# The Position of Frisian in the Germanic Language Area

*Charlotte Gooskens and Wilbert Heeringa*

## 1. Introduction

Among the Germanic varieties the Frisian varieties in the Dutch province of Friesland have their own position. The Frisians are proud of their language and more than 350,000 inhabitants of the province of Friesland speak Frisian every day. Heeringa (2004) shows that among the dialects in the Dutch language area the Frisian varieties are most distant with respect to standard Dutch. This may justify the fact that Frisian is recognized as a second official language in the Netherlands. In addition to Frisian, in some towns and on some islands a mixed variety is used which is an intermediate form between Frisian and Dutch. The variety spoken in the Frisian towns is known as Town Frisian[1].

The Frisian language has existed for more than 2000 years. Genetically the Frisian dialects are most closely related to the English language. However, historical events have caused the English and the Frisian language to diverge, while Dutch and Frisian have converged. The linguistic distance to the other Germanic languages has also altered in the course of history due to different degrees of linguistic contact. As a result traditional genetic trees do not give an up-to-date representation of the distance between the modern Germanic languages.

In the present investigation we measured linguistic distances between Frisian and the other Germanic languages in order to get an impression of the effect of genetic relationship and language contact for the position of the modern Frisian language on the Germanic language map. We included six Frisian varieties and one Town Frisian variety in the investigation. Furthermore, eight Germanic standard languages were taken into account. Using this material, we firstly wished to obtain a hierarchical classification of the Germanic varieties. From this classification the position of (Town) Frisian became clear. Secondly, we ranked all varieties with respect to each

of the standard Germanic languages as well as to (Town) Frisian. The rankings showed the position of (Town) Frisian with respect to the standard languages and the position of the standard languages with respect to (Town) Frisian.

In order to obtain a classification of varieties and establish rankings, we needed a tool that can measure linguistic distances between the varieties. Bolognesi and Heeringa (2002) investigated the position of Sardinian dialects with respect to different Romance languages using the Levenshtein distance, an algorithm with which distances between word pronunciations are calculated. In our investigation we used the same methodology.

In Section 2, we will present the traditional ideas about the genetic relationship between the Germanic languages and discuss the relationship between Frisian and the other Germanic languages. At the end of the section we will discuss the expected outcome of the linguistic distance measurements between Frisian and the other Germanic languages. In Section 3 the data sources are described and in Section 4 the method for measuring linguistic distances between the language varieties is presented. The results are presented in Section 5, the discussion of which is presented in Section 6.

## 2. Frisian and the Germanic languages

### 2.1. History and classification of the Germanic languages[2]

The Germanic branch of the Indo-European languages has a large number of speakers, approximately 450 million native speakers, partly due to the colonization of many parts of the world. However, the number of different languages within the Germanic group is rather limited. Depending on the definition of what counts as a language there are about 12 different languages. Traditionally, they are divided into three subgroups: East Germanic (Gothic, which is no longer a living language), North Germanic (Icelandic, Faeroese, Norwegian, Danish, and Swedish), and West Germanic (English, German, Dutch, Afrikaans, Yiddish, and Frisian). Some of these languages are so similar that they are only considered independent languages because of their position as standardized languages spoken within the limits of a state. This goes for the languages of the

Scandinavian countries, Swedish, Danish and Norwegian, which are mutually intelligible. Other languages consist of dialects which are in fact so different that they are no longer mutually intelligible but are still considered one language because of standardization. Northern and southern German dialects are an example of this situation.

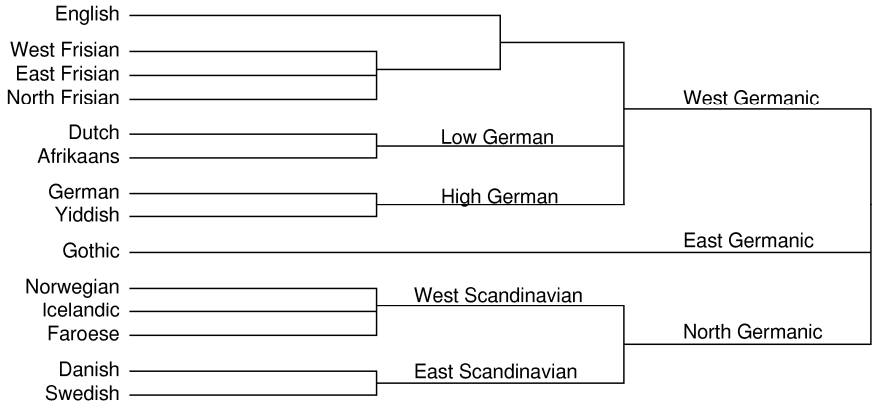


*Figure 1.* The genetic tree of Germanic languages.

In Figure 1, a traditional Germanic genetic tree is shown. We constructed this tree on the basis of data in the literature. The tree gives just a rough division, and linguistic distances should not be derived from this tree. It is commonly assumed that the Germanic languages originate from the southern Scandinavian and the northern German region. After the migration of the Goths to the Balkans towards the end of the pre-Christian era, North-West Germanic remained uniform till the 5th century AD, after which a split between North and West Germanic occurred owing to dialectal variation and the departure of the Anglo-Saxons from the Continent and the colonization of Jutland.

During the Viking Age, speakers of North Germanic settled in a large geographic area, which eventually led to the five modern languages (see above). Of these languages, Icelandic (and to a lesser degree Faeroese), which is based on the language of southwestern Norway where the settlers came from, can be considered the most conservative language (Sandøy, 1994). Of the three mainland Scandinavian languages, Danish has moved farthest away from the common Scandinavian roots due to influences from the south.

The parentage of the West Germanic languages is less clear. Different tribal groups representing different dialect groups spread across the area, which eventually resulted in the modern language situation. Historically Frisian and English both belong to the Ingwaeonic branch of the West Germanic language group. Originally the Frisian speech community extended from the present Danish-German border along the coast to the French-Belgian border in the south. However, expansion from Saxons and Franconians from the east and the south throughout the medieval period resulted in a loss of large Frisian areas and a division into three mutually intelligible varieties: West Frisian (spoken in the northern Dutch province of Friesland by more than 350,000 people), East Frisian or Saterlandic (spoken by a thousand speakers in three villages west of Bremen) and North Frisian (spoken by less than ten thousand people on the islands on the north-western coast of Germany).

The English language came into being as a result of immigrations of tribal Anglo-Saxon groups from the North Sea coast during the fifth and sixth centuries. Whereas other insular Germanic varieties are in general rather conservative, the English insularity lacked this conservatism. English is considered most closely related to Frisian on every linguistic level due to their common ancestorship and to continued language contact over the North Sea.

The German language is spoken in many European countries in a large number of dialects and varieties, which can be divided into Low German and High German. Yiddish, too, can be regarded as a German variety. Dutch is mainly based on the western varieties of the low Franconian area but low Saxon and Frisian elements are also found in this standard language. Scholars disagree about the precise position of Dutch and Low German in the language tree. They can be traced back to a common root often referred to as the Ingwaeonic language group, but are often grouped together with High German as a separate West Germanic group. This grouping with High German might be the best representation of the modern language situation given that the individual dialects spoken in the area in fact form a dialect continuum. Afrikaans, finally, is a contemporary West Germanic language, developed from seventeenth century Dutch as a result of colonization, but with influences from African languages.

## 2.2. The relationship between Frisian and the other Germanic languages.

This short outline of the relationships among the Germanic languages shows that English is the language which is genetically closest to Frisian, and still today English is considered to be most similar to Frisian. For example *The Columbia Encyclopedia* (2001) says: "Of all foreign languages, [Frisian] is most like English". Pei (1966, p. 34) summarizes the situation as follows: "Frisian, a variant of Dutch spoken along the Dutch and German North Sea coast, is the foreign speech that comes closest to modern English, as shown by the rhyme: 'Good butter and good cheese is good English and good Fries'". This rhyme refers to the fact that the words for butter and cheese are almost the same in the two languages. However, in the course of history, contact with other Germanic languages has caused Frisian to converge to these languages. The Frisians have a long history of trade and in early medieval times they were one of the leading trading nations in Europe due to their strategic geographic position close to major trade routes along the rivers and the North Sea. Also, the Vikings and the English were frequent visitors of the Frisian language area. This intensive contact with both English and the North Germanic languages, especially Danish, resulted in linguistic exchanges (see Feitsma, 1963; Miedema, 1966; Wadstein, 1933). Later in history, the Frisian language was especially influenced by the Dutch language (which itself contains many Frisian elements). For a long period, Frisian was stigmatized as a peasant language and due to the weak social position of the Frisian language in the Dutch community it was often suppressed, resulting in a strong Dutch impact on the Frisian language. Nowadays, Dutch as the language of the administration still has a large influence on the media and there has been substantial immigration of Dutch speaking people to Friesland. However, the provincial government has decided to promote Frisian at all levels in the society.

When investigating the position of the Frisian language within the Germanic language group, there are clearly two forces which should be taken into account. On the one hand, Frisian and English are genetically closely related and share sound changes which do not occur in the other Germanic languages. This yields the expectation that the linguistic distance between these two languages is relatively small. On the other hand, the close contact with Dutch makes it plausible that the Dutch and the Frisian languages have converged. Also the distance to Danish might be smaller than expected from the traditional division of Germanic into a North

Germanic and a West Germanic branch at an early stage because of the intensive contacts in the past.

## 3. Data sources

In this section, we will first give a short characterization of the language varieties and the speakers who were recorded for our investigation. Next, we will present the nature of the recordings and the transcriptions which formed the basis for linguistic distance measurements.

### 3.1. Language varieties

Since our main interest was the Frisian language and its linguistic position within the Germanic language group we wished to represent this language as well as possible. For this reason, we included seven Frisian varieties, spread over the Frisian language area. Furthermore, our material contained eight Germanic standard languages. First, we will describe the Frisian varieties and next the standard languages.

As far as the Frisian varieties are concerned, we chose varieties from different parts of the province, both from the coastal area and from the inland. The varieties are spoken in different dialect areas according to the traditional classification (see below) and they represent different stages of conservatism. The precise choice of the seven varieties was determined by speaker availability for recordings in our vicinity and at the Fryske Akademy in Leeuwarden. In Figure 2, the geographical position of the seven Frisian language varieties in the province of Friesland is shown.

Due to the absence of major geographical barriers, the Frisian language area is relatively uniform. The major dialectal distinctions are primarily phonological. Traditionally, three main dialect areas are distinguished (see e.g. Hof, 1933; Visser, 1997): Klaaifrysk (clay Frisian) in the west, Wâldfrysk (forest Frisian) in the east and Súdwesthoeksk (southwest quarter) in the southwest. In our material Klaaifrysk is represented by the dialects of Oosterbierum and Hijum, Wâldfrysk by Wetsens and Westergeest, and Súdwesthoeksk by Tjerkgaast. Hindeloopen is in the area of Súdwesthoeksk. However, this dialect represents a highly conservative area. The phonological distance between Hindeloopen and the main dialects is substantial (van der Veen, 2001). Finally, our material contains

the variety spoken in Leeuwarden (see note 1). This is an example of Town Frisian, which is also spoken in other cities of Friesland. Town Frisian is a Dutch dialect strongly influenced by Frisian but stripped of the most characteristic Frisian elements (Goossens, 1977).

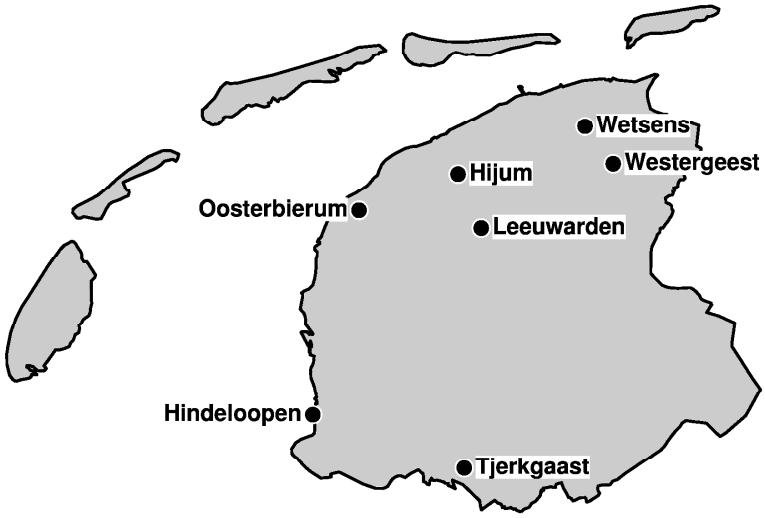


*Figure 2.* The geographical position of the seven Frisian language varieties in the province of Friesland.

In addition to the Frisian dialects, the following eight standard languages were included: Icelandic, Faroese, Norwegian, Swedish, Danish, English, Dutch, and German. We had meant to include all standard Germanic languages in our material. However, due to practical limitations a few smaller languages were not included.

As for Norwegian, there is no official standard variety. The varieties spoken around the capital of Oslo in the southeast, however, are often considered to represent the standard language. We based the present investigation on prior research on Norwegian dialects (see Heeringa and Gooskens, 2003; Gooskens and Heeringa, submitted), and we chose the recording which to Norwegians sounded most standard, namely the Lillehammer recording[3]. It was our aim to select standard speakers from all countries, but it is possible that the speech of some speakers contains slight regional influences. The speakers from Iceland, the Faroe Islands and

Sweden spoke the standard varieties of the capitals. The Danish speaker came from Jutland, the German speaker from Kiel, the English speaker from Birmingham and the Dutch speaker had lived at different places in the Netherlands, including a long period in the West during adolescence.

### 3.2. Phonetic transcriptions

The speakers all read aloud translations of the same text, namely the fable 'The North Wind and the Sun'. This text has often been used for phonetic investigations; see for example *The International Phonetic Association* (1949 and 1999) where the same text has been transcribed in a large number of different languages. A database of Norwegian transcriptions of the same text has been compiled by J. Almberg (see note 3). As mentioned in the previous section, we only used the transcription of Lillehammer from this database. In future, we would like to investigate the relations between Norwegian and other Germanic varieties, using the greater part of the transcriptions in this database. Therefore, our new transcriptions should be as comparable as possible with the existing Norwegian ones. To ensure this, our point of departure was the Norwegian text. This text consists of 91 words (58 different words) which were used to calculate Levenshtein distances (see Section 4). The text was translated word for word from Norwegian into each of the Germanic language varieties. We are aware of the fact that this may result in less natural speech: sentences were often syntactically wrong. However, it guarantees that for each of the 58 words a translation was obtained. The words were not recorded as a word list, but as sentences. Therefore in the new recordings words appear in a similar context as in the Norwegian varieties. This ensures that the influence of assimilation phenomena on the results is as comparable as possible.

Most new recordings were transcribed phonetically by one of the authors. To ensure consistency with the existing Norwegian transcriptions, our new transcriptions were corrected by J. Almberg, the transcriber of the Norwegian recordings. In most cases we incorporated the corrections. The transcription of the Faroese language was completely done by J. Almberg. The transcriptions were made in IPA as well as in X-SAMPA (eXtended Speech Assessment Methods Phonetic Alphabet). This is a machine-readable phonetic alphabet, which is also readable by people. Basically, it maps IPA-symbols to the 7 bit printable ASCII/ANSI characters[4]. The

transcriptions were used to calculate the linguistic distances between varieties (see Section 4).

## 4. Measuring distances between varieties

In 1995 Kessler introduced the use of the Levenshtein distance as 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 on, this approach was applied by Nerbonne, Heeringa, Van den Hout, Van der Kooi, Otten, and Van de Vis (1996) to Dutch dialects. They assumed that distances between all possible pairs of segments are the same. E.g. the distance between an [l] and an [e] is the same as the distance between the [l] and [ə]. Both Kessler (1995) and Nerbonne and Heeringa (1997) also experimented with more refined versions of the Levenshtein algorithm in which gradual segment distances were used which were found on the basis of the feature systems of Hoppenbrouwers (1988) and Vieregge et. al. (1984).

In this paper we use an implementation of the Levenshtein distance in which sound distances are used which are found by comparing spectrograms. In Section 4.1 we account for the use of spectral distances and explain how we calculate them. Comparisons are made on the basis of the audiotape *The Sounds of the International Phonetic Alphabet* (Wells and House, 1995). In Section 4.2 we describe the Levenshtein distance and explain how spectral distances can be used in this algorithm.

### 4.1. Gradual segment distances

When acquiring language, children learn to pronounce sounds by listening to the pronunciation of their parents or other people. The acoustic signal seems to be sufficient to find the articulation which is needed to realize the sound. Acoustically, speech is just a series of changes in air pressure, quickly following each other. A spectrogram is a "graph with frequency on the vertical axis and time on the horizontal axis, with the darkness of the graph at any point representing the intensity of the sound" (Trask, 1996, p. 328).

In this section we present the use of spectrograms for finding segment distances. Segment distances can also be found on the basis of phonological

or phonetic feature systems. However, we prefer the use of acoustic representations since they are based on physical measurements. In Potter, Kopp and Green's (1947) *Visible Speech*, spectrograms are shown for all common English sounds (see pp. 54-56). Looking at the spectrograms we already see which sounds are similar and which are not. We assume that visible (dis)similarity between spectrograms reflects perceptual (dis)similarity between segments to some extent. In Figure 3 the spectrograms of some sounds are shown as pronounced by John Wells on the audiotape *The Sounds of the International Phonetic Alphabet* (Wells and House, 1995). The spectrograms are made with the computer program PRAAT[5].

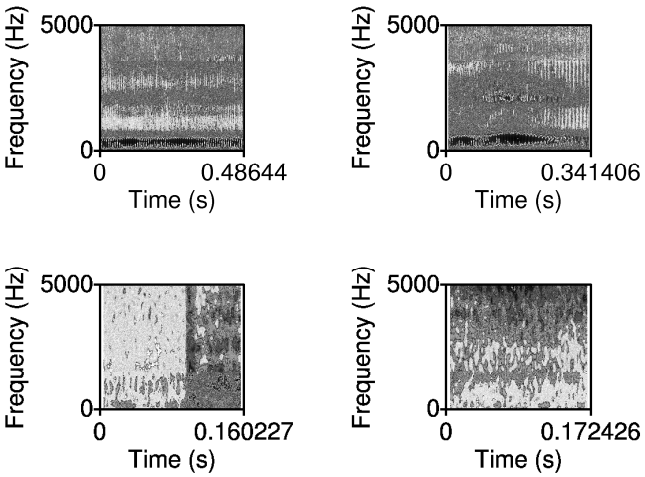


*Figure 3.* Spectrograms of some sounds pronounced by John Wells. Upper the [i] (left) and the [e] (right) are shown, and lower the [p] (left) and the [s] (right) are visualized.

### *4.1.1. Samples*

For finding spectrogram distances between all IPA segments we need samples of one or more speakers for each of them. We found the samples on the tape *The Sounds of the International Phonetic Alphabet* on which all

IPA sounds are pronounced by John Wells and Jill House. On the tape the vowels are pronounced in isolation. The consonants are sometimes preceded, and always followed by an [a]. We cut out the part preceding the [a], or the part between the [a]'s. We realize that the pronunciation of sounds depends on their context. Since we use samples of vowels pronounced in isolation and samples of consonants selected from a limited context, our approach is a simplification of reality. However, Stevens (1998, p. 557) observes that

> "by limiting the context, it was possible to specify rather precisely the articulatory aspects of the utterances and to develop models for estimating the acoustic patterns from the articulation".

The burst in a plosive of the IPA inventory is always preceded by a period of silence (voiceless plosives) or a period of murmur (voiced plosives). When a voiceless plosive is not preceded by an [a], it is not clear how long the period of silence which really belongs to the sounds lasts. Therefore we always cut out each plosive in such a way that the time span from the beginning to the middle of the burst is equal to 90 ms. Among the plosives which were preceded by an [a] or which are voiced (so that the real time of the start-up phase can be found) we found no sounds with a period of silence or murmur which was clearly shorter than 90 ms.

In voiceless plosives, the burst is followed by an [h]-like sound before the following vowel starts. A consequence of including this part in the samples is that bursts often do not match when comparing two voiceless plosives. However, since aspiration is a characteristic property of voiceless sounds, we retained aspiration in the samples. In general, when comparing two voiced plosives, the bursts match. When comparing a voiceless plosive and a voiced plosive, the bursts do not match.

To keep trills comparable to each other, we always cut three periods, even when the original samples contained more periods. When there were more periods, the most regular looking sequence of three periods was cut.

The Levenshtein algorithm also requires a definition of 'silence'. To get a sample of 'silence' we cut a small silent part on the IPA tape. This assures that silence has approximately the same background noise as the other sounds.

To make the samples as comparable as possible, all vowel and extracted consonant samples are monotonized on the mean pitch of the 28 concatenated vowels. The mean pitch of John Wells was 128 Hertz; the

mean pitch of Jill House was 192 Hertz. In order to monotonize the samples the pitch contours were changed to flat lines. The volume was not normalized because volume contains too much segment specific information. For example it is specific for the [v] that its volume is greater than that of the [f].

### *4.1.2. Acoustic representation*

In the most common type of spectrogram the linear Hertz frequency scale is used. The difference between 100 Hz and 200 Hz is the same as the difference between 1000 Hz and 1100 Hz. However, our perception of frequency is non-linear. We hear the difference between 100 Hz and 200 Hz as an octave interval, but also the difference between 1000 Hz and 2000 Hz is perceived as an octave. Our ear evaluates frequency differences not absolutely, but relatively, namely in a logarithmic manner. Therefore, in the Barkfilter, the Bark-scale is used which is roughly linear below 1000 Hz and roughly logarithmic above 1000 Hz (Zwicker and Feldtkeller, 1967).

In the commonly used type of spectrogram the power spectral density is represented per frequency per time. The power spectral density is the power per unit of frequency as a function of the frequency. In the Barkfilter the power spectral density is expressed in decibels (dB's). "The decibel scale is a way of expressing sound amplitude that is better correlated with perceived loudness" (Johnson, 1997, p. 53). The decibel scale is a logarithmic scale. Multiplying the sound pressure ten times corresponds to an increase of 20 dB. On a decibel scale intensities are expressed relative to the auditory threshold. The auditory threshold of 0.00002 Pa corresponds with 0 dB (Rietveld and Van Heuven, 1997, p. 199).

A Barkfilter is created from a sound by band filtering in the frequency domain with a bank of filters. In PRAAT the lowest band has a central frequency of 1 Bark per default, and each band has a width of 1 Bark. There are 24 bands, corresponding to the first 24 critical bands of hearing as found along the basilar membrane (Zwicker and Fastl, 1990). A critical band is an area within which two tones influence each other's perceptibility (Rietveld and Van Heuven, 1997). Due to the Bark-scale the higher bands summarize a wider frequency range than the lower bands.

In PRAAT we used the default settings when using the Barkfilter. The sound signal is probed each 0.005 seconds with an analysis window of 0.015 seconds. Other settings may give different results, but since it was

not a priori obvious which results are optimal, we restricted ourselves to the default settings. In Figure 4 Barkfilters for some segments are shown.

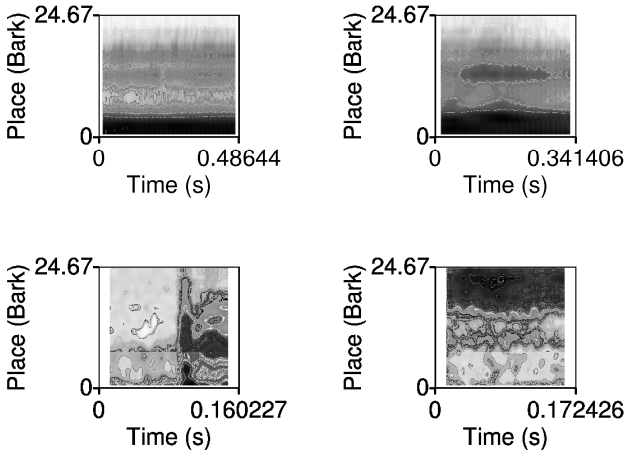


*Figure 4.* Barkfilter spectrograms of some sounds pronounced by John Wells. Upper the [i] (left) and the [e] (right) are shown, and lower the [p] (left) and the [s] (right) are visualized.

### *4.1.3. Comparison*

In this section, we explain the comparison of segments in order to get distances between segments that will be used in the Levenshtein distance measure. In a Barkfilter, the intensities of frequencies are given for a range of times. A spectrum contains the intensities of frequencies at one time. The smaller the time step, the more spectra there are in the acoustic representation. We consistently used the same time step for all samples.

It appears that the duration of the segment samples varies. This may be explained by variation in speech rate. Duration is also a sound-specific property. E.g., a plosive is shorter than a vowel. The result is that the number of spectra per segment may vary, although for each segment the same time step was used. Since we want to normalize the speech rate and regard segments as linguistic units, we made sure that two segments get the same number of spectra when they are compared to each other.

When comparing one segment of *m* spectra with another segment of *n* spectra, each of the *m* elements is duplicated *n* times, and each of the *n* elements is duplicated *m* times. So both segments get a length of $m \times n$.

In order to find the distance between two sounds, the Euclidean distance is calculated between each pair of corresponding spectra, one from each of the sounds. Assume a spectrum $e1$ and $e2$ with *n* frequencies, then the Euclidean distance is:

*Equation 1.* Euclidean distance

$$d(e1, e2) = \sqrt{\sum_{i=1}^{n}(e1_i - e2_i)^2}$$

The distance between two segments is equal to the sum of the spectrum distances divided by the number of spectra. In this way we found that the greatest distance occurs between the [a] and 'silence'. We regard this maximum distance as 100%. Other segment distances are divided by this maximum and multiplied by 100. This yields segment distances expressed in percentages. Word distances and distances between varieties which are based on them may also be given in terms of percentages.

In perception, small differences in pronunciation may play a relatively strong role in comparison with 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*).*

### *4.1.4. Suprasegmentals and diacritics*

The sounds on the tape *The Sounds of the International Phonetic Alphabet* are pronounced without suprasegmentals and diacritics. However, a restricted set of suprasegmentals and diacritics can be processed in our system.

Length marks and syllabification are processed by changing the transcription beforehand. In the X-SAMPA transcription, extra-short

segments are kept unchanged, sounds with no length indication are doubled, half long sounds are trebled, and long sounds are quadrupled. Syllabic sounds are treated as long sounds, so they are quadrupled.

When processing the diacritics *voiceless* and/or *voiced*, we assume that a voiced voiceless segment (e.g. [t̬]) and a voiceless voiced segment (e.g. [d̥]) are intermediate pronunciations of a voiceless segment ([t]) and a voiced segment ([d]). Therefore we calculate the distance between a segment *x* and a voiced segment *y* as the average of the distance between *x* and *y* and the distance between *x* and the voiced counterpart of *y*. Similarly, the distance between a segment *x* and a voiceless segment *y* is calculated as the mean of the distance between *x* and *y* and the distance between *x* and the voiceless counterpart of *y*. For voiced sounds which have no voiceless counterpart (the sonorants), or for voiceless sounds which have no voiced counterpart (the glottal stop) the sound itself is used.

The diacritic *apical* is only processed for the [s] and the [z]. We calculate the distance between [s˚] and e.g. [f] as the average of the distance between [s] and [f] and [̪] and [f]. Similarly, the distance between [z˚] and e.g. [v] is calculated as the mean of [z] and [v] and [̪] and [v].

The thought behind the way in which the diacritic *nasal* is processed is that a nasal sound is more or less intermediate between its non-nasal version and the [n]. We calculate the distance between a segment *x* and a nasal segment *y* as the average of the distance between *x* and *y* and the distance between *x* and [n].

## 4.2. Levenshtein distance

Using the Levenshtein distance, two dialects are compared by comparing the pronunciation of a word in the first dialect with the pronunciation of the same word 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 *afternoon* is pronounced as [ɐ ɑ̃«⅓t«ɛɪn¬ʃn] in the dialect of Savannah, Georgia, and as [ɛɪɑ̃⅓₇«₅ɐ ₁₈ʃ ₁] in the dialect of Lancaster, Pennsylvania[6]. Changing one pronunciation into the other can be done as in table 1 (ignoring suprasegmentals and diacritics for this moment)[7]:

*Table 1.* Changing one pronunciation into another using a minimal set of operations.

| | | |
|---|---|---|
| ɑ«⅓t«n¬n | delete « | 1 |
| ɑ⅓t«n¬n | insert r | 1 |
| ɑ⅓t«rn¬n | subst. ¬/ɐ | 1 |
| ɑ⅓ɜ«ɔɪɐɪ | | |
| | | 3 |

In fact many sequence operations map [ɡ ɑ«⅓t«ɛɪn¬ʃn] to [ɛɪɑ⅓ɜ«ɔɡ ɪɐʃ ɪ]. 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 this does not accord with the idea that words are linguistic units, 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:

*Table 2.* Alignment which gives the minimal cost. The alignment corresponds with table 1.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ɑ | « | ⅓ | t | « | | n | ¬ | n |
| ɑ | | ⅓ | ɜ | « | ɔ | ɪ | ɐ | ɪ |
| | 1 | | | | 1 | | 1 | |

The total cost of 3 (1+1+1) is now divided by the length of 9. This gives a word distance of 0.33 or 33%.

In Section 3.1.3 we explained how distances between segments can be found using spectrograms. This makes it possible to refine our Levenshtein algorithm by using the spectrogram distances as operation weights. Now the cost of insertions, deletions and substitutions is not always equal to 1, but varies, i.e., it is equal to the spectrogram distance between the segment

and 'silence' (insertions and deletions) or between two segments (substitution).

To reckon with syllabification in words, the Levenshtein algorithm is adapted so that only a vowel may match with a vowel, a consonant with a consonant, the [j] or [w] with a vowel (or opposite), the [i] or [u] with a consonant (or opposite), and a central vowel (in our research only the schwa) with a sonorant (or opposite). In this way unlikely matches (e.g. a [p] with a [a]) are prevented.

In our research we used 58 different words. When a word occurred in the text more than once, the mean over the different pronunciations was used. So when comparing two dialects we get 58 Levenshtein distances. Now the dialect distance is equal to the sum of 58 Levenshtein distances divided by 58. 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 15 language varieties are arranged in a $15 \times 15$ matrix.

## 5. Results

The results of the Levenshtein distance measurements are analyzed in two ways. First, on the basis of the distance matrix we applied hierarchical cluster analysis (see Section 5.1). 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 alternatives 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 ($r$=0.9832), see Sokal and Rohlf (1962).

Second, we ranked all varieties in order of relationship with the standard languages, Frisian and Town Frisian (see Section 5.2). When ranking with relation to Frisian, we looked at the average over all Frisian dialects. Since the ratings with respect to each of the Frisian varieties individually were very similar averaging was justified.

5.1. The classification of the Germanic languages

Looking at the clusters of language varieties in Figure 5 we note that our results reflect the traditional classification of the Germanic languages to a large extent (see Figure 1). On the highest level there is a division between English and the other Germanic languages. When we examine the group of other Germanic languages, we find a clear division between the North Germanic languages and the West Germanic languages. Within the North Germanic group, we see a clear division between the Scandinavian languages (Danish, Norwegian and Swedish) on the one hand and the Faroese and Icelandic on the other hand. In the genetic tree (see Figure 1), Norwegian is clustered with Icelandic and Faroese. However, due to the isolated position of Iceland and the Faroes and intensive language contact between Norway and the rest of Scandinavia, modern Norwegian has become very similar to the modern languages of Denmark and Sweden. All varieties spoken in the Netherlands, including the Frisian varieties, cluster together, and German clusters more closely to these varieties than English.

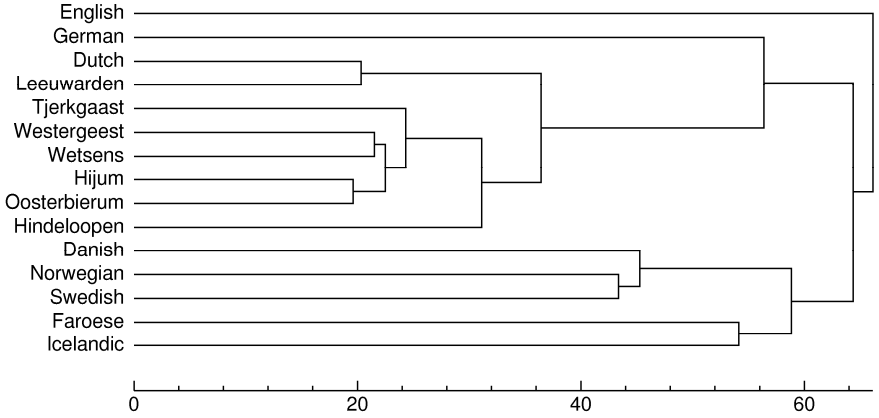


*Figure 5.* Dendrogram showing the clustering of the 14 language varieties in our study. The scale distance shows average Levenshtein distances in percentages.

All Frisian dialects form a cluster. This clustering corresponds well with the traditional classification as sketched in Section 3.1. The dialects of Hijum and Oosterbierum belong to Klaaifrysk and these dialects form a cluster. The Wâldfrysk dialects of Westergeest and Wetsens also cluster together. The Levenshtein distance between the four dialects is small,

ranging from 19.6% between Hijum and Oosterbierum and 23.8% between Oosterbierum and Westergeest. Also the Súdwesthoeksk dialects, represented by the Tjerkgaast dialect, are rather close to the Klaaifrysk and Wâldfrysk dialects (distances between 21.6% and 26.4%). The highly conservative dialect of Hindeloopen is more deviant from the other dialects (distances between 29.8% and 32.5%) and this is also the case for the Town Frisian dialect of Leeuwarden which is more similar to Dutch (20.3%) than to Frisian (between 32.3% and 35.8%) which confirms the characterization of Town Frisian by Kloeke (1927) as 'Dutch in Frisian mouth'.

## 5.2. The relationship between Frisian and the other Germanic languages

From Table 3 and 4 it is possible to determine the distance between all Germanic standard languages. We are especially interested in the position of Frisian within the Germanic language group. For this purpose the mean distance over the 6 Frisian dialects (excluding the dialect of Leeuwarden which is considered Dutch) has been added. This makes it possible to treat Frisian as one language. Examining the column which shows the ranking with respect to Frisian, we find that Dutch is most similar to Frisian (a mean distance of 38.7%). Clearly the intensive contact with Dutch during history has had a great impact on the distance between the two languages. Moreover, German appears to be closer to Frisian than any other language outside the Netherlands. Looking at the ranking with respect to Dutch, it appears that Town Frisian is most similar (Leeuwarden 20.3%), followed by the Frisian varieties (average of 38.7%). Next, German is most similar, due to common historical roots and continuous contact (a distance of 53.3%).

As discussed in the introduction, Friesland has a long history of language contact with the Scandinavian countries, and traces of Scandinavian influences can be found in the Frisian language. The impact of this contact is reflected in our results only to a limited extent. Remarkably, the distances to the mainland Scandinavian languages (Danish, Norwegian and Swedish) are smaller (between 60.7% and 63.3%) than to English (65.3%) even though the Frisian language is genetically closer related to English than to Scandinavian (see Section 2.1).

*Table 3.* Ranked Levenshtein distances in percentages between each of the five West Germanic languages and the other language varieties in the investigation.

| Frisian | | Leeuwarden | | Dutch | | English | | German | |
|---|---|---|---|---|---|---|---|---|---|
| | | Dutch | 20.3 | Leeuw | 20.3 | *Hindel* | 63.1 | Dutch | 53.3 |
| | | *Wetsens* | 32.3 | *Hindel* | 37.5 | *Wetsens* | 64.4 | Leeuw | 54.2 |
| | | *Westerg* | 32.7 | *Westerg* | 37.7 | Dutch | 64.7 | *Hindel* | 56.2 |
| | | Frisian | 34.2 | *Wetsens* | 38.3 | Swedish | 64.9 | *Westerg* | 56.9 |
| | | *Oosterb* | 34.3 | *Tjerkg* | 38.5 | Leeuw | 65.1 | *Oosterb* | 57.2 |
| | | *Hindel* | 34.9 | Frisian | 38.7 | *Tjerkg* | 65.2 | *Tjerkg* | 57.3 |
| Leeuw | 34.2 | *Tjerkg* | 35.3 | *Hijum* | 38.9 | Frisian | 65.3 | Frisian | 57.3 |
| Dutch | 38.7 | *Hijum* | 35.8 | *Oosterb* | 41.3 | *Hijum* | 65.8 | *Hijum* | 57.5 |
| German | 57.3 | German | 54.2 | German | 53.3 | *Westerg* | 65.8 | *Wetsens* | 58.6 |
| Swedish | 60.7 | Swedish | 59.2 | Swedish | 60.9 | Danish | 66.7 | Swedish | 61.0 |
| Norweg | 60.9 | Norweg | 60.0 | Norweg | 61.4 | Faroese | 67.1 | Danish | 63.5 |
| Danish | 63.3 | Danish | 61.1 | Danish | 63.4 | *Oosterb* | 67.2 | Norweg | 64.0 |
| English | 65.3 | English | 65.1 | English | 64.7 | German | 68.1 | Faroese | 67.1 |
| Faroese | 67.7 | Faroese | 67.5 | Faroese | 66.1 | Norweg | 68.6 | English | 68.1 |
| Icelandic | 70.0 | Icelandic | 69.6 | Icelandic | 69.2 | Icelandic | 69.1 | Icelandic | 68.5 |

*Table 4.* Ranked Levenshtein distances in percentages between each of the five North Germanic languages and the other language varieties in the investigation.

| Danish | | Swedish | | Norwegian | | Icelandic | | Faroese | |
|---|---|---|---|---|---|---|---|---|---|
| Norweg | 43.8 | Norweg | 43.4 | Swedish | 43.4 | Faroese | 54.1 | Swedish | 53.6 |
| Swedish | 47.0 | Danish | 47.0 | Danish | 43.8 | Swedish | 58.7 | Icelandic | 54.1 |
| Faroese | 58.5 | Faroese | 53.6 | Faroese | 57.2 | Norweg | 62.6 | Norweg | 57.2 |
| Leeuw | 61.1 | Icelandic | 58.7 | *Westerg* | 59.6 | Danish | 62.7 | Danish | 58.5 |
| *Westerg* | 62.2 | *Hindel* | 59.2 | Leeuw | 60.0 | German | 68.5 | Dutch | 66.1 |
| *Wetsens* | 62.3 | Leeuw | 59.2 | *Hindel* | 60.2 | *Tjerkg* | 69.1 | *Hindel* | 67.0 |
| Icelandic | 62.7 | *Westerg* | 59.6 | *Tjerkg* | 60.6 | English | 69.1 | English | 67.1 |
| *Hijum* | 62.9 | *Tjerkg* | 60.0 | *Wetsens* | 60.7 | Dutch | 69.2 | German | 67.1 |
| Frisian | 63.3 | Frisian | 60.7 | Frisian | 60.9 | Leeuw | 69.6 | *Westerg* | 67.4 |
| *Hindel* | 63.4 | Dutch | 60.9 | Dutch | 61.4 | *Hijum* | 69.8 | Leeuw | 67.5 |
| Dutch | 63.4 | German | 61.0 | *Oosterb* | 61.9 | Frisian | 70.0 | *Tjerkg* | 67.5 |
| German | 63.5 | *Wetsens* | 61.1 | *Hijum* | 62.6 | *Wetsens* | 70.1 | Frisian | 67.5 |
| *Tjerkg* | 63.8 | *Oosterb* | 61.4 | Icelandic | 62.6 | *Hindel* | 70.1 | *Oosterb* | 67.7 |
| *Oosterb* | 65.2 | *Hijum* | 62.7 | German | 64.0 | *Oosterb* | 70.3 | *Wetsens* | 68.1 |
| English | 66.7 | Icelandic | 64.9 | English | 68.6 | *Westerg* | 70.3 | *Hijum* | 68.2 |

So, when looking at the results from a Frisian perspective, the close genetic relationship with English is not reflected in our results. Of the Germanic languages in our investigation, only Icelandic and Faroese are less similar to Frisian than English. However, when looking at the results from an English perspective, we discover that of all Germanic language varieties in our material the Frisian dialect of Hindeloopen is most similar to English. As mentioned before, this dialect is highly conservative and furthermore it is spoken in a coastal place, which provides for easy contact with England. Also the Frisian dialect of Wetsens is more similar to English than the remaining Germanic languages. The other Frisian varieties are found elsewhere in the middle of the ranking. Among the non-Frisian varieties, Dutch appears to be most similar to English. However, all Germanic languages, including Frisian and Dutch, show a large linguistic distance to English, all distances being above 60%. The development of the English language has thus clearly taken place independently from the other Germanic languages, which can be explained by the strong influence from non-Germanic languages, especially French.

Also Icelandic shows a large distance to all other Germanic languages (from 54.1% to 70.0%), but in the Icelandic case this is explained by the conservative nature of this language rather than by language contact phenomena. Faroese is somewhat less conservative, but still shows rather large distances to the other languages (between 53.6% and 67.7%). The distances between the other Nordic languages are smaller (between 43.4% and 47%), as was expected given that the three Scandinavian languages are mutually intelligible.

## 6. Conclusions and discussion

Overall, the classification of the Germanic languages resulting from our distance measurements supports our predictions. This goes for the classification of the Frisian dialects and also for the rest of the Germanic languages. We interpret this as a confirmation of the suitability of our material showing that it is possible to measure Levenshtein distances on the basis of whole texts with assimilation phenomena typical of connected speech and with a rather limited number of words.

The aim of the present investigation was to get an impression of the position of the Frisian language in the Germanic language area on the basis of quantitative data. The fact that Frisian is genetically most closely related

to English yields the expectation that these two languages may still be linguistically similar. However, the distance between English and the Frisian dialects is large. We can thus conclude that the close genetic relationship between English and Frisian is not reflected in the linguistic distances between the modern languages. Geographical and historical circumstances have caused the two languages to drift apart linguistically. Frisian has been strongly influenced by Dutch whereas English has been influenced by other languages, especially French.

It would have been interesting to include these languages in our material. This would have given an impression of their impact on the English language. At the same time it would also have given us the opportunity to test the Levenshtein method on a larger language family than the Germanic family with its relatively closely related languages. It would also be interesting to include Old English in our material since this would give us an impression of how modern Frisian is related to the English language at a time when it had only recently separated from the common Anglo-Saxon roots to which also Old Frisian belonged.

For many centuries Frisian has been under the strong influence from Dutch and the Frisian and Dutch language areas share a long common history. It therefore does not come as a surprise that Dutch is the Germanic language most similar to the language varieties spoken in Friesland.

It may be surprising that the linguistic distances between Dutch and the Frisian dialects are smaller than the distances between the Scandinavian languages (a mean difference of 6%). Scandinavian languages are known to be mutually intelligible. This means that when, for example, a Swede and a Dane meet, they mostly communicate each in their own language. This kind of communication, which is known as semi-communication (Haugen, 1966), is not typical in the communication between Dutch-speaking and Frisian-speaking citizens in the Netherlands. The two languages are considered so different that it is not possible for a Dutch-speaking person to understand Frisian and consequently the Frisian interlocutor will have to speak Dutch to a non-Frisian person. Our results raise the question whether semi-communication would also be possible in a Dutch-Frisian situation. If this is not the case, we may explain this by linguistic and non-linguistic differences between the Frisian-Dutch situation and the Scandinavian situation. The Levenshtein distance processes lexical, phonetic and morphological differences. All three types are present in our transcription, since word lists are derived from running texts. Syntactic characteristics are completely excluded from the analysis. It might be the case that certain

characteristics play a larger role for the Levenshtein distances than desirable in the case of the Scandinavian languages if we were to use the method for the explaining mutual intelligibility. For example, it is well-known among the speakers of Scandinavian languages that many words end in an 'a' in Swedish while ending in an 'e' in Danish. Probably people use this knowledge in an inter-Scandinavian situation. However, this difference is included in the Levenshtein distances between Swedish and Danish. It is possible that Frisian-Dutch differences are less predictable or less well-known by speakers of the two languages. It is also possible that the difference in communication in the Netherlands and in Scandinavia should be sought at the extra-linguistic level. Scandinavian research on semi-communication has shown that the willingness to understand and the belief that it is possible to communicate play a large role for mutual intelligibility between speakers of closely related languages.

Staying with the Scandinavian languages, it should be noted that the mainland Scandinavian languages are in fact closer to Frisian than English, even though the Scandinavian languages belong genetically to another Germanic branch than English and Frisian. This can probably be explained by intensive contacts between Frisians and Scandinavians for many centuries. However, the common idea among some speakers of Frisian and Scandinavian that the two languages are so close that they are almost mutually intelligible is not confirmed by our results, at least not as far as the standard Scandinavian languages are concerned. Probably this popular idea is built on the fact that a few frequent words are identical in Frisian and Scandinavian. It is possible, however, that this picture would change if we would include more Danish dialects in our material. For example, it seems to be relatively easy for fishermen from Friesland to speak to their colleagues from the west coast of Denmark. Part of the explanation might also be that fishermen share a common vocabulary of professional terms. Also the frequent contact and a strong motivation to communicate successfully are likely to be important factors.

As we mentioned in the introduction, among dialects in the Netherlands and Flanders, the Frisian varieties are most deviant from Standard Dutch. However, among the varieties which are recognized as languages in the Germanic language area, Frisian is most similar to Dutch. The smallest distance between two languages, apart from Frisian, was found between Norwegian and Swedish: 43.4%. The distance between Frisian and Dutch is smaller: 38.7%. The Town Frisian variety of the capital of Friesland (Leeuwarden) has a distance of only 20.3% to Dutch. Although the

recognition of Frisian as second official language in the Netherlands is right in our opinion, we found that the current linguistic position of Frisian provide too little foundation for becoming independent from the Netherlands, as some Frisians may wish[8].

## Acknowledgements

This research would have been impossible without informants who were willing to translate the story of 'the Northwind and the Sun'. We wish to thank G. Blom (Hindeloopen), J. Spoelstra (Hijum) and W. Visser (Oosterbierum). All of them are affiliated with the Fryske Akademy in Leeuwarden. We also thank S. van Dellen (Wetsens), T. de Graaf (Leeuwarden), F. Postma (Tjerkgaast) and O. Vries (Westergeest), all of them employees of the University of Groningen. We thank J. Allen (England), A. Mikaelsdóttir (Iceland), Vigdis Petersen (the Faroes), R. Kraayenbrink (the Netherlands), K. Sjöberg (Sweden) and R. Schmidt (Germany). We are also very grateful to Jørn Almberg for making available the recording of Lillehammer (Norway). The recordings and transcriptions of the Frisian transcriptions are made by the second author, and those of the standard languages (except Norway and the Faroes) by the first author. The transcriptions subsequently were checked by Jørn Almberg who we thank gratefully for correcting our transcriptions. Furthermore, we wish to express our gratitude to Peter Kleiweg for his software for creating the map (Figure 2) and visualizing the dendrogram (Figure 5). Finally we thank Maartje Schreuder for reading an earlier version of this article and giving useful comments and Angeliek van Hout for reviewing our English.

## Notes

1 Dr. Tjeerd de Graaf, the central figure in this volume, was born in Leeuwarden, the capital of Friesland. Leeuwarden is one of the places where Town Frisian is spoken. Tjeerd de Graaf is a native speaker of this dialect, but later on he also learned (standard) Frisian. The Leeuwarden speaker in the present investigation was Tjeerd de Graaf (see Section 3.1).

2 Most of this section is based on König and Van der Auwera (1994).

3 The Lillehammer recording can be found at http://www.ling.hf.ntnu.no/nos/ together with 52 recordings of other Norwegian dialects.

[4] Since our material included two toneme languages, Swedish and Norwegian, also the two tonemes I and II were transcribed. For the other varieties primary stress was noted. Stress and tonemes were, however, not included for calculation of linguistic distances.

[5] 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.

[6] The data is taken from the *Linguistic Atlas of the Middle and South Atlantic States* (LAMSAS) and available via: http://hyde.park.uga.edu/lamsas/.

[7] The example should not be interpreted as a historical reconstruction of the way in which one pronunciation changed into another. From that point of view it may be more obvious to show how [ɛɪɑ̃⅓₇«₅ɡ₁₈ʃ₁] changed into [ɡɑ̃«⅓t«ɛɪn¬ʃn]. 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.

[8] Tjeerd de Graaf has never taken such an extreme position. Possibly speakers of Town Frisian have a more moderate opinion towards this issue since Town Frisian is more closely related to standard Dutch, as appeared in Figure 5 and Table 3.

## References

Bolognesi, R. and 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.). *Gramma/TTT: tijdschrift voor taalwetenschap*. Nijmegen University Press, Nijmegen, 9 (1): 45-84.

Feitsma, T. (1963). Sproglige berøringer mellem Frisland og Skandinavien. *Sprog og kultur*, 23: 97-121.

Gooskens, Ch. and W. Heeringa (submitted). Perceptive Evaluation of Levenshtein Dialect Distance Measurements Using Norwegian Dialect Data. (submitted to *Language Variation and Change*).

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

Haugen, E. (1966). Semicommunication: The Language Gap in Scandinavia. *Sociological Inquiry,* 36 (2): 280-297.

Heeringa, W. (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Doctoral dissertation. University of Groningen.

Heeringa, W. and C. Gooskens (2003). Norwegian Dialects Examined Perceptually and Acoustically. In: J. Nerbonne and W. Kretzschmar (eds.). *Computers and the Humanities*. Kluwer Academic Publishers, Dordrecht, 37 (3): 293-315.

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

Hoppenbrouwers, C and G. Hoppenbrouwers (1988). De featurefrequentie methode en de classificatie van Nederlandse dialecten. *TABU: Bulletin voor Taalwetenschap*, 18 (2): 51-92.

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

Johnson, K. (1997). *Acoustic and Auditory Phonetics*. Blackwell Publishers, Cambridge etc..

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*. Nijhoff, 's-Gravenhage.

König, E. and J. van der Auwera (1994). eds. *The Germanic Languages*. Routledge, London.

Miedema, H.T.J. (1966). *Van York naar Jorwerd. Enkele problemen uit de Friese taalgeschiedenis*. J.B. Wolters, Groningen.

Nerbonne, J., W. Heeringa, E. van den Hout, P. van der Kooi, S. Otten, and 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*. Antwerpen. University of Antwerp, Center for Dutch Language and Speech, 185-202.

Nerbonne, J. and W. Heeringa (1997). Measuring dialect distances phonetically. In: J. Coleman (ed.). *Workshop on Computational Phonology*. Madrid, 11-18.

Pei, M. (1966). *The story of language*. Allen & Unwin, London.

Potter, R.K., G.A. Kopp and H.C. Green (1947). *Visible Speech*. The Bell Telephone Laboratories Series. Van Nostrand, New York.

Rietveld, A.C.M. and V.J. Van Heuven (1997). *Algemene fonetiek*. Coutinho, Bussum.

Sandøy, H. (1994). Utan kontakt og endring? In: U.-B. Kotsinas and J. Helgander (eds.). *Dialektkontakt, språkkontakt och språkförändring i Norden*. Almqvist & Wiksell International, Stockholm, 38-51.

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

Stevens, K.N. (1998). *Acoustic Phonetics*. MIT Press, Cambridge.

The Columbia Encyclopedia (2001). www.bartleby.com/65/fr/Frisianl.html

The International Phonetic Association (1949). *The principles of the International Phonetic Association: being a description of the International Phonetic Alphabet and the manner of using it, illustrated by texts in 51 languages.* International Phonetic Association, London.

The International Phonetic Association (1999). *Handbook of the International Phonetic Association: a guide to the use of the International Phonetic Alphabet*. Cambridge University Press, Cambridge.

Trask, R.L. (1996). *A Dictionary of Phonetics and Phonology*. Routledge, London and New York.

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

Vieregge, W. H., A.C.M. Rietveld and C. Jansen (1984). A distinctive feature based system for the evaluation of segmental transcription in Dutch. In: M.P.R. van den Broecke and A. Cohen. *Proceedings of the 10th International Congress of Phonetic Sciences*. Foris Publications, Dordrecht and Cinnaminson, 654-659.

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

Wadstein, E. (1933). *On the Relations between Scandinavians and Frisians in Early Times.* University of London, London.

Wells, J. and J. House (1995). The sounds of the International Phonetic Alphabet. UCL, London.

Zwicker, E. and H. Fastl (1990). *Psychoacoustics and Models*. Springer Verlag, Berlin.

Zwicker, E. and R. Feldtkeller (1967). *Das Ohr als Nachrichtemfänger. Monographien der elektrischen Nachrichtentechnik*. 19, 2nd revised edition. Hirzel, Stuttgart.