to the separate identity of Afrikaners *vis-à-vis* their English-speaking compatriots, there is a longstanding fear of those international loanwords with cognate forms in English such that they are now regarded as dispensable anglicisms with preference being given to indigenous synonyms. (...) (Donaldson 1994: 503).

Due to these historical and language political developments, the Dutch reader will be confronted with fewer unknown content words in the Afrikaans texts than the South African reader in the Dutch texts. The asymmetry in the percentages of non-related content words automatically leads to the asymmetry in the percentages of related content words in measure A. Of course in this case, the percentage is higher from the perspective of the Dutch reader (94.6%) than from the perspective of the South African reader (83.1%).

On the other hand, the lower percentage of non-related function words from the perspective of the Dutch reader (0.8% of all function words) relative to that of the South African reader (17.7% of all function words) has to do with the fact that many more function words in the Afrikaans texts could be interpreted via paradigmatic relationships (23.7% of the function words) than in the Dutch texts (no occurrences). So, in this case there is a trade-off between measure C and measure B. A clear case are the pronouns. The Afrikaans pronominal system is very much reduced compared to the Dutch pronominal system. For example, in Dutch there are two forms for the first person plural of the personal pronoun, namely *wij* 'we' for the subject and *ons* 'us' for the object. Afrikaans only has *ons* for both cases. Afrikaans *ons* will be familiar to the Dutch reader because of its paradigmatic relationship with *wij* in his own language, whereas the South African reader will be confronted with Dutch *wij* which has no related form in Afrikaans. Many similar examples can be added.

Table 3 Measures A, B, and C. Percentages of related and non-related words (tokens) from the perspective of the South African reader and from the perspective of the Dutch reader, broken down for functions words, content words, and all words together.

		South African subjects reading Dutch			Dutch subjects reading Afrikaans		
		function	content	all	function	content	all
A	related directly or via a synonym	82.3	83.1	82.7	75.5	94.6	84.1
B	related via a paradigm	-	-	-	23.7	1.7	13.8
C	non-related	17.7	16.9	17.3	0.8	3.7	2.1

Until now we have been counting the percentages of word pairs falling within different categories of relatedness. We will now look at these word pairs more closely. To express the degree of similarity, or rather dissimilarity, between corresponding words in the Dutch and Afrikaans texts in an objective way, we used the Levenshtein method (measure D). In principle, the Levenshtein distance between two languages A and B is symmetrical, i.e. it does not make a difference whether language A is taken as the point of departure or language B. However, in our case some non-related words in the texts were replaced by a synonym cognate. We just saw that this occurred more frequently from the perspective of the Dutch reader than from the perspective of the South African reader. As is to be expected, this asymmetry is reflected in the Levenshtein distances between Dutch and Afrikaans when they are computed for the texts as a whole, including the synonym cognates. The Levenshtein distance between the two languages is smaller when considered from the perspective of the Dutch reader (24.9%) than when considered from the perspective of the South African reader (33.0%). However, when considered for the three categories A, B, and C separately, the Levenshtein distances are symmetrical, and therefore of no further interest to us here.

The results for the last three measures E (lexical opacity), F (grammatical opacity) and G (total opacity) are presented in Table 4. The original values have been rescaled to percentages in order to facilitate comparison. The reader is reminded that in these measures the non-related words (category C in Table 3) were excluded. We will focus on the differences between the values as a function of the perspective taken (the Dutch or the South African reader). A *t*-test for independent samples showed that these were all significant at the 5% level, except for the opacity of the total meaning (measure G) of the content words and of all words together.¹³ The pattern is clear. For the South African subjects reading the Dutch texts the lexical meaning of related words (measure E) is less transparent than for the Dutch subjects reading the Afrikaans texts. This holds for the function words (22.6% versus 13.3%) and content words (32.3% versus 21.0%) separately as well as for all words in the texts together (27.3% versus 16.7%). With respect to the grammatical meaning, we find asymmetries in the opposite direction. How can these findings be explained?

With respect to the asymmetrical opacity of the lexical meaning, orthography appears to be an important factor. The Afrikaans orthography presents few problems to the Dutch reader, whereas there are several aspects of the Dutch orthography that may confuse the South African reader. A clear example is the phoneme /ei/, which, when evolved from West Germanic long /i/, is written in Afrikaans as *y* and in Dutch as *ij*. Dutch readers automatically interpret the letter *y* as an alternative spelling for *ij*, whereas South African readers may interpret *ij* as a sequence of a vowel and a consonant. The asymmetry is reflected in the lexical opacity scores for word pairs such as Afr. *byna*

(0 point for the Dutch reader) and Du. *bijna* (1 point for the South African reader) ‘almost’. A similar asymmetry is found in the spelling of /x/. The spelling *g* in Afrikaans occurs in Dutch as well, and will therefore automatically be interpreted correctly. To the South African reader, however, the Dutch spelling *ch* must be obscure. This is reflected in asymmetrical lexical opacity ratings for a word pair such as Afr. *regte* (1 point for the Dutch reader) and *rechte* (2 points for the South African reader) ‘straight’. The problems presented by the Dutch spelling were checked with a few speakers of Afrikaans, but have, of course, to be tested experimentally in future research.

The asymmetrical opacity of the grammatical meaning (measure F) is due to the fact that many Dutch words have been preserved in Afrikaans, but with a more general grammatical meaning. Above we gave the example of the personal pronoun *ons*, which in Dutch can only be used for the object form and not for the subject form, whereas in Afrikaans it can denote both cases. The South African reader is not confronted with an unfamiliar grammatical meaning, whereas the Dutch reader is. The situation is thus more confusing for the Dutch reader than for the South African reader. The verb system presents many similar asymmetries.

Taking into account that the lexical meaning is more important for text comprehension than grammatical meaning, the results for the communication-oriented measures again make plausible why the South African subjects were less successful in completing the cloze text for L2 than the Dutch subjects.

Table 4 Measures E, F, and G. Mean opacity (in %) of Dutch words for South African readers and of Afrikaans words for Dutch readers, broken down for function words, content words, and all words together. Higher values express greater opacity.

		South African subjects reading Dutch			Dutch subjects reading Afrikaans		
		function	content	all	function	content	all
E	lexical opacity	22.6	32.3	27.3	13.3	21.0	16.7
F	grammatical opacity	0.5	0.0	0.5	17.5	10.5	14.5
G	total opacity	14.0	19.4	16.4	15.4	16.4	15.8

5 Conclusion

A cloze test for two newspaper articles of average difficulty yielded higher percentages of correct answers for Afrikaans interpreted by Dutch speaking-subjects than for Dutch interpreted by Afrikaans-speaking subjects. In the literature, asymmetrical relationships in intelligibility are often attributed to language attitudes. Listeners or readers who have a relatively positive attitude towards L2 would have fewer problems understanding L2. This would have to do with motivation rather than with competence (e.g. Maurud (1976), Bø (1978) and Börestam (1987) for asymmetries in intelligibility in the Scandinavian area and Wolff (1959) for Western Africa). In the present study, attitudes were measured by means of judgment scales related to the languages themselves, the speakers of the languages, and the countries where the languages are spoken. A few significant correlations were found between attitude ratings and intelligibility scores, in the expected direction. However, in terms of explained variance there is enough room for other factors to have played a role.

The first important linguistic factor that we hold responsible for the asymmetrical

intelligibility is the percentage of non-related words in L2 for which there is no related synonym in L1. Due to the large number of content words imported in Dutch after Afrikaans developed and the loss of many Dutch function words in Afrikaans, the South African reader encountered more of these than the Dutch reader. The second important factor is the opacity of the relationship of the related words, which is greater for the Afrikaans words than for the Dutch words. With respect to lexical opacity this is mainly due to differences in spelling; with respect to grammatical opacity this is mainly due to differences between the morphological systems.

The asymmetries in the linguistic relationships appeared very clearly from the measures which were done completely 'by hand', based on linguistic knowledge and intuitions, namely the percentages of (non-) related words and the lexical and grammatical opacity scores. We also used a variant of the computer-based Levenshtein algorithm, which measured orthographical similarity on the basis of the number of operations that were needed to transform the word in the one language into the corresponding word in the other. Prior to applying the Levenshtein distance measure, the words in the Afrikaans and Dutch texts had been aligned by hand, with some adaptations in the word order. Moreover, non-related words in L2 had been replaced by a synonym in L1 whenever this was possible. When correlating the Levenshtein distances with the scores for lexical opacity for the category of words that are related directly or via a synonym (measure A), we got significant coefficients of .79 for Dutch from the perspective of the South African reader and .52 for Afrikaans from the perspective of the Dutch reader. We can think of several reasons for the fact that the correlation is lower from the perspective of the Dutch readers than from the perspective of the South African readers. In the first place, there is less variance in the lexical opacity scores for Afrikaans than for Dutch. In Afrikaans there were only three words with a lexical opacity score of 3, whereas in Dutch there were 18. In the second place, in Afrikaans a large Levenshtein distance often corresponded with great lexical opacity, as is to be expected. However, there were also some cases where a large Levenshtein distance corresponded with little lexical opacity. For example, the Levenshtein distance between Afr. *sy* and Du. *zij* 'she' is 100%, since all letters are different. Nevertheless Afr. *sy* presents few problems for the Dutch reader. Word initial *z* is often pronounced as [s] in Dutch and *y* is a regular alternative in writing for *ij*. The combination of high values in the one variable with both high and low values in the other variable has, of course, a negative effect on the correlation coefficient. In the present study this and other factors have not been included in the Levenshtein distance measure. They depend to a large extent on the specific combination of languages and on the medium (spoken or written).

In future research we will concentrate on the intelligibility of single words in related languages and language varieties rather than texts. One of the things we would like to find out is whether deletions are more detrimental to the recognition of a related word in another language than insertions. In our material we found many examples of syncope, such as Afr. *eie* from Du. *eigen* 'own' and Afr. *teen* from Du. *tegen* 'against'. It is our intuition that the effect on intelligibility is asymmetrical in the sense that it will hinder the South African reader to a higher extent than the Dutch reader. However, this hypothesis has to be verified experimentally. We are also interested in non-linguists' intuitions about plausible sound correspondences and the role these play in written communication between speakers of related languages. To what degree do readers use silent speech in decoding unfamiliar word forms? Our ultimate goal is to develop a variant of the Levenshtein

distance measure that is better able to predict the mutual intelligibility of word pairs between related languages and language varieties.

References

- Björnsson, C. H.** (1968). *Läsbarhet* [Readability]. Stockholm: Liber.
- Bø, I.** (1978). *Ungdom og naboland* [Youth and neighboring country]. Stavanger: Rogalandsforskning (report 4).
- Börestam, U.** (1987). *Dansk-svensk språkgemenskap på undantag* [Danish-Swedish language community as a special case]. Uppsala: Uppsala Universitet.
- Braunmüller, K. and L. Zeevaert** (2001). *Semikommunikation, rezepitive Mehrsprachigkeit und verwandte Phänomene. Eine bibliographische Bestandsaufnahme* [Semicommunication, receptive multilingualism and related phenomena. A bibliographical overview]. Arbeiten zur Mehrsprachigkeit – Folge B, nr. 19. Universität Hamburg.
- Broerse, A.C. and E.J. Zwaan** (1966). The information value of initial letters in the identification of words. *Journal of Verbal Learning and Verbal Behavior*, 5, 441-446.
- Budovičová, V.** (1987). Literary languages in contact (A sociolinguistic approach to the relation between Slovak and Czech today). In: J. Chloupek, J. Nekvapil et al. (eds.), *Reader in Czech sociolinguistics*. Amsterdam/Philadelphia: John Benjamins, 156-175.
- Donaldson, B.** (1994). Afrikaans. In: E. König and J. van der Auwera (eds.), *The Germanic languages*. London and New York: Routledge.
- Haugen, E.** (1966). Semicommunication: The language gap in Scandinavia. *Sociological Inquiry* 36, 280-297.
- Heeringa, W.** (2004). *Measuring dialect pronunciation differences using Levenshtein distances*. Groningen: Groningen dissertations in linguistics (Grodil).
- 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* [Mutual intelligibility of languages in Scandinavia. A study of the mutual comprehension of written and spoken language in Denmark, Norway and Sweden]. Stockholm: Nordiska rådet.
- Pierce, J. E.** (1952). Dialect distance testing in Algonquian. *International Journal of American Linguistics* 18, 203-210.
- Ponelis, F.** (1997). Afrikaans. In: M. van den Toorn, W. Pijnenburg, J. van Leuvensteijn, and J. van der Horst (eds.), *Geschiedenis van de Nederlandse taal* [A history of the Dutch language]. Amsterdam: Amsterdam University Press, 597-645.
- Trudgill, P. and H. Giles** (1978). Sociolinguistic and linguistic value judgments: correctness, adequacy, and aesthetics. In: F. Coppieters and D.L. Goyvaerts (eds.), *Functional studies in language and literature*. Gent: Story-Scientia, 167-190.
- Van Bezooijen, R. and C. Gooskens** (2005). Intertalig tekstbegrip. De begrijpelijkheid van Friese en Afrikaanse teksten voor Nederlandse lezers [The comprehensibility of Frisian and Afrikaanse texts for Dutch readers]. *Nederlandse Taalkunde* 10 (2), pp. 129-152.
- Van Bezooijen, R. and C. Gooskens** (to appear). How easy is it for speakers of Dutch to understand spoken and written Frisian and Afrikaans, and why? *Linguistics in the Netherlands* 22, 13-24.

- Van der Wal, M.** (1992). *Geschiedenis van het Nederlands [A history of Dutch]*. Utrecht: Het Spectrum.
- Webb, V.** (2002). *Language in South Africa. The role of language in national transformation, reconstruction and development*. Amsterdam: John Benjamins.
- Wolff, H.** (1959). Intelligibility and inter-ethnic attitudes. *Anthropological linguistics* 1, 34-41.
- Zeevaert, L.** (2004). *Interskandinavische Kommunikation. Strategien zur Etablierung von Verständigung zwischen Skandinaviern im Diskurs [Interscandinavian communication. Strategies towards establishing the understanding among Scandinavians in discourse]*. Hamburg: Dr. Kovač.

Notes

¹ The quantitative information about Afrikaans is from Webb (2002).

² We aimed at texts with an average level of difficulty. That is why we selected human interest items from a regional newspaper (*De Leeuwarder Courant*), which addresses readers of all educational levels. To make sure that the texts indeed had an average level of difficulty, we applied 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. So, this confirmed that the texts indeed had the intended level of difficulty.

³ The cloze test was developed by William Taylor in 1953 in America. Since then it has been a widely-used tool for measuring the intelligibility of texts. Sometimes the words are placed above the text, like in our investigation, but sometimes the test subjects have to think of words to fill in themselves.

⁴ In principle, we had two options. The first option was to take original texts in both languages. However, in that case the texts might not have been comparable. To assess a possible asymmetry in intelligibility, it is imperative that there be no differences between the texts in the two languages in vocabulary and style. We therefore decided to make use of translation. We took great care to obtain high-quality translations. The Dutch texts were translated into Afrikaans by a South African student of English and corrected by several South African linguists.

⁵ In addition, the subjects were also tested for their listening comprehension (see Van Bezooijen and Gooskens, to appear). The listening tests preceded the cloze tests.

⁶ In fact, we found no relationship between the subjects' performance for the two languages. The Pearson correlation between the scores for the Dutch and Afrikaans texts was $-.09$ for the Dutch subjects and $.06$ for the South African subjects. Neither coefficient is significant at the 5% level.

⁷ We normalized for the difference in level of difficulty of the two texts in the following manner. We determined the mean intelligibility score for each of the Afrikaans texts and next calculated the difference between the two means. The relatively difficult feminist text was on average 2.0% easier than the overall mean and the relatively difficult dating text was on average 2.0% more difficult than the overall average. In order to normalize, all subject who read the Afrikaans feminist text had 2.0% added to their score and subjects who read the Afrikaans dating text had 2.0% subtracted from

their score. The same procedure was followed for the Dutch texts. In this case 0.15% was added or subtracted.

⁸ Sometimes the frequency of word pairs related via a synonym may differ somewhat in the two languages. However, we have taken the view that as long as the word can be assumed to be known by readers of the age group at hand, the frequency does not have to be identical.

⁹ We thank Bart Alewijnse for calculating the Levenshtein distances.

¹⁰ We measured the distances on the basis of binary differences. Different symbols increase the distance, identical symbols do not increase the difference. For spoken texts, a more advance method can be used, where the phonetic distances between the individual sounds are taken into account.

¹¹ We thank Vincent van Heuven for his help.

¹² We thank Sulette Bruwer for her help.

¹³ It was not possible to carry out a *t*-test for dependent samples since the Dutch text were aligned differently with the Afrikaans texts than the Afrikaans texts with the Dutch texts.