[0816] Dialect variation in and around Frisia: classification and relationships

Wilbert Heeringa

1. Introduction

1.1 Position of Frisian

Different varieties of Frisian are found in the Netherlands and Germany along the North Sea. In the Dutch province of Friesland a variety is spoken known as West Lauwers Frisian. In the municipality of Saterland in the North of Germany an East Frisian or Sater Frisian variety is spoken (Fort, 2001). On the west coast of Schleswig-Holstein roughly between the Dano-German border in the north and the small town of Bredstedt in the south, and on islands west of the northern part of the coast in this area, as well as on the island of Heligoland, a variety is spoken which is known as North Frisian (Walker, 2001).

Siebs (1901) distinguishes West Frisian, East Frisian and North Frisian. In the 13th century the West Frisian area was found between the Flie and Lauwers, i.e. the Dutch province of Friesland. The East Frisian area was then found between the Lauwers and the Weser. It could be further divided into Weser Frisian and Ems Frisian. Ems Frisian included the present German municipality of Saterland and the Dutch province of Groningen (see pp. 1166-1174). Siebs divides North Frisian into coastal varieties on the one hand and island varieties on the other. The different Frisian groups descended from Old Frisian. “Of the continental Old Germanic dialects, it is Old Frisian that is most closely related to Old English” (Visser, 1997, following Nielsen, 1981).

In this article we focus on West Lauwers Frisian (in brief, Frisian) and other varieties spoken in West Frisia (Fryslân/Friesland), a province in the northwest of the Netherlands, and in a small part of the province of Groningen, east of West Frisia. Heeringa (2004) shows that among the dialects in the Dutch language area the Frisian varieties are linguistically most distant from standard Dutch. This may justify the fact that Frisian is recognized as a second official language in the Netherlands. Today, in West Frisia, probably

about 74% of the population is able to speak Frisian. This means that there are about 400,000 speakers of West-Frisian (Gorter, 2001).

1.2 Classification of Frisian

Probably the oldest known source in which different Frisian varieties are distinguished, is *Hulde aan Gysbert Japiks II*, written by J.H. Halbertsma and published in 1827. This publication is discussed extensively by Breuker (1973). Halbertsma distinguishes rural Frisian (Lânfrysk), the dialect of Hindeloopen (Hynljippersk), Old Frisian, the language of Gysbert Japicx, the dialect of the island of Schiermonnikoog (Skiermuontseagersk), Town Frisian (Stedfrysk), the dialect of Workum (Warkumersk), South Corner (Súdhoeksk) and the dialect of Molkwerum (Molkwardersk) (see Breuker (1973), p. 45). Old Frisian may be interpreted as the protolanguage from which the other varieties originated. The language of Gysbert Japicx is a mix of rural Frisian and South Corner (see Breuker (1973), p. 25 and p. 46). Town Frisian is much closer to Dutch than any of the other varieties. The dialects of Hindeloopen, Molkwerum and Workum belong to the South Corner group (see Breuker (1973), p. 25). Halbertsma writes that the dialects of Molkwerum, Hindeloopen and the island dialect (Eilandsch) belong to the same language group (Breuker (1973), p. 45). We assume that the dialect of Schiermonnikoog belongs to the island dialect. On the basis of these data we constructed a tree which is shown in Figure 1.

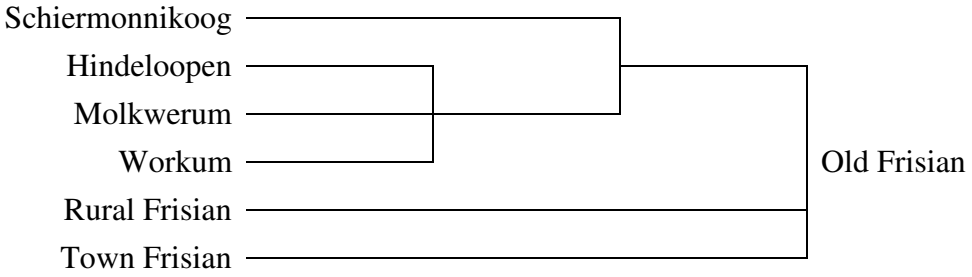

Figure 1. Classification of West Lauwers Frisian dialects, obtained on the basis of data given by J. H. Halbertsma in his Hulde aan Gysbert Japiks II*, published in 1827. The different varieties originated from Old Frisian.*

In 1867 Halbertsma distinguishes three main groups: the dialects of South Corner, Clay Corner and the Seven Woods. He also mentioned the dialects of Hindeloopen, Warns, Workum and Makkum, possibly as a group of

archaic varieties (see Breuker (1973), p. 47, who cites Halbertsma (1867), p. 12). The division in three main groups is also found in Hof's map (1933). Hof's map is published in his *Friesche Dialectgeographie* (Frisian dialect geography), which is still considered the standard work (De Haan, 2001). The map is based on four bundles of phonological isoglosses (van der Veen, 2001, p. 113).

Related but different is van der Veen's map (2001, p. 99), reflecting a division based on computational processing of isoglosses. These isoglosses are based on the most frequent words spoken in West Frisia (nearly 200 in total). He gathered the words from different sources (e.g. the *Reeks Nederlandse Dialectatlassen*). "By means of cluster analysis he was able to indicate the mutual similarity and distinction." (p. 113). The map is based on van der Veen (1994) (see also van der Veen, 1986).

1.3 This paper

Is the classification of Frisian dialects suggested by Hof in 1993 still valid when using more recent data? To find an answer to this question we will classify the varieties spoken in and around Frisia on the basis of material of the *Reeks Nederlandse Dialectatlassen* (Blancquaert, 1925-1982). Using this material we measure linguistic distances between the varieties using the Levenshtein distance. On the basis of these distances, we performed cluster analysis. Subsequently we will compare our results with both Hof's map and van der Veen's map. We will also consider the relation of the varieties with respect to Standard Frisian and Standard Dutch. We will especially focus on the question whether the dialect of the Frisian cities is a Dutch or a Frisian dialect.

In Section 2 we discuss our data source: the *Reeks Nederlandse Dialect-tlassen (RND)*. Section 3 is a brief explanation of our tool for the measurement of linguistic distances: the Levenshtein distance. On the basis of the Levenshtein distances, the varieties can be classified. The role of transcriber differences is discussed in Section 4. We will also show how the influence of transcriber differences can be eliminated. A detailed classification is described in Section 5. We compare our results with classification results in literature, especially the classifications of Hof and van der Veen. In Section 6 we consider the relation of the varieties with respect to Standard Dutch, Standard Frisian and to the Groningen dialect. Finally we draw some conclusions in Section 7.

2. Data source

The *Reeks Nederlandse Dialectatlassen* is a series of atlases covering the Dutch language area. The Dutch area comprises the Netherlands, the northern part of Belgium (Flanders), a smaller northwestern part of France and the German county Bentheim. The RND contains 1956 varieties, which can be found in 16 volumes. The first volume appeared in 1925, the last in 1982. E. Blancquaert initiated the project. When Blancquaert passed away before all the volumes were finished, the project was finished under the direction of W. Peé.

In the RND, the same 141 sentences are translated and transcribed in phonetic script for each dialect. Since digitizing the phonetic texts is time-consuming on the one hand, and since the Levenshtein distance is a word-based method on the other hand, we selected only 125 words from the text. The words represent (nearly) all the vowels (monophthongs and diphthongs) and consonants. The words are listed in Table 1.

Part 15 of the RND contains transcriptions of Frisian varieties and Frisian mixed varieties. For the most part they are found in the province of Friesland, but a few are also found in the adjacent Westerkwartier area in the province of Groningen.

Figure 2 shows our selection of 54 varieties of this atlas part. Dialects in the northern part were recorded and transcribed by G. van der Woude, and those in the southern part by K. Boelens. The transcriptions of Dokkum, Franeker and Sneek were joint recordings of E. Blancquaert, K. Boelens and G. van der Woude. The recordings were made in de period 1950-1951.

For the Frisian locations of Tjalleberd, Donkerbroek and Appelscha two texts are given in the RND. In Tjalleberd, most people spoke Frisian when the RND recordings in Friesland were made. However, a small part of the population spoke Tjalleberds (or 'Gietersk'), a variety introduced by peat laborers from northwestern Overijssel (Giethoorn and surroundings). We process the Tjalleberd variety as language island. The Frisian variety is referred to as 'Tjalleberd 1', and the Tjalleberd variety as 'Tjalleberd 2'.

Both a Frisian and a Stellingwerf variety are spoken in Donkerbroek. In Daan & Blok's map (1969) it can be seen that the river Kuinder (or Tjonger) is the boundary between Frisian (west) and Low Saxon (east). Since Donkerbroek is located west of this river, we regarded the Frisian variety as part of the Frisian language continuum and the Stellingwerf variety as a Low Saxon language island in the Frisian language continuum. Below 'Donker-

Figure 2. Distribution of the 54 RND varieties in the province of Friesland and a small part of the province of Groningen. The northern varieties are transcribed by G. van der Woude, and the southern ones by K. Boelens. The varieties marked with a star, are joint recordings of E. Blancquaert, K. Boelens and G. van der Woude.

broek 1' refers to the Frisian variety and 'Donkerbroek 2' to the Stellingwerf variety. Appelscha is located in the Stellingwerf area. In addition to a Low Saxon Stellingwerf variety, a Frisian variety is spoken, introduced by Frisian laborers who moved to Appelscha at the time of peat-diggings. We process the Frisian variety as a language island. Below, 'Appelscha 1' refers to the Stellingwerf variety and 'Appelscha 2' to the Frisian variety.

For more details about the RND data, and some preprocessing of it to increase consistency, see Heeringa (2004, pp. 214-226). A detailed discussion about the selection of varieties and words from the RND can be found in Heeringa (2001).

3. Measuring linguistic distances

The maps of Hof and van der Veen are based on phonological isoglosses. Van der Veen (2001, p. 113) writes that "classification maps on other levels than the phonological one have never been drawn". Maps based on morphological, lexical or syntactic differences do not exist. In this section we present the Levenshtein distance as a tool for measuring linguistic distances. Levenshtein distance processes lexical, phonetic and morphological differences. While the use of isoglosses gives a very categorical view of dialect differences: either a dialect is different from another dialect or it is not (see further Chambers & Trudgill 1998, pp. 89-103), the use of Levenshtein distance allows us to note gradual linguistic distances.

In 1995 Kessler introduced the use of the Levenshtein distance as a tool for measuring linguistic distances between language varieties. The Levenshtein distance is a string edit distance measure and Kessler applied this algorithm to the comparison of Irish dialects. Later the same technique was successfully applied to Dutch (Nerbonne et al., 1996, Heeringa, 2004, pp. 213-278), Sardinian (Bolognesi & Heeringa, 2002), Norwegian (Gooskens & Heeringa, 2004b) and German (Nerbonne & Siedle, 2005). Below, we give a brief explanation of the methodology. For a more extensive explanation see Heeringa (2004, esp. pp. 121-135).

3.1 Algorithm

Using the Levenshtein distance, two dialects are compared by comparing the pronunciation of words in the first dialect with the pronunciation of the same words in the second. It is determined how one pronunciation is changed into the other by inserting, deleting or substituting sounds. Weights are assigned to these three operations. In the simplest form of the algorithm, all operations have the same cost, e.g. 1. Assume *candle* (Frisian: *kears*, Dutch: *kaars*) is pronounced as [kjɛs] in the dialect of Makkum, and as [keəs] in the dialect of Nes[1]. Changing one pronunciation into the other can be done as follows (ignoring suprasegmentals and diacritics for this moment)[2]:

1. The data is taken from sentence 59 of part 15 of the *Reeks Nederlandse Dialectatlassen* and available via: http://www.let.rug.nl/~heeringa/ dialectology/atlas/rnd/.
2. The example should not be interpreted as a historical reconstruction of the way in which one pronunciation changed into another. We just show that the distance between two arbitrary pronunciations is found on the basis of the least costly set of operations mapping one pronunciation into another.

kjɛs	delete j	1
kɛs	subst. ɛ/e	1
kes	insert ə	1
keəs		
		3

In fact many sequence operations map [kjɛs] to [keəs]. The power of the Levenshtein algorithm is that it always finds the cost of the cheapest mapping.

Comparing pronunciations in this way, the distance between longer pronunciations will generally be greater than the distance between shorter pronunciations. The longer the pronunciation, the greater the chance for differences with respect to the corresponding pronunciation in another variety. Because we prefer to view words as contributing equally to dialect pronunciation differences, we normalize pronunciation difference based on word length. To achieve this the sum of the operations is divided by the length of the longest alignment which gives the minimum cost. The longest alignment has the greatest number of matches. In our example we have the following alignment:

k	j	ɛ		s
k		e	ə	s
	1	1	1	

The total cost of 3 (1+1+1) is now divided by the length of 5. This gives a word distance of 0.6 or 60%.

3.2 Gradual weights

The simplest versions of this method are based on a notion of phonetic distance in which phonetic overlap is binary: non-identical phones contribute to phonetic distance, identical ones do not. Thus the pair [i,ɒ] counts as different to the same degree as [i,ɪ]. In more sensitive versions phones are compared on the basis of their feature values, so the pair [i,ɒ] counts as

more different than [i,ɪ]. However, it is not always clear how to weight the various contributions of the different features. The version which we use in this paper is based on the acoustic signal which realizes the sounds and which is normally visualized as a spectrogram. A spectrogram is the visual representation of the acoustical signal, and the visual differences between the spectrograms are reflections of the acoustical differences. When using spectrograms we do make decisions about the weight of the different features. The spectrograms were made on the basis of recordings of the sounds of the International Phonetic Alphabet as pronounced by John Wells and Jill House on the cassette *The Sounds of the International Phonetic Alphabet* from 1995[3]. The different sounds were isolated from the recordings and monotonized at the mean pitch of each of the two speakers with the program PRAAT.[4] Next, with PRAAT a spectrogram was made for each sound using the so-called Barkfilter which is a perceptually oriented model. On the basis of the Barkfilter representation, segment distances were calculated. The way in which this was done is described extensively in Heeringa (2004 pp. 79-119), and more briefly in Gooskens and Heeringa (2004a).

3.3 Logarithmic weights

In perception, small differences in pronunciation may play a relatively strong role in comparison to larger differences. Therefore we used logarithmic segment distances. The effect of using logarithmic distances is that small distances are weighed relatively more heavily than large distances. Since the logarithm of 0 is not defined, and the logarithm of 1 is 0, distances are increased by 1 before the logarithm is calculated. To obtain percentages, we calculate:

$$(ln(distance + 1) / ln(maximum\ distance + 1)) * 100$$

3.4 Allowed matches

To reckon with syllabification in words, the Levenshtein algorithm is adapted so that only vowels may match with vowels, consonants with con-

3. See http://www.phon.ucl.ac.uk/home/wells/cassette.htm.
4. The program PRAAT is a free public-domain program developed by Paul Boersma and David Weenink at the Institute of Phonetic Sciences of the University of Amsterdam and available at http://www.fon.hum.uva.nl/ praat.

sonants, the [j] or [w] with a vowel (or vice versa), the [i] or [u] with a consonant (or vice versa), and a central vowel (in our research only the schwa) with a sonorant (or vice versa). So the [i], [u], [j] and [w] align with anything, and otherwise vowels align with vowels and consonants with consonants. In this way unlikely matches (e.g. a [p] with a [a]) are prevented.

3.5 Distances

As mentioned in Section 2, the dialect comparisons are made on the basis of 125 words. When comparing two dialects we get 125 Levenshtein distances. Now the dialect distance is equal to the sum of 125 Levenshtein distances divided by 125. When the word distances are presented in terms of percentages, the dialect distance will also be presented in terms of percentages. All distances between the 54 language varieties are then arranged in a 54 × 54 matrix. Since distance is symmetric ($d(a,b)=d(b,a)$) half of the matrix is redundant.

3.6 Classification

On the basis of the distance matrix we applied hierarchical cluster analysis. The goal of clustering is to identify the main groups. The groups are called *clusters*. Clusters may consist of subclusters, and subclusters may in turn consist of subsubclusters, etc. The result is a hierarchically structured tree in which the dialects are the leaves (Jain and Dubes, 1988). Several alternative clustering techniques exist. We used the *Unweighted Pair Group Method using Arithmetic averages* (UPGMA), since dendrograms generated by this method reflected distances which correlated most strongly with the original Levenshtein distances (r=0.9679) (See Sokal and Rohlf, 1962, and Heeringa, 2004, p. 150-153).

4. The role of transcriber differences
4.1 Detecting transcriber groups

In Section 1 above we mentioned Hof's map and van der Veen's map. We will compare our results with these two maps and with other data in literature. Our classification obtained on the basis of Levenshtein distances can

be found in Figure 5. The main division consists of a cluster of Frisian varieties (the cluster Hindeloopen…Schiermonnikoog) and a cluster containing Low Saxon and Town Frisian varieties (the cluster Tjalleberd2...Midsland).

Most distinct within the Frisian group are Schiermonnikoog, Oosterend, West-Terschelling and Hindeloopen. Apart from these four varieties, we find a division in a northern and a southern group. We labeled the northern group as 'W' (van der Woude), and the southern group as 'B' (Boelens) in the dendrogram in Figure 5. From the map in Figure 2 appears that this division perfectly reflects the division between the two transcribers who made the recordings in the Frisian area. The problem of finding transcriber borders was also found and discussed by Nerbonne & Kleiweg (2003) for the North-American LAMSAS data.

Looking at the Low Saxon/Town Frisian cluster, we find a Low Saxon cluster (Tjalleberd 2…Zoutkamp) and a Town Frisian cluster (Franeker…Midsland). In the Town Frisian cluster, Midsland is most distinct. Apart from this dialect, we again find a perfect division in transcriber groups. The cluster Franeker…Sint Annaparochie contains van der Woude varieties, and the cluster Bolsward…Heerenveen contains Boelens varieties. Exceptions are Franeker, Sneek and Dokkum, which are joint recordings of Blancquaert, Boelens and van der Woude.

4.2 Eliminating transcriber influences

In both Hof's map and van der Veen's map the Frisian varieties are divided in three groups: Clay Frisian, Wood Frisian and South Corner (see Figure 3 and 4). The same three groups can be found in our results, notwithstanding the division in transcriber clusters. Considering the van der Woude cluster, the subcluster Bakkeveen…Rottevalle contains Wood Frisian varieties, and the subcluster Holwerd…Hallum Clay Frisian varieties. Looking at the Boelens cluster, the subcluster Lemmer…Koudum contains mainly South Corner varieties, the subcluster Tjalleberd 1…Donkerbroek1 contains Wood Frisian varieties and the subcluster Makkum…Spannum contains Clay Frisian varieties.

Can we eliminate the transcriber border? This question was also discussed by John Nerbonne in a paper, which he presented on the NWAVE32 conference (Philadelphia, 2003). Using the LAMSAS data he showed that the effect of transcriber borders can be eliminated by converting the linguis-

Figure 3. Division of the our set of RND varieties according to Hof (1933). In grey areas a Frisian dialect is spoken, in white areas a non-Frisian dialect. The islands are not taken into account in Hof's map.

tic distances to z-scores. Using z-scores, the mean of a group will be equal to 0 and the standard deviation will be equal to 1. When having two groups, three sets of z-scores are distinguished: z-scores in the one group, z-scores in the other group and z-scores between the two groups. Unfortunately we cannot apply this methodology to our data set since the number of dialects per transcriber group is too small. We will use a different, but related methodology which relies on the assumption that a division in Clay Frisian, Wood Frisian and South Corner may be supposed beforehand.

In attempting to eliminate the transcriber border, we will focus on Clay Frisian and Wood Frisian, and ignore the South Corner groups since this group is only found in the van der Woude group. Our basic strategy is to at-

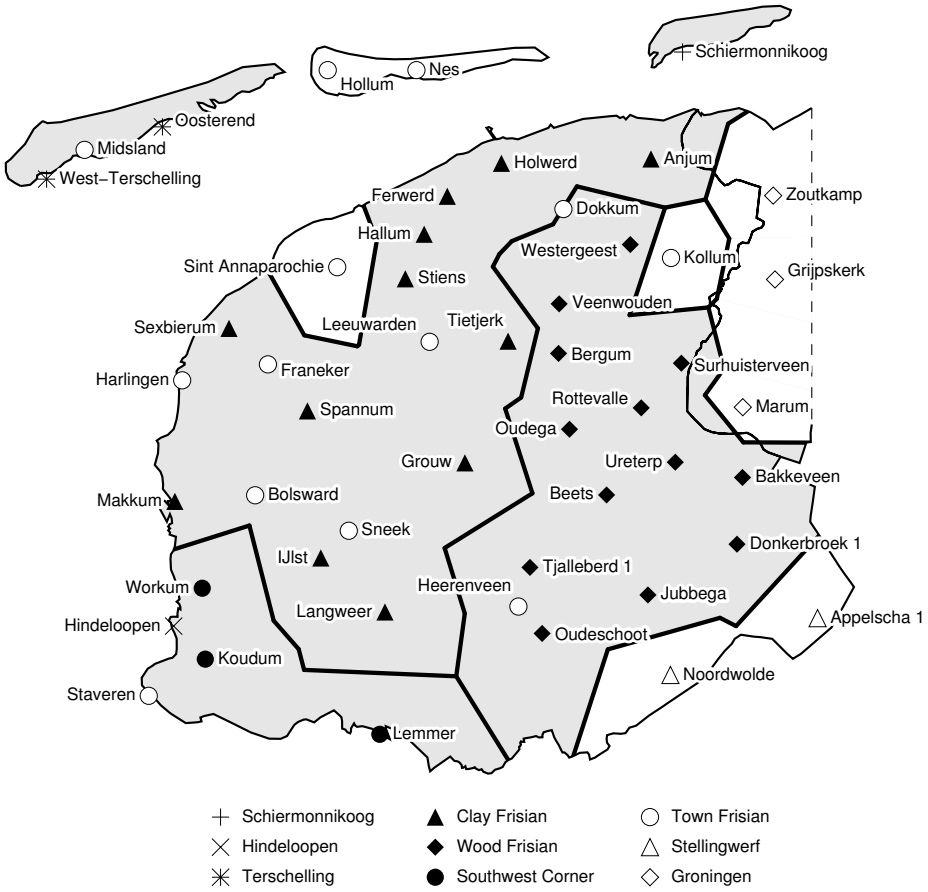

Figure 4. Division of our set of RND varieties according to van der Veen (2001). In grey areas a Frisian dialect is spoken, in white areas a Town Frisian or Low Saxon dialect.

tempt to eliminate the effect of those words which appear particularly subject to transcriber distortions. We describe now how we attempted to identify those words.

We calculated the average distance between the transcriber groups (Boelens group and van der Woude group) per word, and we calculated the average distance between two dialect groups (Clay Frisian and Wood Frisian)

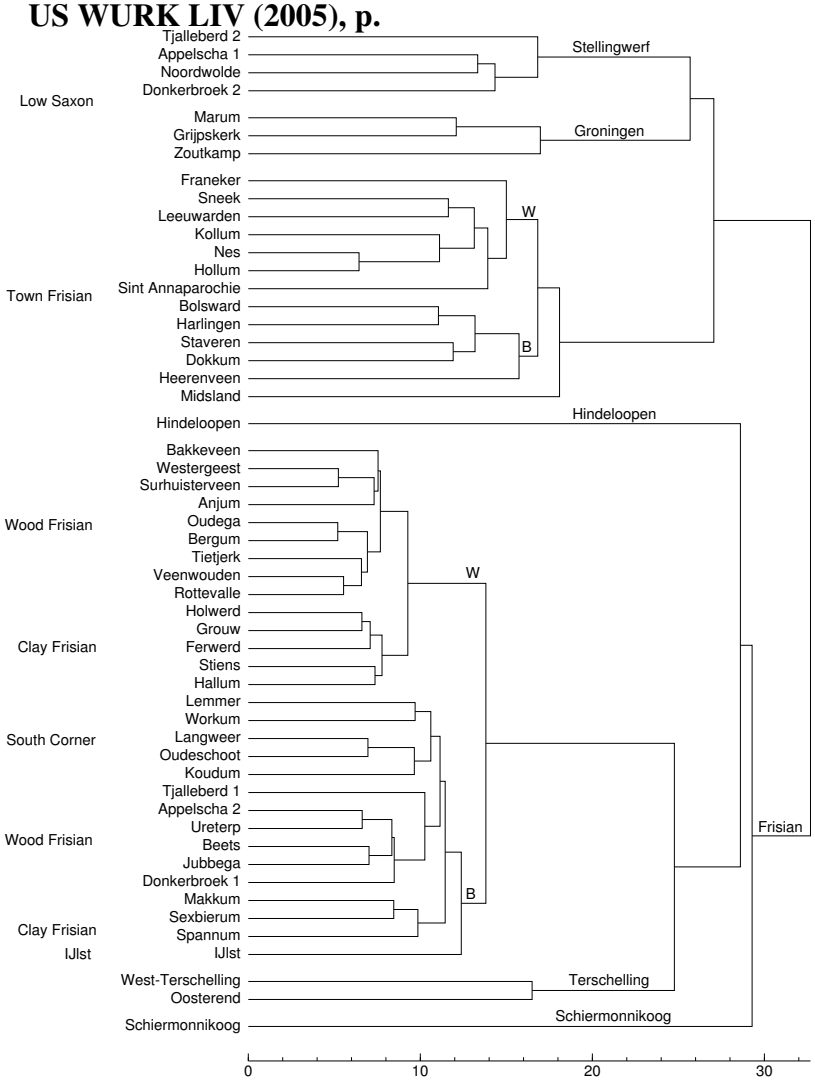

Figure 5. Classification of the 54 Frisian and Frisian mixed varieties of the RND. The scale distance shows percentages. The varieties of a cluster with a 'B' are transcribed by K. Boelens, and varieties of a cluster with a 'W' are transcribed by G. van der Woude. Exceptions are the varieties of Franeker, Sneek and Dokkum. The transcriptions of these varieties are joint recordings of Blancquaert, Boelens and van der Woude.

per word. Next we subtracted the distance between dialect groups from the distance between transcriber groups for each word. This results in positive and negative values. A positive value for one word means that transcriber differences were greater than dialect group differences for that word and,

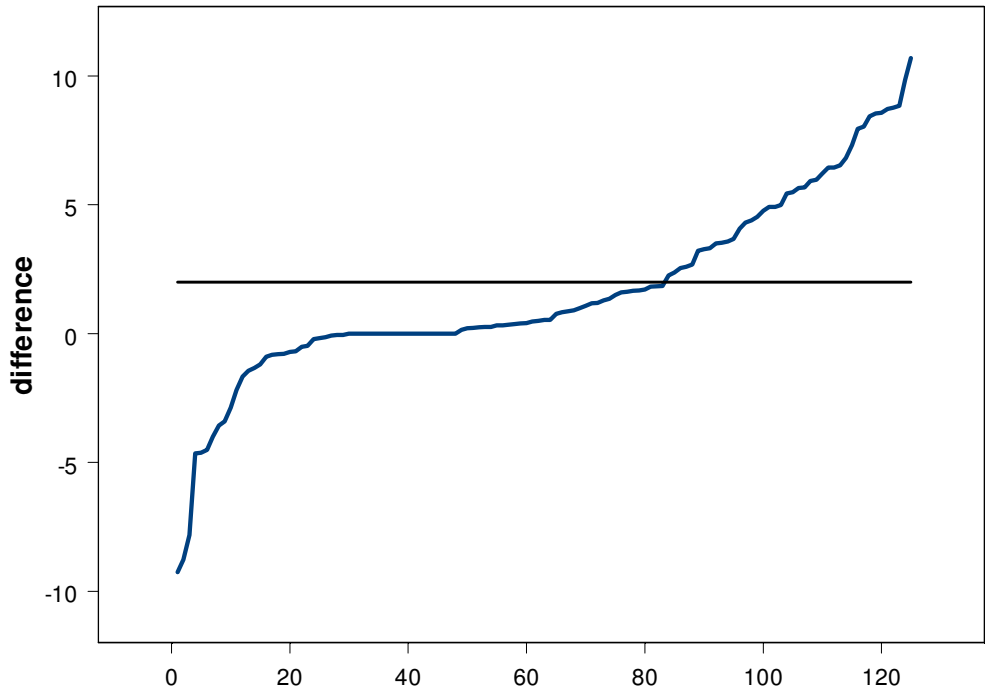

Figure 6. Per word the distance between two transcriber groups (Boelens group and van der Woude group) and the distance between two dialect groups (Clay Frisian and Wood Frisian) was calculated. Next we subtracted the distance between dialect groups from the distance between transcriber groups for each word. In the graph these distance differences are shown for each word in sorted order. The x-axis gives the rank order numbers of the words. When ignoring words with distance differences greater than 1.9% (distances above the horizontal line, 42 words), i.e. words subject to major transcriber differences, we obtain a less distorted picture of dialect pronunciation differences.

also that the variation of the word suggests a north-south division rather than a division in Clay Frisian and Wood Frisian. In case of a negative value it is the other way around. In the graph in Figure 6 these distance differences are shown for each word in sorted order. The x-axis gives the ranks of the words.

We studied the word variation of the words with the highest distance differences, in order to find out which phenomena are responsible for the north-south division. The word *litten* (English: 'to let', Dutch: 'laten') was transcribed by Boelens as [lɪtˀnˀ], and by van der Woude as [lɪtn̩]. Also for words with similar endings like *roppen* (English: 'called', Dutch: 'ge-

roepen') and *potten* (English: 'jars', Dutch: 'potten') we found a similar difference. Another difference is found in the word *jier* (English: 'year', Dutch: 'jaar'), transcribed by Boelens as [jiˑᵊr], and by van der Woude as [iˑᵊr]. We found also differences in the notation of length. The word *see* (English: 'sea', Dutch: 'zee') is transcribed by Boelens as [se:], and by van der Woude as [seˑ]. We found something similar in words like *twa* (English: 'two', Dutch: 'twee') and *hûs* (English: 'house', Dutch: 'huis'). We found also differences in the notation of nasalized centering diphthongs. The word *gean* (English: 'to go', Dutch: 'gaan') is transcribed as [geˑᵊ] by Boelens, and as [gẽˑẽ] by van der Woude.

We may be sure that in the cases just mentioned, the different notations of syllabification, whether the [i] is preceded by [j] or not, length differences and whether a vowel is followed by a schwa or not, are just transcriber differences, not real differences between northern and southern dialects. From the examination of the phenomena which are responsible for the north-south border, we may conclude that this border is not significant. Nevertheless we are hesitant to simply ignore these phenomena since they can represent real dialect differences (e.g. between Clay Frisian and SouthCorner). We rather prefer to leave out the *words* which cause the transcriber division. In that case, the same phenomena are still found in the remaining words, leading to right divisions.

We found that transcriber effects are eliminated when ignoring words with distance differences higher than 1.9%. These are the last 42 words in the graph in Figure 6. If we eliminate these, we obtain a division into a Clay Frisian group, a Wood Frisian group, and a South Corner group.

4.3 Using fewer words

The results in the rest of this paper are based on measurements based on 125-42=83 words. To be sure that the number of words is a sufficient basis for a reliable Levenshtein analysis, we calculated Cronbach's α (see Heeringa 2004, pp. 170-173). When using 125 words we found that Cronbach's α=0.98, and when using 83 words Cronbach's α=0.97. For both 125 words and 83 words we get a very high Cronbach's α. When using 83 words Cronbach's α is hardly lower than when using 125 words, so there should be no problem in basing our results on the subset of 83 words.

5. Classification

In Section 3 we described the way in which distances between varieties are calculated using Levenshtein distance. In Section 4 we presented a dendrogram obtained on the basis of these distances. Examining the dendrogram we found the influence of transcriber differences. Fortunately we found a way to eliminate the influence of transcriber differences. On the basis of distances which are not influenced by transcriber differences we obtained a new dendrogram. This new dendrogram is given in Figure 7. The main groups of this dendrogram are also geographically visualized in Figure 8. We will compare our results with classification results in literature. However, we have to be aware of the fact that the classification results of literature are in some cases older, and in other cases newer than our RND data set. We consulted mainly Hof (1933, pp. 4-6), Gorter (2001, p. 75) and van der Veen (2001, p. 98-102, 113). In Figures 2 and 3 the maps of Hof and van der Veen respectively are given. However the maps are simplified, i.e. they give the classification according to Hof and van der Veen applied to our set of RND varieties. In our maps the course of the borders depends on the density of the grid. For the exact course of the borders one should consult the maps of Hof (2001, p. 14a) and van der Veen (2001, p. 99). The latter is also found in Visser (1997, p. 3). Both Hof and van der Veen emphasize that sharp boundaries between the main dialect groups do not exist (Hof, 1933, p. 2-3, van der Veen, 2001, p. 113).

The main division consists of a cluster of Frisian varieties (the cluster Hindeloopen…Schiermonnikoog) and a cluster containing Low Saxon and Town Frisian varieties (the cluster Tjalleberd2…Heerenveen). Looking at the Low Saxon/Town Frisian cluster, we find a Low Saxon cluster (Tjalleberd 2…Zoutkamp) and a Town Frisian cluster (Midsland…Heerenveen). In Section 5.1 we discuss the Frisian group, in Section 5.2 the Town Frisian group and in Section 5.2 the Low Saxon group.

5.1 Frisian

5.1.1. Archaic varieties

According to Visser (1997) the dialects of Schiermonnikoog, western and eastern Terschelling and Hindeloopen have not undergone certain innova-

tions which other Frisian varieties have. Winkler (1874, I) supposed that the inhabitants of Schiermonnikoog belonged to another tribe than those of mainland West Frisia (p. 452). Possibly the Schiermonnikoog dialect is a relic of the Frisian dialect which was spoken in the province of Groningen before the Reformation. When the Saxons invaded the province of Groningen and mixed with the original inhabitants, they never reached the island of Schiermonnikoog (p. 453). Hof (1919) on the other hand suggests that the Schiermonnikoog dialect originated from the northeastern part of the province of Friesland. He studied proverbs in rhyme, versified by Reyner Bogerman around 1500 and probably written in the dialect of Anjum. Hof found striking similarities between the Bogerman Frisian and the Schiermonnikoog variety when examining vocalism. Furthermore in both varieties a singular t-apocope was found.

The hypothesis of Winkler and the finding of Hof do not necessarily contradict each other since Anjum is very close to the province of Groningen. To determine the origin of the Schiermonnikoog dialect, the dialect should be systematically compared with both Old West Frisian and Old East Frisian.

In the work of Hof (1933) the islands Schiermonnikoog and Terschelling are ignored. According to Gorter (2001, p. 75) these dialects "can be thought of as being heavily influenced by both Dutch and Frisian, or as a sort of mixed language." In 1874 Winkler wrote that the dialect of the western part of Terschelling is closer to Dutch than the dialect of the eastern part of this island (Winkler, 1874, II, p. 15).

In Hof (1933) no attention is paid to the dialect of Hindeloopen. But Hoppenbrouwers and Hoppenbrouwers (2001) describe Hindeloopen as an isolated place inhabited by fishermen with an archaic dialect (p. 99). The phonological distance between Hindeloopen and the main dialects is substantial (van der Veen, 2001).

5.1.2 Clay Frisian, Wood Frisian, South Corner

Gorter (2001) mentions that "the West Frisian speech community is basically homogeneous and all the dialect varieties are mutually comprehensible." (p. 75). In Hof's map the Frisian area is divided in three areas: Clay Frisian (Frisian: Klaaifrysk), Wood Frisian (Frisian: Wâldfrysk) and South Corner (Frisian: Súdhoeksk). These three groups are also found in van der Veen's map. Apart from Schiermonnikoog, West-Terschelling, Oosterend

and Hindeloopen we found the same three groups in our Frisian group as well. The subcluster Koudum…Workum represents South Corner, the subcluster Stiens…Spannum represents Clay Frisian and the subcluster Tjalleberd 1…Tietjerk represents Wood Frisian.

Hof's map and van der Veen's map differ with regard to the northeastern part of West Frisia. In Hof's map this area belongs to Wood Frisian, but in van der Veen's map it is a part of the Clay Frisian area. Gorter (2001, p. 75) also pays attention to this northeastern group: is it a separate variety, or does it belong to Clay Frisian or Wood Frisian? On p. 3 Hof writes, that he regards it as a separate variety, first calling it North Clay Frisian, but later recanting this point. On his map on p. 14a North Clay Frisian and Wood Frisian form one group. However, just as Hof did earlier, Visser (1997) distinguishes four groups again, where North Clay Frisian (or Northeast corner) is the fourth group. He judges North Clay Frisian to be closer to Wood Frisian than to Clay Frisian. This accords with Hof's map (1993). Visser based his classification on five phonological phenomena (see p. 4).

In our data set the northeastern area is represented by Anjum, Holwerd, Ferwerd, Hallum and Stiens. In the dendrogram the dialect of Anjum is found among the Wood Frisian varieties which accords with Hof and Visser. The dialects of Holwerd, Hallum and Stiens form a subcluster in the Clay Frisian group and also the dialect of Ferwerd is found among the Clay Frisian varieties. This accords with van der Veen. Especially with regard to the position of 'North Clay Frisian', our division is closer to that of van der Veen. We have the most confidence in van der Veen's map because of his well-considered choice and weighting of the isoglosses.

Considering our border between Clay Frisian and Wood Frisian in Figure 8, we found the dialects of Beets and Oudeschoot unexpected members of Clay Frisian. In both Hof's map and van der Veen's map they belong to Wood Frisian. Possibly the two varieties are transition dialects. Both Hof (1933, p. 2-3) and van der Veen (2001, p. 113) allude to the vagueness of dialect borders.

Winkler (1874, I) wrote that in Workum, Hindeloopen, Staveren, Makkum, Molkwerum, Koudum, Warns, Scharl and Laaksum a South Corner dialect was originally spoken. The South Corner was a separate area. People in Makkum, Workum, Hindeloopen and Molkwerum made their living from seatrade with the Baltic. Later on, the sea trade disappeared, and agriculture and cattle breeding became the main means of sustenance. This may have led to a closer orientation with the rest of West Frisia. The result was that

South Corner Frisian dissolved into 'common' Frisian. Only in Hindeloopen was the original South Corner dialect kept alive (p. 436). In the maps of Hof and van der Veen a South Corner area is nevertheless still distinguished. However, Makkum is not longer found in the South Corner area. Obviously in the West the border has been shifted a little bit to the south. Our results may thus be seen to agree with both Hof and van der Veen.

5.2 Town Frisian

To keep a clear overview, we labeled the cluster Midsland…Heerenveen as 'Town Frisian' in our dendrogram in Figure 7. Actually this is not completely right, since only the varieties of Leeuwarden, Sneek, Bolsward, Franeker, Harlingen, Staveren and Dokkum are Town Frisian. These cities are dialect islands in the Frisian dialect continuum. Town Frisian dates from the 16th century. The existence of this variety is the result of "a change in government (new rulers with immigrant civil servants) and increased trade contacts with the towns in the province Holland" (Gorter, 2001, Jonkman, 1993, Van Bree, 2001). According to Hof, the phonetic system is Dutch, but articulation and syntax are Frisian. Semantically they are closest to Frisian as well. Of all Dutch varieties, Town Frisian is originally most closely related to the North Holland dialects (Van Bree, 2001). Sint Annaparochie, Midsland, Hollum, Nes and Kollum are not Town Frisian varieties. Nevertheless, they are very close to Town Frisian, as can be seen in the dendrogram. The Town Frisian or Town Frisian related varieties can be found in the map in Figure 8.

Most distinct in our 'Town Frisian' cluster is the dialect of Heerenveen. In Daan & Blok's dialect map (1969) the dialect of Heerenveen is also marked as a Town Frisian dialect. During, but mainly after peat digging, many merchants came from the Frisian cities to Heerenveen. Later on, Heerenveen became the chief town of neighboring municipalities (Schoterland and Eangwirden), and a district court, a cantonal court and tax offices were settled there. The result was that Frisian disappeared nearly completely in the 19th century. The Heerenveen dialect is more related to Town Frisian, characterized by Hof as 'Dutch, spoken with Frisian mouth'. Although Hof pays attention to the Heerenveen variety in a separate paragraph, it is not mentioned by Gorter (2001).

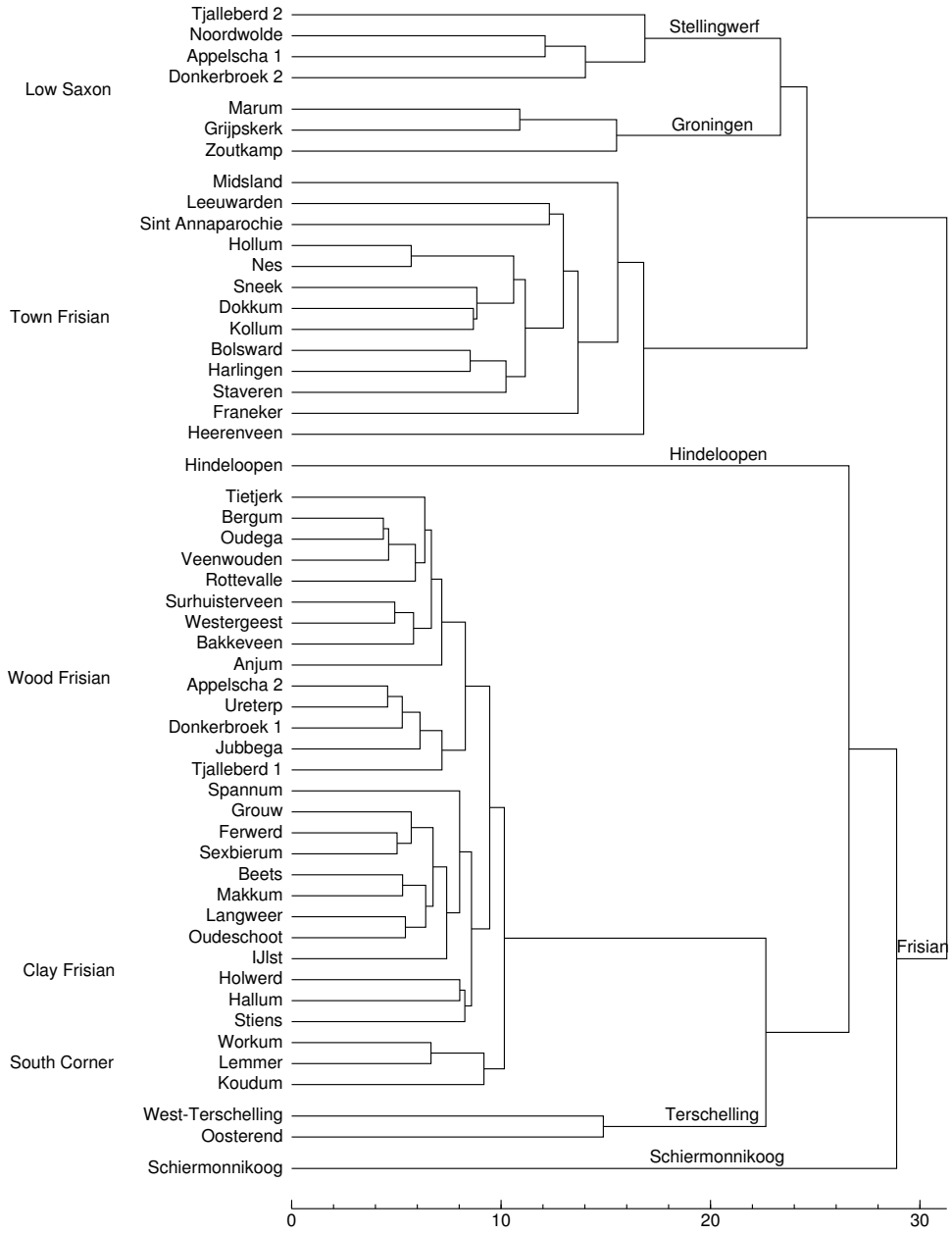

Figure 7. Classification of the 54 Frisian and Frisian mixed varieties of the RND after eliminating words particularly susceptible to transcriber effects. The scale distance shows percentages. In contrast to the dendrogram in Figure 5, the influence of different transcribers cannot be found.

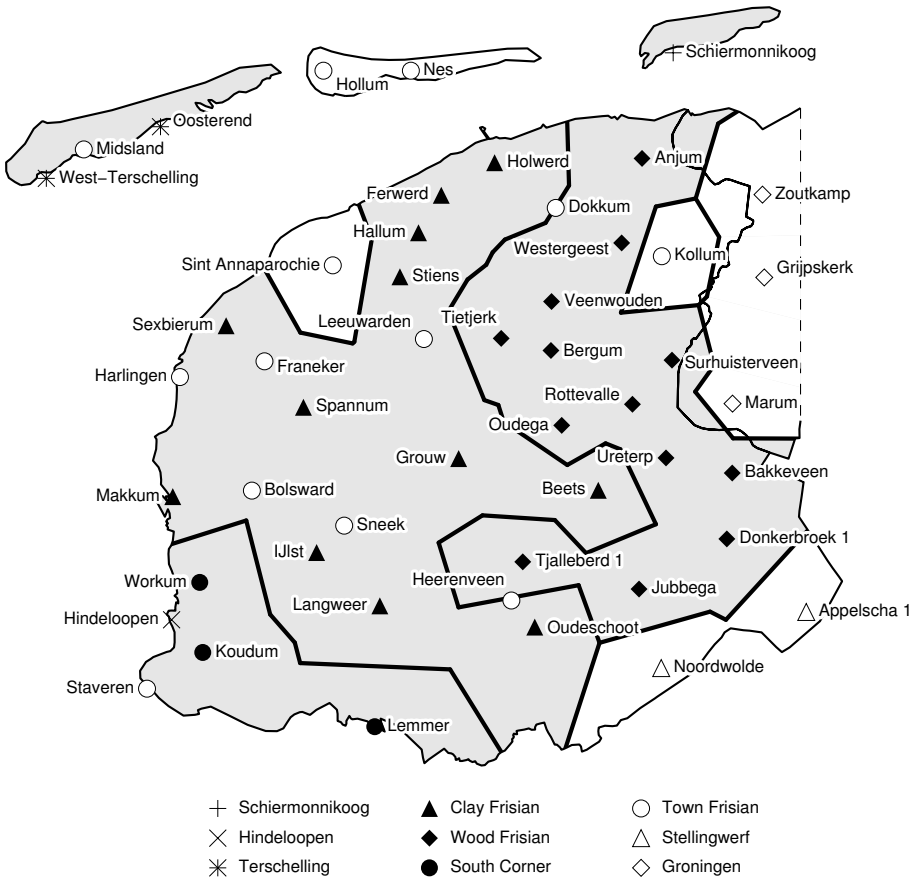

Figure 8. Division of our set or RND varieties, derived from the dendrogram in Figure 7. In grey areas a Frisian dialect is spoken, in white areas a Town Frisian or Low Saxon dialect. In Appelscha and Donkerbroek, both a Frisian and a Stellingwerf dialect is spoken. In Tjalleberd both a Frisian and an Overijssel dialect is spoken.

Going one level deeper in the Town Frisian cluster, we find the Midsland variety most distinct. On the islands of Terschelling and Ameland, Frisian was the original language (van der Veen, 2001, p. 112). Later on this situation changed. In Midsland, the middle part of the island of Terschelling, a dialect is spoken which is strongly related to Town Frisian. For centuries, Midsland was the administrative and economic center of the island, comparable with the cities on the Frisian mainland (Hoppenbrouwers, 2001). From the 16th century on the island lived under Dutch rule.

Going a further level deeper, we find a subcluster Leeuwarden…Sint Annaparochie and a subcluster Hollum…Franeker. The variety of Sint Annaparochie represents the dialect of Het Bildt. Het Bildt is a small area in the Northwest of the province of West Frisia and "consists mainly of land reclaimed from the Middle Sea by the beginning of the 16th century" (Gorter, 2001, p. 75). According to Hof, both Frisian and Dutch colonists settled in this area. Gorter (2001) mentioned that farmers from the province of South Holland came to this area after it was impoldered. Despite the different history of the Town Frisian dialect of Leeuwarden and the Het Bildt dialect of Sint Annaparochie, they appear to be rather close. Since the varieties are geographically rather close, we speculate that they may have influenced each other.

In the subcluster Hollum…Staveren we find the dialects of Hollum and Nes clustered closely together. The dialects are found on the island of Ameland. We mentioned already that Frisian was the original language on the islands of Terschelling and Ameland (van der Veen, 2001, p. 112). Afterwards on the entire island of Ameland a variety was spoken which is close to Town Frisian. In 1825 J.H. Halbertsma wrote in his *Over de uitspraak van het Landfriesch* that on the island of Ameland the same dialect is spoken as for example in Dokkum and Leeuwarden (Breuker, 1973, p. 50). Van Bree (2001, p. 135) explains the closeness with Town Frisian by the fact that inhabitants of Ameland stayed for some time in Holland, and later on they returned to Ameland (*om-utens*-going). Van Bree (2001, p.135) pays also attention to the fact that the island was autonomous until 1798. Ameland is represented by Hollum and Nes in our data set (see Figure 2). According to van der Veen (2001, p. 113) the western variety of Ameland resembles Frisian more closely than the eastern one.

Furthermore the dialects of Sneek, Dokkum and Kollum appear as a small group. Dokkum and Kollum are geographically rather close, but Sneek is distant from the other two. The dialect of Kollum represents the Kollumerland area. Apart from Kollum this area includes the eastern part of Kollumerland. In this area Frisian was still spoken at the end of the Middle Ages (van der Veen, 2001, p. 99). Later on, rural Frisian was dispelled by a dialect which is related to Town Frisian. A possible explanation is given by Kloeke (1927). On the one hand Kollum is rather distant from the city of Groningen, so a Low Saxon dialect was not adopted. On the other hand the Kollum inhabitants do not regard themselves as real Frisians, so they do not speak Frisian. They found their identity in contacts with the outside world,

especially Holland. Therefore, they speak Town Frisian (see also Van Bree, 2001, p. 135). To the east the dialect changes in the dialect of Burum, which is close to the dialect of the neighboring Westerkwartier area. The dialect of Westerkwartier belongs to the province of Groningen and is a Low Saxon variety.

We found another small group consisting of the Town Frisian varieties of Bolsward, Staveren and Franeker. The three varieties are found along the west coast of Frisia. Possibly the Staveren dialect is the oldest Town Frisian variety (Hof, 1933, p. 5).

In 1874 Winkler wrote that the upper social classes in Workum also speak Town Frisian. However, in the maps of Hof (1933), Daan & Blok (1969) and van der Veen (2001), Workum is not marked as a Town Frisian city. This accords with our results, where Workum belongs to the South Corner group. Possibly the Town Frisian dialect has been completely displaced by Frisian.

5.3 Low Saxon

The Low Saxon cluster consists of a Stellingwerf group (Tjalleberd 2…Donkerbroek 2) and a Groningen group (Marum…Zoutkamp). Both the Stelllingwerf varieties and the Groningen varieties can be found in the map in Figure 8. In Section 5.3.1 we discuss the Stellingwerf group, and in Section 5.3.2 the Groningen group.

5.3.1 Stellingwerf

Within the Stellingwerf group, Tjalleberd 2 is most distinct. Tjalleberd is located in the Frisian dialect continuum. As mentioned in Section 2, both a Frisian and a Low Saxon dialect are spoken in Tjalleberd, where Tjalleberd 2 refers to the Low Saxon variety. Hof (1933, p. 6) mentions that in Tjalleberd-Luinjeberd and in Oldeouwer an Overijssel dialect is spoken which is strongly intermixed with Frisian. The Tjalleberd dialect is also known as the dialect of Giethoorn, or *Gietersk*. According to Daan & Blok's map (1969) Tjalleberd belongs to the Frisian dialect continuum. Is it not marked as a dialect island.

In the Frisian volume of the RND (volume 15), the author (K. Boelens) mentions that the dialect of Tjalleberd and Luinjeberd is closer to the Stellingwerf dialect than to Frisian. Inhabitants of Luinjeberd call their dialect

the dialect of Tjalleberd as well. Boelens tells us further that Dutch is spoken only in a few families while 90% of the population speaks Frisian and 10% speaks the dialect of Tjalleberd. Parents speak the dialect of Tjalleberd with each other, but Frisian with their children. In the schoolyards, the children speak Frisian.

Hof (1933, p. 6) describes that the Drenthe-Overijssel element can be observed strongly in the towns of St. Johannesga, Rotsterhaule, Rotstergaast, Delfstrahuizen, Echten and Oosterzee, although it is decreasing since the peat digging ended or diminished and no new foreign peat diggers came to these towns. None of these locations are included in our data set.

Apart from Tjalleberd 2, we find a cluster of 'real' Stellingwerf varieties: Noordwolde, Appelscha 1 and Donkerbroek 2. The Stellingwerf area is found in the Southeast of the province of West Frisia. Both Winkler (1874, I, p. 496) and Hof (1933, p. 4,5) consider this area in the narrow sense. Construed narrowly, the area consists of the municipalities West- and Oost-Stellingwerf. These municipalities include the varieties in our data set: Appelscha and Noordwolde (see Figure 8). In Daan & Blok's map (1969) the Stellingwerf area also includes a part of the provinces of Drenthe and Overijssel. So in the awareness of the dialect speakers, the Stellingwerf dialect does not halt at the Frisian province border. Winkler (1874, I, p. 496) describes the Stellingwerf dialect as a Drenthe dialect, intermixed with Frisian sounds, words, forms, expressions and phrasings, pronounced with a different degree of a Frisian accent. Hoppenbrouwers and Hoppenbrouwers (2001, p. 101) also cite Sassen (1953, p. 101) who describes the dialect as a Drenthe variety which has become more Frisian, where the process of becoming more Frisian possibly started in the 13th century or earlier.

Appelscha is marked as a Frisian dialect island in the Stellingwerf dialect area in Daan & Blok's map (1969). As mentioned in Section 2 both a Frisian and a Low Saxon variety is spoken in this location, where Appelscha 1 refers to the Low Saxon variety. When peat digging begun, Frisian laborers came to Appelscha, introducing different Wood Frisian dialects. Hof (1933) tells that as the peat digging came to a finish, the use of Frisian was decreasing. Most inhabitants speak Frisian in the family. The RND transcription of the Stellingwerf variety in Appelscha was recorded by K. Boelens. He gives us some results of a school survey. The results of this survey were provided by H.J. Bergveld. In 135 families (49%), Frisian is the domestic everyday language, in 129 families (47%) the Stellingwerf dialect is the colloquial

speech, and in 9 families (3%) the Dutch language is used in daily speech. Nowadays the situation is likely to be different.

In Section 2 we mentioned that in Donkerbroek both a Frisian and a Low Saxon variety is spoken. Donkerbroek 2 refers to the Low Saxon variety. According to Daan & Blok's map (1969), Donkerbroek belongs to the Frisian area. However, Hof (1933) writes that the language boundary is not so sharp in this area. In Donkerbroek, most people speak Frisian, but the Stellingwerf dialect is spoken as well. In the RND K. Boelens mentioned that older native inhabitants speak the Stellingwerf dialect. He estimated that the half of the population speaks Frisian. In about 20 families Dutch is the domestic everyday language. In the schoolyards, children speak Frisian.

5.3.2 Groningen

East of the province of West Frisia, we find the province of Groningen. Originally Frisian was also spoken in this province, with the exception of the city of Groningen. In the city of Groningen a Low Saxon variety has spoken ever since the arrival of the Saxons from the province of Drenthe (Hoekstra, 2001, p. 139, Niebaum, 2001, p. 431). Later on, the Saxonization emanating from the city of Groningen was reinforced by the immigration of settlers who came from the east where Low German was spoken. This immigration had to do with the reclamation of land in East Groningen (Hoekstra, 2001, p. 139).

In the Westerkwartier area, the western part of the province Groningen which borders on West Frisia, the Frisian variety probably disappeared later than in the rest of the province, probably in the early part of the 16th century. In the Westerkwartier variety more Frisian elements can be found than in other Groningen dialects (Winkler, 1874, I, p. 417).

In our dendrogram the Westerkwartier area is represented by Marum and Grijpskerk. The 'pure' Groningen variety of Zoutkamp is clustered together with the two Westerkwartier varieties.

6. Comparison among Dutch, Frisian and Groningen

In this section we investigate the relation of Frisian, Town Frisian and Low Saxon varieties with respect to Standard Dutch, Standard Frisian and Groningen. To be able to compare the varieties with respect to Standard Dutch, we added a transcription of Standard Dutch. To assure consistency with the existing RND transcriptions, the Standard Dutch transcription is based on

the *Tekstboekje* of Blancquaert (1939). However, we transcribed words such as *komen*, *rozen* and *open* as [koˑmə], [roːzə] and [oˑpə]. In the *Tekstboekje* of Blancquaert these words ended in an [n], just as suggested by the spelling. For more details see Heeringa (2001). There is also no transcription of Standard Frisian available in the RND. Since Standard Frisian is closest to Grouw, we used the dialect of Grouw as Standard Frisian. Furthermore we compare Frisian, Town Frisian and Low Saxon varieties with the dialect of a former Frisian area, the province of Groningen. For this purpose we use the dialect of Zoutkamp, which is included in our RND data set.

In Section 6.1 we compare the Frisian, Town Frisian and Low Saxon varieties with Standard Dutch, in Section 6.2 we compare the same varieties with the Grouw variety, and in Section 6.3 a comparison is made with respect to the dialect of Zoutkamp.

6.1 Comparison with Standard Dutch

Figure 9 shows the distances of varieties and groups with respect to Standard Dutch. For the Clay Frisian, Wood Frisian, South Corner and Town Frisian varieties, we used the average per group. The groups are the same as found in Figure 7 and shown in Figure 8.

In the graph we find the dialect of Heerenveen closest to Dutch (26.1%). As explained in Section 5.2, the dialect of Heerenveen is of more recent origin than the dialects of the other Town Frisian cities. Heerenveen is followed by Kollum (28.6%), a dialect close to the Low Saxon area. Sint Annaparochie represents the dialect of Het Bildt. The dialect is just a little bit closer to Dutch than Town Frisian (28.8% versus 28.9%). Nes and Hollum are both found on the island of Ameland. Hollum is closer to Dutch than Nes (29.0% versus 29.4%). The Ameland varieties are followed by Midsland, which is found on the island of Terschelling (31.0%). Next we find Low Saxon varieties. The Westerkwartier varieties are clearly closer to Dutch than the Stellingwerf varieties (31.0% versus 35.5%). The Stellingwerf varieties are followed by the Groningen dialect of Zoutkamp (36.0%). We found the Overijssel dialect of Tjalleberd the most distant Low Saxon variety (39.5%).

In Section 5.1.1 we cited Gorter (2001) who describes the island dialects as “being heavily influenced by both Dutch and Frisian, or as a sort of mixed language." (p. 75). In our graph the dialect of West-Terschelling ap-

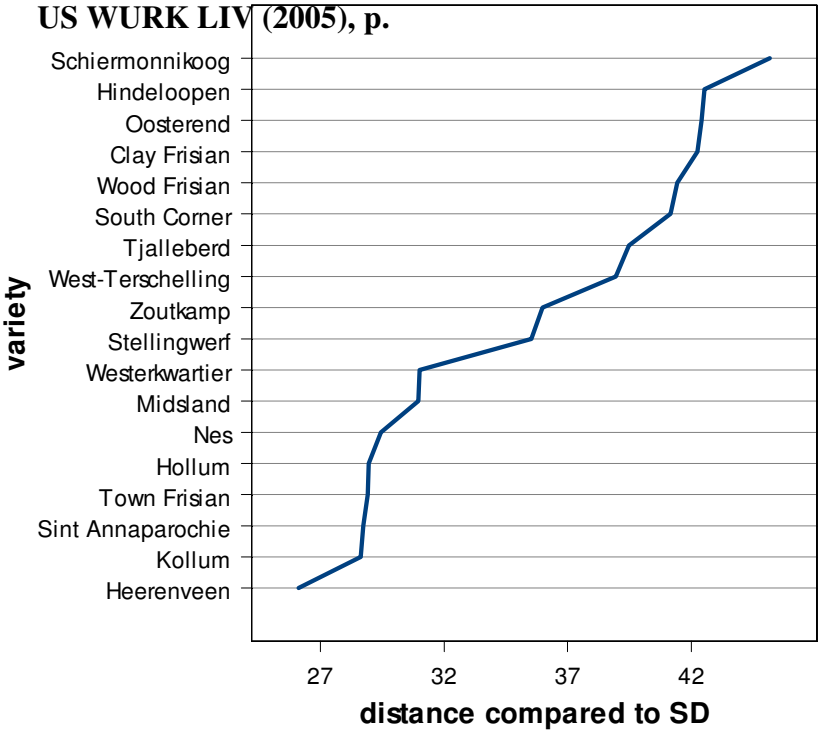

Figure 9. Distances compared to Standard Dutch (SD). The scale distance shows percentages.

pears as the Frisian variety which is closest to Dutch (39.0%). We found the dialect of Oosterend much more distant from Dutch (42.4%). The closeness of West-Terschelling to Dutch may be explained by the fact that the island Terschelling belonged to the province of North Holland until 1942. In Section 5.1.1 we mentioned that Winkler wrote that the western part of Terschelling is closer to Dutch than the eastern part. Our results, based on material of about three-quarter of a century later, show the same tendency.

We found the South Corner varieties closer to Dutch than both the Wood Frisian and Clay Frisian varieties (41.2% versus 41.4% and 42.2%). Since Dutch is close to the North Holland dialects, this outcome is geographically not unexpected. Wood Frisian is in turn closer to Dutch than Clay Frisian. Because of its geographic situation, Clay Frisian may be minimal influenced by other varieties such as Westerkwartier, Stellingwerf or North Holland varieties.

Most distant are the archaic varieties of Hindeloopen and Schiermonnikoog (42.5% and 45.2%). As described in Section 5.1.1, Hindeloopen is a relic of the former South Corner dialect. The distant position of Schiermon-

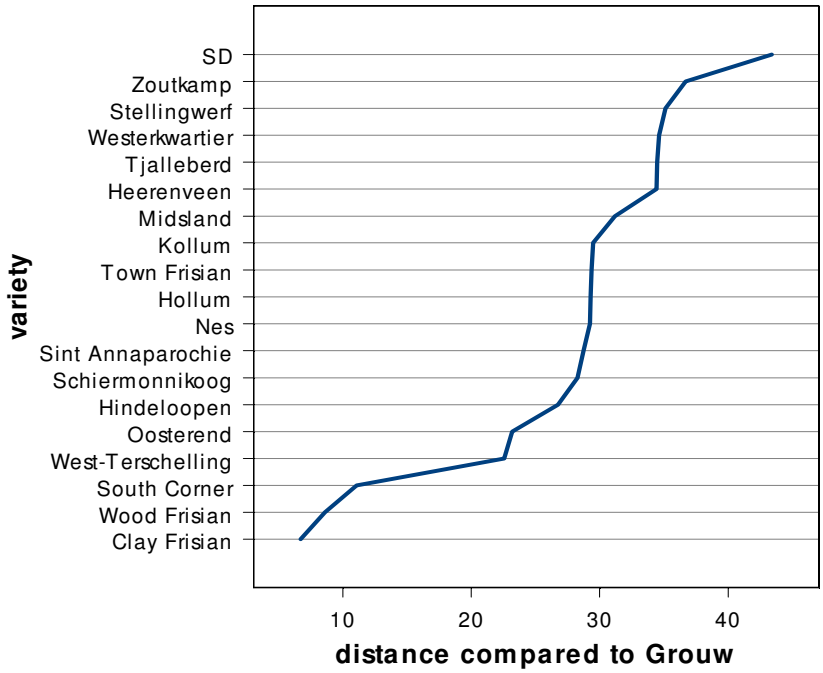

Figure 10. Distances compared to Grouw. The scale distance shows percentages.

nikoog contradicts Gorter (2001) who describes the island dialects as "being heavily influenced by both Dutch and Frisian, or as a sort of mixed language." (p. 75). Winkler wrote that of all Dutch dialects Schiermonnikoog is most distant to the written Standard Dutch language. Although written Dutch differs from spoken Dutch, of course they are strongly related. In Heeringa (2004, p. 274-276) an experiment is described in which 360 RND dialects of the complete Dutch language are compared to the spoken Dutch language. In that experiment the dialect of the island of Schiermonnikoog was found to be most distant.

6.2 Comparison with Grouw

In the graph in Figure 10 the distances compared to Grouw are visualized. Just as in Section 6.1, we used the averages of the Clay Frisian, Wood Frisian, South Corner and Town Frisian groups.

According to Figures 6 and 7 Grouw belongs to the Clay Frisian group of van der Woude. Therefore it is not surprising that we found Clay Frisian closest to Grouw in the graph (6.7%). However, when excluding Grouw from the Clay Frisian group, its average distance with respect to Grouw is still smallest (7.3%). In the graph Wood Frisian appears to be closer to Grouw than South Corner (8.6% versus 11.1%).

Next we find a very sharp break in the graph. The 'common' Frisian varieties are followed by the 'archaic' Frisian varieties: West-Terschelling (22.6%), Oosterend (23.2%), Hindeloopen (26.8%) and Schiermonnikoog (28.3%). In the dendrogram in Figure 7 we already found these varieties to be very distinct from the other Frisian varieties.

The island varieties are followed by Town Frisian varieties or varieties strongly related with Town Frisian. Closest is Sint Annaparochie, which represents the dialect of Het Bildt (28.8%), followed by the two Ameland varieties of Nes and Hollum (29.3% and 29.3%). This outcome does not confirm van der Veen (2001, p. 113) who states that the western variety of Ameland resembles Frisian more closely than the eastern one. The Ameland varieties are followed by Town Frisian (29.4%) and Kollum (29.5%). More distant is the Terschelling variety of Midsland (31.2%) and the dialect of Heerenveen (34.4%). In Section 6.1 we found Heerenveen closest to Dutch which explains its distant position compared to Frisian (Grouw).

Next we find the Low Saxon varieties: the Overijssel variety of Tjalleberd (34.5%), the Westerkwartier varieties (34.6%), the Stellingwerf varieties (35,1%) and the Groningen dialect of Zoutkamp (36.7%). Note that the Westerkwartier varieties are closer to Grouw than the Stellingwerf varieties.

Finally Standard Dutch appears to be the most distant variety. Its distance compared to Grouw is 43.4%.

6.3 Comparison with Zoutkamp

In Section 5.3.2 we explained that a Frisian dialect was originally spoken in the province of Groningen. Later on the province was saxonized by settlers who came from the east where Low German was spoken (Hoekstra, 2001, p. 139). The Saxons mixed with the original inhabitants. Therefore we want to investigate the relation between the present Groningen dialects and the Frisian varieties. The graph in Figure 11 shows the distances of Frisian, Town Frisian and Low Saxon varieties compared to Zoutkamp. Just as in Section

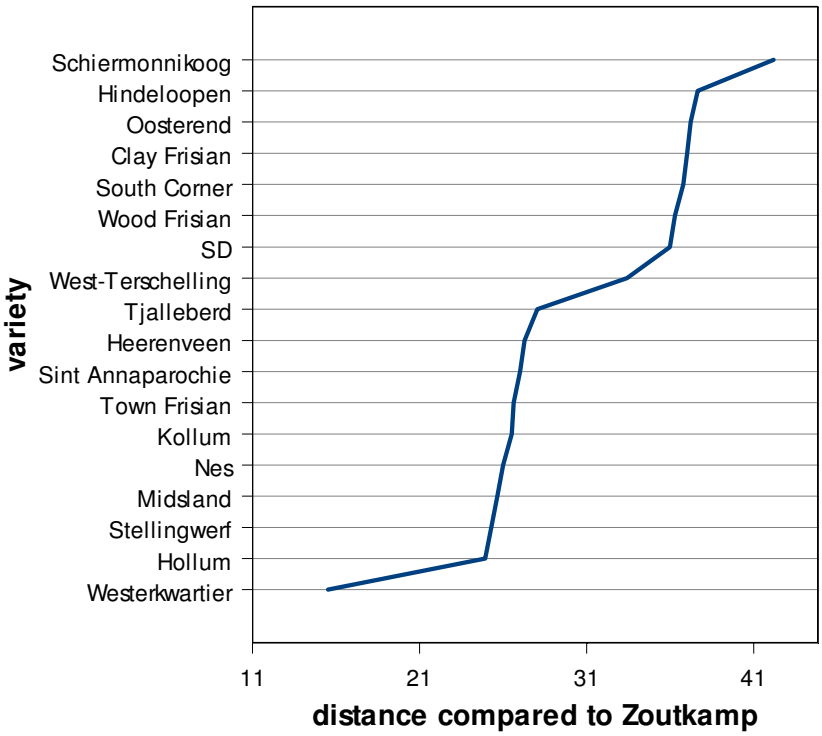

Figure 11. Distances compared to Zoutkamp. The scale distance shows percentages.

6.1 and 6.2, we used the averaged distances of the Clay Frisian, Wood Frisian, South Corner and Town Frisian groups.

Closest are the Westerkwartier varieties (15.5%). This is not surprising, since both the dialect of Zoutkamp and the Westerkwartier varieties are spoken in the province of Groningen. All other varieties are found in the province of Friesland.

After a substantial leap, we find the other Low Saxon varieties, Town Frisian varieties and varieties closely related to Town Frisian. Some of the closest varieties are the Low Saxon Stellingwerf varieties. Most distant is the Overijssel variety of Tjalleberd. In Section 6.1 we found Tjalleberd most distant to Standard Dutch among all Low Saxon varieties, and in Section 6.2 we found Tjalleberd closest to Frisian (Grouw) among all Low Saxon varieties. For the rest, we did not find a meaningful order.

Next we find the Frisian varieties, and Dutch among them. Closest is the dialect of West-Terschelling (33.4%), but Oosterend, which is found on the same island, is much more distant (37.2%). Wood Frisian is closer to

Zoutkamp than both South Corner and Clay Frisian (36.3% versus 36.8% and 37.0%). This is exactly what might be expected on the basis of geography.

Finally we get a second leap and find the archaic varieties of Hindeloopen and Schiermonnikoog most distant (37.7% and 42.2%). In Section 6.1 we found them also most distant to Dutch, and in Section 6.2 we found them most distant to Grouw among the Frisian varieties. These outcomes suggest the independent position of these dialects.

In this section we found that the Groningen dialect is obviously closer to Town Frisian, Town Frisian related varieties and to Dutch than to 'pure' Frisian. Heeringa (2004, p. 227, pp. 272-273) found a sharp boundary between Frisian and Groningen. His results are also based on Levenshtein distance measurements, i.e. based on lexical, phonological and some morphological differences (see Section 3). However, Hoekstra (1998) shows that traces of Frisian can still be found in the Groningen dialect, especially on the syntactic and morphological level. Phonological similarities can hardly be found. Hoekstra explains this by the theory of language contact of Van Coetsem (1988). Someone who does not have a full command of a foreign language will revert to the grammar of his native language. The novel phonology will be borrowed first. Therefore, when the Frisian people in Groningen switched from Frisian to Low Saxon, they adopted the Low Saxon phonology, but held on to Frisian syntactic and morphological elements. Hoekstra gives many syntactic and morphological examples, but we are not sure to what extent a generalization may be made. Further research on the syntactic and morphological level may be interesting.

6.4 Frisian mixed varieties: Frisian or Dutch?

De Haan (2001, p. 33-34) mentioned the classical problem of West Lauwers Frisian linguistics, concerning the genesis of the dialects of the Frisian cities of Leeuwarden, Franeker, Dokkum, Harlingen, Bolsward, Staveren and Sneek in the 16th and following centuries. Is Town Frisian a Frisian or Dutch language variety, or is it a mixed variety, i.e. a language in its own right? (See also Van Bree, 2001, p. 131.) De Haan (2001) mentioned that this discussion started before World War II, and was recently resumed in the eighties and nineties of the twentieth century. Kloeke characterized Town Frisian as Dutch in Frisian mouth (Kloeke, 1927, p. 81), while Gosses (1929, p. 270) characterized Town Frisian as Frisian intermixed with a little Dutch. According to Gorter (2001) Town Frisian is basically "a Dutch dia-

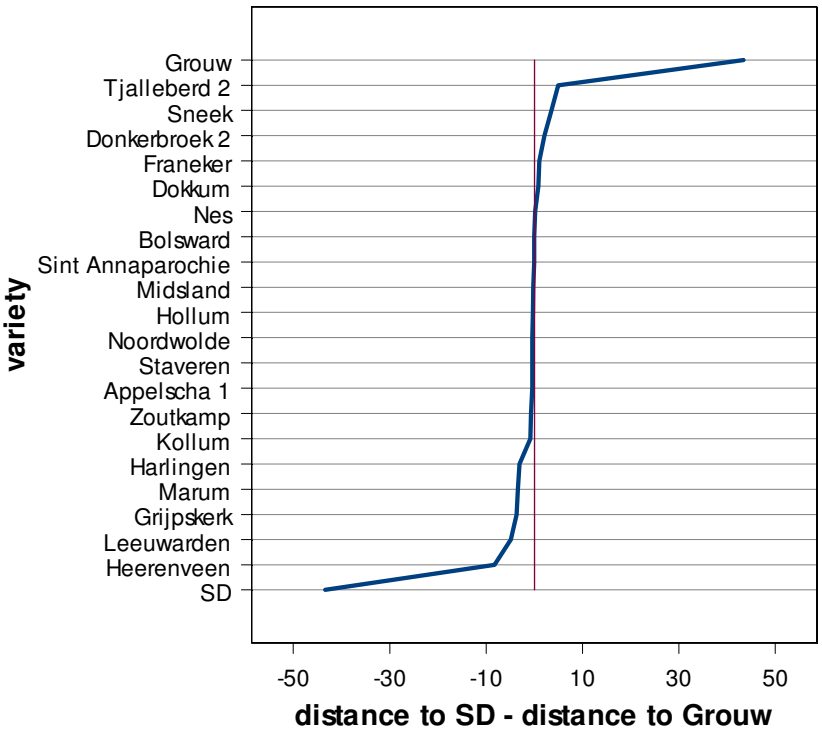

Figure 12. For each dialect the distance compared to Grouw is subtracted from the distance compared to Standard Dutch (SD). The scale distance shows percentages. The vertical line in the middle of the graph represents the case that the distance to Dutch is equal to the distance to Grouw.

lect, although it has been strongly influenced by West Frisian, especially in its lexicon and pronunciation".

In this section we will address the classical question of whether Town Frisian dialects are Frisian, Dutch or mixed. While attempting to answer this question, we will also consider varieties related to Town Frisian, and Low Saxon varieties which are more or less related to Frisian. We calculated distances compared to Standard Dutch, and distances compared to Grouw. Next we subtracted the distances compared to Grouw from the distances compared to Dutch. This resulted in positive and negative values. A positive value means that the variety is closer to Dutch, and a negative value means that the variety is closer to Frisian (i.e. to Grouw). The results are shown in the graph of Figure 12. In the graph we find most of the distance differ-

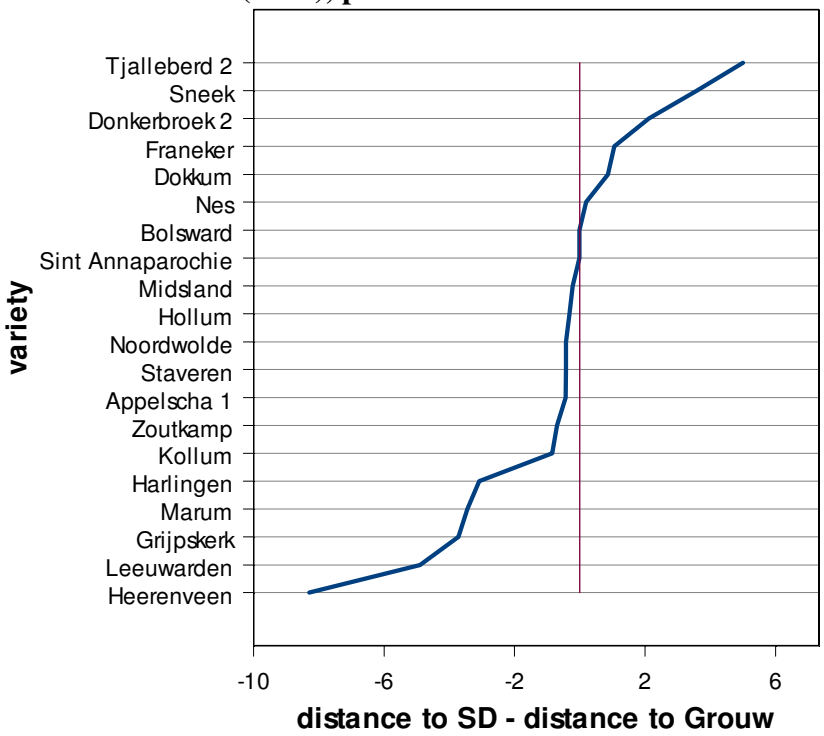

Figure 13. For each dialect the distance compared to Grouw is subtracted from the distance compared to Standard Dutch (SD). Dutch and Grouw are left out in this graph, to get a more sensitive view. The scale distance shows percentages. The vertical line in the middle of the graph represents the case that the distance to Dutch is equal to the distance to Grouw.

ences close to zero. This means that the varieties are not clearly Frisian or Dutch. We are attracted to the third possibility mentioned by De Haan (2001) who regards Town Frisian as "a mixed variety, i.e. a language in its own right." (p. 33-34) When leaving out Dutch and Grouw, the distance differences can be examined in more detail. This can be seen in the graph in Figure 13. We find the Town Frisian varieties of Sneek, Franeker and Dokkum closer to Grouw, while the varieties of Staveren, Harlingen and Leeuwarden are closer to Dutch. Bolsward is equidistant from Dutch and Grouw (29.6%). These findings confirm our conclusion that Town Frisian should be considered as a mixed variety.

Looking at the Town Frisian related varieties, the Ameland variety of Nes is a little bit closer to Grouw, and the Ameland variety of Hollum is just closer to Dutch. The Terschelling variety of Midsland and the dialect of

Kollum are also closer to Dutch than to Grouw. The dialect of Heerenveen is obviously closer to Dutch. The dialect of Sint Annaparochie (representing Het Bildt) has the same distance from Dutch as from Grouw (28.8%). Just as Town Frisian it should be considered as a mixed variety.

When examining the Low Saxon varieties, we find the varieties of Tjalleberd and Donkerbroek closer to Grouw than to Dutch. Both varieties are dialect islands in the Frisian dialect continuum and in both places the Low Saxon variety shares its position with a Frisian dialect. The varieties of Noordwolde, Appelscha, Zoutkamp, Marum and Grijpskerk belong to the Low Saxon dialect continuum. We find these varieties closer to Dutch.

We should be aware of the fact that the graph show distance *differences*, distances compared to Frisian may still be large. For example, the dialect of the Groningen dialect of Zoutkamp is 0.7% closer to Dutch than to Grouw, but its distance compared to Dutch is still large: 36.0% (see also Figure 9).

7. Conclusions

The authoritative dialect maps in Frisian dialectology are Hof's map (1933) and van der Veen's map (2001). The two maps are based on isoglosses. However, with isoglosses degrees of difference cannot be expressed. Furthermore, only phonological isoglosses are used. With Levenshtein distance gradual word distances are calculated. Lexical, phonetic and morphologic differences all contributed to the distance. We applied this method to transcriptions of 54 varieties in and close to the province of Friesland, recorded in 1950-1951. Despite the influence of transcriber differences, we were able to find meaningful results.

As the main division we found a Frisian group, a Town Frisian group and a Low Saxon group. Within the Frisian group, the dialects of Hindeloopen, Schiermonnikoog, West-Terschelling and Oosterend were very distinct. This accords with the descriptions in scholarly literature. Apart from these varieties, we found a Clay Frisian group, a Wood Frisian group and a South Corner group, just as in Hof's map and van der Veen's map. Sometimes Frisian linguists distinguish a fourth group: North Clay Frisian. We concluded that this group is a part of Clay Frisian. Our results accords better with van der Veen's map than with Hof's map.

In the Town Frisian group we found also the dialects of Het Bildt, Ameland, Kollum and Heerenveen. Despite their different history they form one group. In the Low Saxon group we found a group of Groningen dialects

including the Westerkwartier varieties, and a group containing the Stelllingwerf varieties and the Overijssel dialect of Tjalleberd.

We compared the varieties with Standard Dutch, Frisian (Grouw) and Groningen dialect (Zoutkamp). The Town Frisian (related) varieties were closest to Dutch, followed by the Low Saxon varieties. The Frisian varieties were most distant. South Corner was closest, followed by Wood Frisian and Clay Frisian. This finding is in accordance with geography, since Dutch is close to the North Holland dialects. More distant were the archaic varieties of Hindeloopen and Schiermonnikoog, where Schiermonnikoog was most distant. This is in accordance with Winkler (1874, I, p. 455). Of all Dutch dialects he judged Schiermonnikoog to be most distant to the written Dutch language.

When comparing to Grouw, we found more or less the opposite order. Clay Frisian was closest, followed by Wood Frisian and South Corner. After a sharp break we found the archaic varieties of West-Terschelling, Oosterend, Hindeloopen and Schiermonnikoog closest. Next the Town Frisian or Town Frisian related varieties follow. Still more distant are the Low Saxon varieties, where the Groningen dialect of Zoutkamp is most distant. Most distant of all varieties was Standard Dutch.

When ranking the varieties with respect to the Groningen dialect of Zoutkamp, we found the Westerkwartier varieties closest, followed by the Stellingwerf varieties and the Town Frisian (related) varieties. Next the Low Saxon dialect of Tjalleberd was closest, followed by Standard Dutch, Wood Frisian, South Corner and Clay Frisian. Most distant were Hindeloopen and Schiermonnikoog.

We also offered an answer to the classical question whether Town Frisian is a Frisian or Dutch dialect. Some Town Frisian varieties were a little bit closer to Dutch, other were a little bit closer to Frisian. Since distances compared to Dutch are about the same as distances compared to Frisian, we find a possibility mentioned by De Haan (2001) congenial, namely that Town Frisian is “a mixed variety, i.e. a language in its own right.” (p. 33-34)

In this article we showed that with the use of Levenshtein distance many dialectological questions, found in literature, can be answered. We found that transcriber differences may distort results. However, we found a way to eliminate the influence of transcriber differences. As future work it may be useful to investigate the Frisian area and its surrounding on the basis of the *Goeman-Taeldeman-Van Reenen-project* (GTRP) data. Although in this

data the influence of transcriber differences possibly can still be found, the data is known to be much more consistent than the RND data.

Acknowledgements

I thank Peter Kleiweg for letting me use the programs which he developed for the visualization of the maps and dendrograms in the present article, and for his help for some maps. I would like to thank John Nerbonne, Pieter Breuker and some anonymous reviewers for valuable comments. I thank John Nerbonne for reviewing the English. Remaining errors are for my own responsibility. This research was carried out within the framework of the The Determinants of Dialectal Variation *project, which is supported by a grant (360-70-121) from the Netherlands Organisation of Scientific Research (NWO).*

Rijksuniversiteit Groningen
Afdeling Alfa-Informatica

REFERENCES

Blancquaert, E. (1939). *Tekstboekje.* De Sikkel, Antwerpen, 2nd edition. Nederlandse Fonoplaten van Blancquaert en van de Plaetse, Eerste Reeks.

Blancquaert, E. & W. Pée, editors (1925-1982). *Reeks Nederlandse Dialect-atlassen.* De Sikkel, Antwerpen.

Bolognesi, R. & W. Heeringa (2002). De invloed van dominante talen op het lexicon en de fonologie van Sardische dialecten. In: D. Bakker, T. Sanders, R. Schoonen and Per van der Wijst (eds.). In: *Gramma/TTT: tijdschrift voor taalwetenschap*, 9 (1), 45-84. Available at: http://www.let.rug.nl/~heeringa/dialectology/ papers/.

Breuker, P. (1973). *J.H. Halbertsma, Gysbert Japicx en de Fryske dialekten; in pear opmerkingen oer Halbertsma syn filologyske wurksumheden tusken 1822-1827*. Rige Fryske Dialektstúdzjes 3, Fryske Akademy 447, Ljouwert/Leeuwarden, Fryske Akademy.

Breuker, P. (2001). West Frisian in Language Contact. In: H.H. Munske (ed.). *Handbuch des Friesischen/Handbook of Frisian Studies*. Niemeyer, Tübingen, 121-129.

Chambers, J.K. & P. Trudgill (1998). *Dialectology.* 2nd edition. Cambridge University Press, Cambridge.

Daan, J. (1969). *Van Randstad tot Landrand; toelichting bij de kaart: Dialecten en Naamkunde*, volume XXXVII of *Bijdragen en mededelingen der Dialectencommissie van de Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam*. Noord-Hollandse Uitgevers Maatschappij, Amsterdam.

De Haan, G.J. (2001). Recent Trends in Frisian Linguistics. In: H.H. Munske (ed.). *Handbuch des Friesischen/Handbook of Frisian Studies*. Niemeyer, Tübingen, 32-47.

Fort, M.C. (2001). Das Saterfriesische. In: H.H. Munske (ed.). *Handbuch des Friesischen/Handbook of Frisian Studies*. Niemeyer, Tübingen, 409-422.

Gooskens, Ch. & W. Heeringa (2004a). The position of Frisian in the Germanic language area. In: Gilbert, D. & Schreuder, M. & Knevel, N. (eds.), *On the boundaries of phonology and phonetics*, 61-87. Klankleergroep, Faculty of Arts, University of Groningen, Groningen. Dedicated to Tjeerd de Graaf. Available at: http://www.let.rug.nl/~heeringa/dialecto logy/papers/.

Gooskens, Ch. & W. Heeringa (2004b). Perceptive Evaluation of Levenshtein Dialect Distance Measurements using Norwegian Dialect Data. In: *Language Variation and Change*, 16, 189-207.

Goossens, J. (1977). *Inleiding tot de Nederlandse Dialectologie*. Wolters-Noordhoff, Groningen.

Gorter, D. (2001). Extent and Position of West Frisian. In: H.H. Munske (ed.). *Handbuch des Friesischen/Handbook of Frisian Studies*. Niemeyer, Tübingen, 73-83.

Gosses, G. (1929), It Stêdfrysk. In: *It Heitelân* 11, 270-274, 303-308.

Halbertsma, J.H. (1827). *Hulde aan Gysbert Japiks bewezen, in de Sint Martinikerk te Bolsward, op den 7 Julij 1823: met de daarbij behoorende stukken / ten voordeele der Commissie van Weldadigheid, uitgegeven door de commissie, met de oprigting van een gedenkstuk voor Gysbert Japiks belast geweest II*, Bolsward, Hessing etc., Leeuwarden, H.C. Schetsberg.

Halbertsma, J.H. (1867). Over de uitspraak van het Land-friesch. In: *De Taalgids; tijdschrift tot uitbreiding van de kennis der Nederlandsche Taal*, 9.

Heeringa, W. (2001), De selectie en digitalisatie van dialecten en woorden uit de Reeks Nederlandse Dialectatlassen. In: *TABU: Bulletin voor taalwetenschap*, 31 (1/2), 61-103.

Heeringa, W. (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance.* Doctoral dissertation. University of Groningen, Groningen. Available at: http://www.let.rug.nl/~heeringa/dialectology/thesis/.

Hoekstra, E. (1998). Oer de oerienkomst tusken de dialekten fan Grinslân en it Frysk. In: *Philologia Frisica Anno 1996, Lêzingen en neipetearen fan it fjirtjinde Frysk Filologekongres, 23, 24 en 25 oktober 1996*. Fryske Akademy, Ljouwert, 117-137.

Hoekstra, E. (2001). Frisian Relics in the Dutch Dialects. In: H.H. Munske (ed.). *Handbuch des Friesischen/Handbook of Frisian Studies*. Niemeyer, Tübingen, 138-142.

Hof, J.J. (1919). Dongeradielster en Skiermûntseager Frysk. In: *Swanneblommen. Tiidskrift for Fryske tael- en skiedkinde*, 1 (5), 161-167.

Hof, J.J. (1933). *Friesche Dialectgeographie*. Noord- en Zuid-Nederlandse dialectbibliotheek 3, Nijhoff, 's Gravenhage.

Hoppenbrouwers, C. & G. Hoppenbrouwers (2001). *De indeling van de Nederlandse streektalen. Dialecten van 156 steden en dorpen geklasseerd volgens de FFM*. Koninklijke Van Gorcum, Assen.

Jain, A.K. & R.C. Dubes (1988). *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, New Yersey.

Kessler, B. (1995). Computational dialectology in Irish Gaelic. In: *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*. EACL, Dublin, 60-67.

Kloeke, G.G. (1927). *De Hollandsche expansie in de zestiende en zeventiende eeuw en haar weerspiegeling in de hedendaagsche Nederlandsche dialecten; proeve eener historisch-dialectgeographische synthese.* Noord- en Zuid-Nederlandse dialectbibliotheek 2, Nijhoff, 's-Gravenhage.

Nerbonne, J., W. Heeringa, E. van den Hout, P. van der Kooi, S. Otten & W. van de Vis (1996). Phonetic Distance between Dutch dialects. In: G. Durieux, W. Daelemans, and S. Gillis (eds.). *CLIN VI, Papers from the sixth CLIN meeting*. University of Antwerp, Center for Dutch Language and Speech, Antwerpen, 185-202. Available at: **Error! Hyperlink reference not valid.**.

Nerbonne, J. & W. Heeringa (1997). Measuring dialect distances phonetically. In: J. Coleman (ed.). *Workshop on Computational Phonology*. Madrid, 11-18. Available at: http://www.let.rug.nl/~heeringa/dialectology/ papers/.

Nerbonne, J. & P. Kleiweg (2003). Lexical Distance in LAMSAS. In: J. Nerbonne and W. Kretzschmar (eds.), *Computational Methods in Dialectometry*. Special issue of *Computers and the Humanities*, 37(3), 339-357. Available at: http://www.let.rug.nl/~nerbonne/paper.html.

Nerbonne, J. & C. Siedle (2005). Dialektklassifikation auf der Grundlage aggregierter Ausspracheunterschiede. Submitted to: *Zeitschrift für Dialektologie und Linguistik*. Available at: http://www.let.rug.nl/ ~nerbonne/paper.html.

Niebaum, H. (2001). Der Niedergang des Friesischen zwischen Lauwers und Weser. In: H.H. Munske (ed.). *Handbuch des Friesischen/Handbook of Frisian Studies*. Niemeyer, Tübingen, 430-442.

Nielsen, H.F. (1981). *Old English and the Continental Germanic Languages: a Survey of Morphological and Phonological Interrelations.* Innsbrucker Beiträge zur Sprachwissenschaft; Band 33. Institut für Sprachwissenschaft der Universität Innsbruck, Innsbruck.

Sassen, A. (1953). *Het Drents van Ruinen*. PhD thesis, Rijksuniversiteit Groningen, Van Gorcum, Assen.

Siebs, Th. (1901). Geschichte der friesischen Sprache. In: Hermann, P., *Grundriss der germanischen Philologie I*, Trübner, Straßburg, 1152-1464.

Sokal, R.R. & F.J. Rohlf (1962). The comparison of dendrograms by objective methods. *Taxon*, 11, 33-40.

Van Bree, C. (2001). »Stadtfriesisch« und andere nichtfriesische Dialekte der Provinz Frylân. In: H.H. Munske (ed.). *Handbuch des Friesischen/ Handbook of Frisian Studies.* Niemeyer, Tübingen, 129-138.

Van Coetsem, F. (1988). *Loan phonology and the two transfer types in language contact*. Foris, Dordrecht.

Van der Veen, K.F. (1986). Yndieling en relative ôfstân fan Fryske plattelânsdialekten. In: *Philologia Frisica Anno 1984, Lêzingen en neipetearen fan it tsiende Frysk Filologekongres, oktober 1984*. Fryske Akademy, Ljouwert, 10-60.

Van der Veen, K.F. (1994). Yndielingskaarten fan Fryske plattelânsdialekten. *It Beaken*, 56, 1-23.

Van der Veen, K.F. (2001). West Frisian Dialectology and Dialects. In: H.H. Munske (ed.). *Handbuch des Friesischen/Handbook of Frisian Studies*. Niemeyer, Tübingen, 83-98.

Visser, W. (1997). *The syllable in Frisian*. Holland Academic Graphics, The Hague.

Walker, A.G.H. (2001). Extent and Position of North Frisian. In: H.H. Munske (ed.). *Handbuch des Friesischen/Handbook of Frisian Studies*. Niemeyer, Tübingen, 263-284.

Winkler, J. (1874). *Algemeen Nederduitsch en Friesch Dialecticon*. Martinus Nijhoff, ‘s Gravenhage.

Appendix

Table 1. List of 125 words which were selected from the 141 RND sentences. On the basis of phonetic transcriptions of translations of these words dialects are compared with each other. The fourth column gives the number of the sentence from which the word was usually taken.

	Dutch	**Frisian**	**English**	**RND**
1	mijn	myn	my	2
2	vriend	freon	friend	2
3	werk	wurk	work	4
4	op	op	on	5
5	schip	skip	ship	5
6	kregen	krigen	got	5
7	brood	bôle	bread	5
8	vinger	finger	finger	6
9	vier	fjouwer	four	10
10	bier	bier	beer	10
11	twee	twa	two	11
12	drie	trije	three	12
13	hij	hy	he	13
14	knuppel	kneppel	cudgel	13
15	ik	ik	I	14
16	knie	knibbel	knee	14
17	gezien	sjoen	seen	14
18	kerel	keardel	fellow	21
19	stenen	stiennen	stones	25
20	breder	breder	broader	25
21	duivel	duvel	devil	28
22	gebleven	bleaun	stayed	28
23	meester	master	master	29
24	zee	see	sea	29
25	graag	graach	gladly	31
26	steel	stôk	handle	33
27	bezem	biezem	broom	33
28	geroepen	roppen	called	35
29	peer	par	pear	36
30	rijp	ryp	ripe	36
31	geld	jild	money	38
32	ver	fier	far	39
33	brengen	bringe	bring	39
34	zwemmen	swimme	swim	42
35	bed	bêd	bed	45
36	springen	springe	spring	47
37	vader	heit	father	53

38	zes	seis	six	53
39	jaar	jier	year	53
40	school	skoalle	school	53
41	laten	litte	let	53
42	gaan	gean	go	53
43	potten	potten	jars	56
44	zijn	binne	are	56
45	veel	folle	much	56
46	maart	maart	March	58
47	nog	noch	yet	58
48	koud	kâld	cold	58
49	kaars	kears	candle	59
50	geeft	jout	gives	59
51	licht	ljocht	light	59
52	paard	hynder	horse	60
53	tegen	tsjin	against	63
54	kaas	tsiis	cheese	66
55	dag	dei	day	68
56	avond	jûn	evening	68
57	barst	barst	crack	70
58	brief	brief	letter	71
59	hart	hert	hart	72
60	spannen	spanne	put	74
61	nieuwe	nije	new	74
62	kar	karre	cart	74
63	zoon	soan	son	76
64	koning	koaning	king	76
65	ook	ek	also	76
66	geweest	west	been	76
67	lange	lange	long	78
68	woord	wurd	word	79
69	kindje	berntsje	baby	80
70	was	wie	was	80
71	dochtertje	dochterke	little daughter	82
72	bos	bosk	wood	82
73	ladder	ljedder	ladder	83
74	mond	mûle	mouth	86
75	droog	droech	dry	86
76	dorst	toarst	thirst	86
77	weg	wei	way	87
78	krom	krom	curved	87
79	liedje	lietsje	ditty	90
80	goed	goed	good	92
81	kelder	kelder	cellar	95
82	voor	foar	for	95
83	moest	moast	must	96
84	drinken	drinke	drink	96

85	broer	broer	brother	98
86	moe	wurch	tired	98
87	dun	tin	thin	100
88	zuur	soer	sour	100
89	put	put	well	101
90	uur	oere	houre	101
91	vuur	fjoer	fire	104
92	duwen	triuwe	push	105
93	hebben	ha	have	106
94	stuk	stik	piece	106
95	brug	brêge	bridge	106
96	veulen	fôle	foal	107
97	komen	komme	come	107
98	deur	doar	door	109
99	gras	gers	grass	111
100	bakken	bakke	bake	113
101	je	do	you	116
102	eieren	aaien	eggs	116
103	krijgen	krije	get	116
104	waren	wiene	were	119
105	vijf	fiif	five	119
106	hooi	hea	hay	122
107	is	is	is	122
108	groen	grien	green	122
109	boompje	beamke	little tree	124
110	wijn	wyn	wine	125
111	huis	hûs	house	126
112	melk	molke	milk	127
113	spuit	spuitet	spouts	127
114	koe	ko	cow	127
115	koster	koster	sexton	128
116	buigen	bûge	bend	129
117	blauw	blau	blue	131
118	geslagen	slein	struck	131
119	saus	sjeu	sauce	132
120	flauw	flau	flat	132
121	sneeuw	snie	snow	133
122	doen	dwaan	do	136
123	dopen	dope	baptize	137
124	dorsen	terskje	thresh	138
125	binden	bine	bind	139