

Chapter 9

Measuring Dutch dialect distances

On the basis of a small data set of Norwegian varieties we validated a range of different computational methods in Chapter 7. We found that the method with the highest score is a variant of the Levenshtein distance, where (i) segment distances are found on the basis of the Barkfilter representation, (ii) four length gradations are used, (iii) diphthongs are represented as a sequence of two segments, (iv) and logarithmic segment distances are used (Section 7.5.1). This method was applied to a larger set of Norwegian varieties in Chapter 8. In this chapter we apply the same method to Dutch dialects. We use data from the *Reeks Nederlandse Dialectatlassen* (RND), a series of Dutch dialect atlases which were edited by Blancquaert and Peé in the period 1925–1982. The goal of this chapter is to show that the application of the Levenshtein distance to the Dutch material of the RND gives interesting and useful results, despite the shortcomings of the RND transcriptions.

In Section 9.1, the RND will be discussed in more detail. In Section 9.2 the selected variant of the Levenshtein distance is applied to this data source, and the resulting distances are discussed. On the basis of these distances, the dialects are classified. Results of cluster analysis are presented in Section 9.3, where a main classification is given. In Section 9.4 each of the groups that are found in the main classification is discussed in more detail. In Section 9.5, results of multidimensional scaling are given. Using this technique the Dutch language area may be viewed as a continuum. The Dutch dialects are also compared to Standard Dutch. A ranking of difference with respect to Standard Dutch is given in Section 9.6. In Section 9.7 we draw some conclusions.

9.1 Data source

The *Reeks Nederlandse Dialectatlassen* (RND) is a series of atlases covering the Dutch dialect area. The Dutch dialect area comprises the Netherlands, the northern part of Belgium, a smaller northwestern part of France and the German county Bentheim. The atlas series consist of 16 parts. Although the Dutch language area consists of 16 provinces as well, the 16 volumes do not exactly correspond with the 16 provinces. In 1925, the first volume appeared, compiled by Blancquaert. The volume contains transcriptions of dialects in Klein-Brabant. The first recordings for this volume were already made in 1921 or 1922 (Goossens, 1997). After this, Blancquaert initiated a project in which recordings were made of varieties in the whole Dutch language area. To speed up the progress, Blancquaert engaged several collaborators. Unfortunately, Blancquaert died before all the volumes were finished. Peé was his successor and finished the project. The last recordings were made by Entjes in 1975. These recordings are found in part 14 (Zuid-Drenthe and Noord-Overijssel), that appeared in 1982 as a last installment.

9.1.1 Words

In the RND, the same 141 sentences are translated and transcribed in phonetic script for each dialect. Blancquaert mentions that the questionnaire was conceived as a range of sentences with words that illustrate particular sounds. The design saw to it that, e.g., possible changes of old-Germanic vowels, diphthongs and consonants are represented in the questionnaire. Morphologic and syntactic phenomena are also represented in the sentences (Blancquaert, 1948, p. 13). 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. These words were digitized for each dialect and used as input for the Levenshtein distance. The words represent (nearly) all vowels (monophthongs and diphthongs) and consonants. Also the consonant combination [sx] is represented, which is pronounced as [sk] in some dialects and as [ʃ] in some other dialects. The words are listed in Appendix B Table B.3.

Since the RND transcriptions consists of sentences, the same word may vary lexically in different dialects. We digitized and processed only forms which were semantically equal, as far as we could judge. Among different lexical nouns it may appear that one form has a determiner, and another has not. We always left out the determiner. In a sentence, assimilation phenomena can be found. So these phenomena are also found in the word transcriptions which we cut from the sentences. When two succeeding words in a sentence were not separated by a space, we tried to find the border between the two words by comparing the transcription with the transcriptions of other nearby varieties in which the words were separated by a space, or by comparing with comparable transcriptions in the

same text. If we found that the last segment of a word was shared with the first segment of the next word, we included the segment in both words when cutting them from the sentence. E.g. we split [ɔsəbludrɪŋkən] in [ɔsəblud] ‘ox-blood’ and [drɪŋkən] ‘drink’.

Sometimes a few of the expected words are missing for certain varieties, e.g. as the consequence of a free translation of some sentences. When two varieties are compared, and for one of the 125 words a translation is missing in one variety or in both varieties, the word is not taken into account in the calculation of the distance (see Section 5.1.10.1). For some words, more than one pronunciation was given, since e.g. an older and a newer form may be in circulation simultaneously. In these cases, the mean distance over the variants of one word is used for calculating the distance (see Section 5.1.10.2).

An extended discussion about the selection of words from the RND can be found in Heeringa (2001). The detailed presentation and discussion of the data is not repeated here, since we focus here on the analysis. The digitized data is publically available at <http://www.let.rug.nl/~heeringa/dialectology/atlas/> with the kind permission of the publisher, De Sikkel.¹

9.1.2 Varieties

The RND contains transcriptions of 1956 Dutch varieties. It would be very time-consuming to digitize all transcriptions. Therefore, we made a selection of 360 dialects. When selecting the dialects the goal was to get a net of evenly scattered dialect locations. First, we selected all locations in the RND which have more than 5,000 and fewer than 10,000 inhabitants. In smaller locations the dialect may be less stable due to moving or deaths, while in larger towns there may exist more than one dialect. Where the density remained too low, smaller locations were also used. Where an irregular pattern arose, the larger locations were sometimes replaced by smaller ones. A denser sampling resulted in the areas of Friesland and Groningen, and in the area in and around Bentheim.

In the map of Hof (1933, p. 14a) the Frisian area is divided into Bildts, Woudfries, Zuidhoeks and Stellingwerfs.² The selection was adjusted so that each of these groups was represented. A special group of Frisian varieties that do not form one geographical area are the ‘town Frisian’ varieties (*Stad(s)fries*). Just as the dialect of *het Bildt*, town Frisian dialects may be regarded as an intermediate form of Dutch and Frisian. Town Frisian varieties are spoken in Midsland, Dokkum, Harlingen, Franeker, Leeuwarden, Bolsward, Sneek, Heerenveen, Staveren and on the island of Ameland. In the map of Daan and Blok (1969) the town Frisian locations appear as language islands in the ‘pure’ Frisian language continuum. All of these locations are included in our data set. The map of Daan suggested that

¹Later on this publisher was taken over by De Boeck, Antwerpen.

²A clearer print of this map is found on the cover of the thesis of Breuker (1993).

the variety of Kollum belongs to Kollumerlands. In our results it will appear that this variety belongs rather to the town Frisian varieties. Therefore, we also regard Kollum as a town Frisian language island. The geographical island of Ameland is represented by Hollum and Nes. Since the collection of ‘pure’ Frisian locations have the same density as the sample in the remaining part of our study, and the town Frisian locations are added to them, we get a relatively higher density in Friesland.

For the Frisian locations of Appelscha, Donkerbroek and Tjalleberd two texts are given in the RND. 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. A Frisian and a *Stellingwerf* variety are also spoken in Donkerbroek. In the map of Daan and Blok (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 ‘Donkerbroek 1’ refers to the Frisian variety and ‘Donkerbroek 2’ to the *Stellingwerf* variety. In Tjalleberd, most people spoke Frisian when the RND recordings in Friesland were made. However, a minor 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 *Stellingwerf* variety as ‘Tjalleberd 2’.

In Reker (1993, p. x) the province of Groningen is divided in West-Groningen, North-Groningen, *Oldambt*, *Westerwolde*, *Veenkoloniën*, and the city of Groningen. The northern part of the province of Drenthe, south of the province of Groningen, is not displayed on this map. However, the varieties of this area are strongly related to the Groningen varieties. Because of personal interest, relatively more varieties are chosen in Groningen and North-Drenthe. The varieties are chosen so that each of the different areas is represented.

In part 14 (Zuid-Drenthe and Noord-Overijssel), varieties of the German county Bentheim are included. In a study about Dutch-German contact in and around Bentheim, the German transcriptions on the one side, and the Dutch transcriptions on the other side of the Dutch-German border are used (Heeringa et al., 2000). Since the same varieties are used in the present study, a higher density is found in and around Bentheim.

The RND includes also some varieties in the Belgium province of Luik. Just as the varieties in Bentheim, these varieties do not belong to the Dutch language area. The dialects are found south of the Dutch province of Limburg in the

northeastern part of the Belgium province of Luik. We selected Aubel and Baelen which belong to the French language area, and Eupen and Raeren which belong to the German language area.

The geographical distribution of the 360 Dutch varieties is shown in Figure 9.1. Since in Appelscha, Donkerbroek and Tjalleberd two varieties are spoken, the map shows only 357 localities. We divided the Dutch language area in a north-western, northeastern, southwestern and southeastern part. Each of these parts are visualized in more detail in the Figures 9.10, 9.13, 9.15, 9.17, 9.19, 9.21, 9.22 and 9.24. More about the selection of varieties from the RND can be found in Heeringa (2001).

To be able to compare the varieties with respect to Standard Dutch, we also 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 [kɔːmə], [roːzə] and [ɔːpə]. In the *Tekstboekje* of Blancquaert these words would end on an [n], just as suggested by the spelling. For more details see Heeringa (2001).

9.1.3 Groups

The most recent dialect map of the Dutch language was published in 1969 and compiled by Jo Daan (Daan and Blok, 1969). The map was already mentioned in Section 2.2.1 where we discussed the arrow method. With the arrow method, dialect borders are found on the basis of the perception of the dialect speakers. In this map the Dutch language area is divided into 28 different groups. The groups are mentioned in Table 9.1. The map in Figure 9.2 shows the classification of our set of varieties according to the map of Daan. For 49 borderline cases we found it unclear to which of the groups they belong. We left them out, so the map is eventually based on 311 varieties. In this set, 26 of the 28 groups are represented. Not represented are Daan's groups 8 and 16. Three small groups were represented by only one variety, namely group 2 (Egmond aan Zee), group 4 (Koog aan de Zaan) and group 12 (Geraardsbergen). In Section 9.3 we will compare the results of the Levenshtein distance to the classification that is given in the map of Daan.

9.1.4 Consistency

As mentioned above the RND consists of 16 different parts. The recordings were made during a time interval of more than 50 years. Therefore, differences in pronunciation may be in some cases the result of differences in time. The volumes of two adjacent areas never differ by more than 30 years, so the effect of temporal difference is mainly found when comparing varieties which are geographically more distant. Sometimes a rather large interval per volume was found as well. We



Figure 9.1: Distribution of the 357 localities, corresponding with 360 different varieties. White diamonds represent language islands and grey diamonds represent localities with two dialects where one of the two dialects is a dialect island. Circles represent small geographic islands. In the Figures 9.10 through 9.24 the different parts of the Dutch language area of shown in more detail.

- 1 Dialect of Zuid-Holland
- 2 Dialect of Kennemerland
- 3 Dialect of Waterland
- 4 Dialect of Zaan region
- 5 Dialect of northern Noord-Holland
- 6 Dialect of the province of Utrecht and the Alblasserwaard region
- 7 Dialect of Zeeland
- 8 Dialect of region between Holland and Brabant dialects
- 9 Dialect of West Flanders and Zeeuws-Vlaanderen
- 10 Dialect of region between West and East Flanders dialects
- 11 Dialect of East Flanders
- 12 Dialect of region between East Flanders and Brabant dialects
- 13 Dialect of the river region
- 14 Dialect of Noord-Brabant and northern Limburg
- 15 Dialect of Brabant
- 16 Dialect of region between Brabant and Limburg dialects
- 17 Dialect of Limburg
- 18 Dialect of the Veluwe region
- 19 Dialect of Gelderland and western Overijssel
- 20 Dialect of western Twente and eastern Graafschap
- 21 Dialect of Twente
- 22 Dialect of the Stellingwerf region
- 23 Dialect of southern Drenthe
- 24 Dialect of central Drenthe
- 25 Dialect of Kollumerland
- 26 Dialect of Groningen and northern Drenthe
- 27 Frisian language
- 28 Dialects of het Bildt, Frisian cities, Midsland, and Ameland Island

Table 9.1: In the map of Daan and Blok (1969) 28 groups are distinguished. In the map in Figure 9.2 the locations of the groups are displayed.

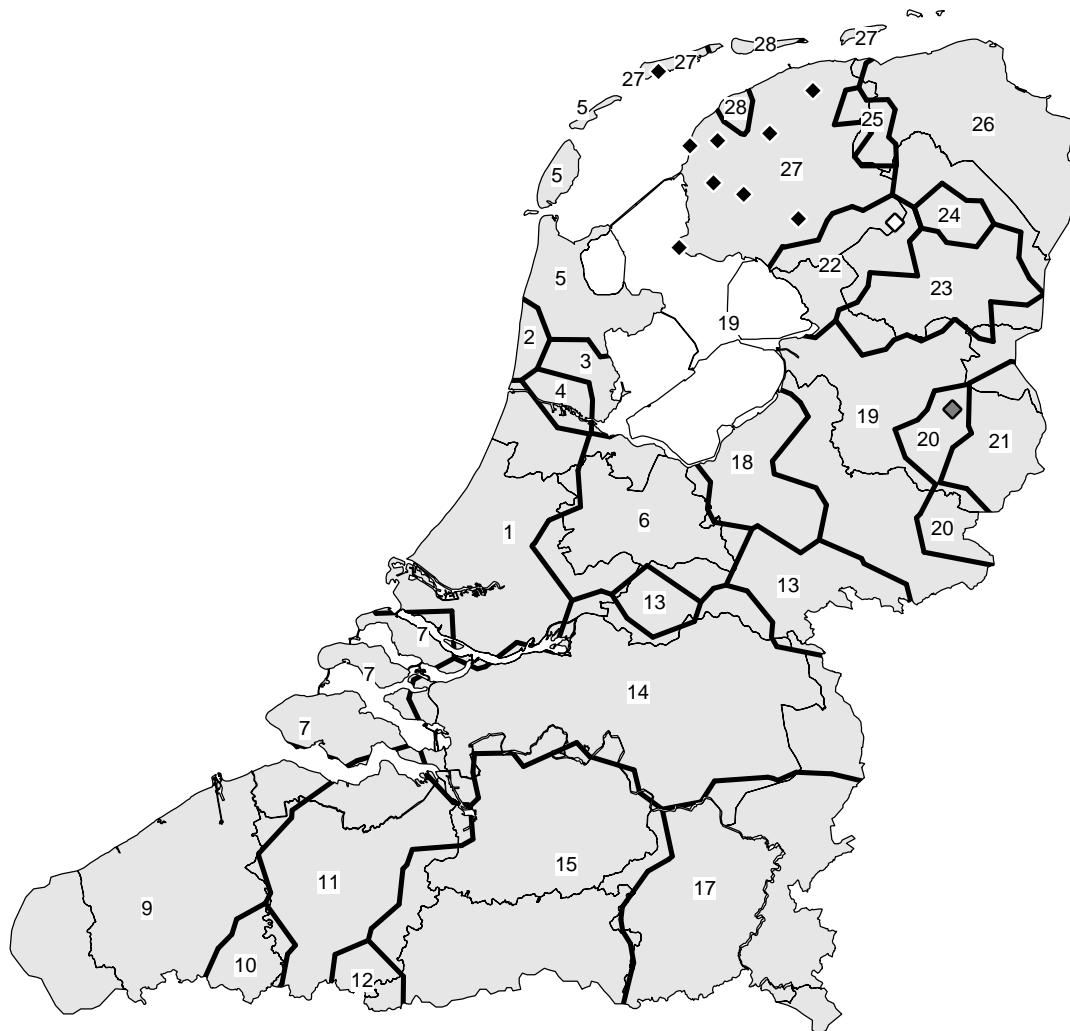


Figure 9.2: Locations of the 28 groups as distinguished in the map of Daan and Blok (1969). Provincial borders are represented by thinner lines and dialect borders by thicker ones. The numbers are explained in Table 9.1. Diamonds represent dialect islands. The black diamonds represent Frisian cities, which belong to group 28. The white diamond represents Appelscha, where both the dialect of group 22 and group 27 is spoken. The grey diamond represents Vriezenveen which contrasts strongly with its surroundings.

found the largest interval in volume 12 (Gelderland and Zuid-Overijssel) which was compiled during a period of 20 years. The 16 volumes of the RND were compiled by 16 different authors (and some assistants). Some authors worked on more than one volume, and some volumes were compiled by more than one author.

In Table 9.2 an overview is given of volumes and authors. The recording periods are borrowed from Reker (1997). The codes refer to the map in Figure 9.3. In this map, the areas per volume and per author are given. When we discuss results in Section 9.3, we can check whether the border we find represent borders between volumes or authors by consulting this map. For volume 2 (the southern part of East Flanders) the recordings are for the greater part made by E. Blancquaert, and for a lesser part by H. Vangassen in cooperation with Blancquaert. In the map, we mention both authors since the volume does not specify which transcriptions were made by which author. All locations in volume 5 (Zeeland Islands) were visited by both E. Blancquaert and P. J. Meertens. Therefore, we mention both authors again. In the map white diamonds represent dialects which were recorded by Blancquaert while surrounding dialects were recorded by one or more other authors. Grey diamonds in the area of volume 11 (Zuid-Holland and Utrecht) represent dialects which were recorded by Blancquaert and L. Oyen. Black diamonds in the area of volume 15 (Friesland) mark dialects recorded by Blancquaert, K. Boelens and G. van der Woude. By means of these joint recordings Blancquaert introduced the new fieldworkers to his system, ensuring a basic level of consistency.

Slightly different questionnaires are used in the RND. Generally speaking a Flemish version (volumes 1 through 8) and a Dutch version (volumes 9 through 16) were used. For volume 6 many words were replaced by French equivalents. For volume 15 in most cases a Frisian questionnaire was used, and for a few cases a Dutch questionnaire. Our list of 125 words contains only words for which equivalents are found in all questionnaires. A questionnaire may direct dialect speakers to a certain degree, especially on the lexical level. Therefore the use of different transcriptions can make the RND material less consistent. On the other hand, questionnaires were adapted so that they accord better with the dialect area for which they were used. This may result in more true transcriptions. A directing effect will be found only in transition zones. More about differences between questionnaires is discussed in Heeringa (2001).

Although the goal was that transcribers should work using Blancquaert's guidelines to ensure consistency, the different transcribers use slightly different notations (see also Goossens (1977, pp. 71–72)). Examining the transcriptions of different varieties, it is not always clear whether differences are transcriber differences or real pronunciation differences, which makes it hard to trace all transcriber differences. However, we found a limited set of differences that were obviously transcriber differences. In this section they will be discussed briefly. For each of them, we describe how we normalized the data for them. In Heeringa

Volume	Author(s)	Period	Code
1	E. Blancquaert	1922–1925	bla 1
2	E. Blancquaert and H. Vangassen	1927–1930	bla/van 2
3	E. Blancquaert	1933–1935	bla 3
4	H. Vangassen	1933–1935	van 4
5	E. Blancquaert and P. J. Meertens	1932–1939	bla/mee 5
6	Willem Pée	1934–1940	pee 6
7	Willem Pée	1946–1953	pee 7
8	J. C. Claessens	1937–1948	cla 8
8	W. Goffin	1937–1948	gof 8
8	A. Stevens	1937–1948	ste 8
9	A. Weijnen	1939–1949	wei 9
10	A. R. Hol	1949–1959	hol 10
10	J. Passage	1949–1959	pa 10
11	L. Oyen	1950–1962	oye 11
12	H. Entjes	1950–1970	ent 12
12	A. R. Hol	1950–1970	hol 12
13	Jo Daan	1950–1962	daa 13
14	H. Entjes	1974–1975	ent 14
15	K. Boelens	1950–1951	boe 15
15	G. van der Woude	1950–1951	wou 15
16	A. Sassen	1956–1961	sas 16

Table 9.2: List of volumes and authors of the RND together with the periods during which the recordings were made. The codes are used in the map in Figure 9.3.



Figure 9.3: Distribution of volumes and authors over the 360 RND varieties. Provincial borders are represented by thinner lines and volume/author areas by thicker ones. Black diamonds represent recordings of E. Blancquaert, white diamonds represent recordings of Blancquaert and L. Oyen, and grey diamonds represent recordings of Blancquaert, K. Boelens and G. van der Woude. The codes are explained in Table 9.2.

(2001) some of the same consistency problems are discussed more extensively, but the way in which they are solved may differ slightly from the way we described in this section.

9.1.4.1 Vowel + trill

In the RND, sometimes the *ee*, *oo* and *eu* before *r* are transcribed as respectively [e:], [o:] and [ø:] (see e.g., Blancquaert (1948)) and sometimes as [r̥], [ʋ̥] and [ɣ̥] (see e.g., the introduction of volume 13). Sometimes one author even used both notations intermixed (see e.g. the introduction of volume 16). To standardize different notations with the same meaning, we could replace each [e], [o] or [ø] before [r] or [ʀ] by respectively [ɪ], [ʊ] or [ɤ]. Since an [r] may also be weakened to a [ə], the [e], [o] and [ø] before [ə] should also be replaced by respectively [ɪ], [ʊ] or [ɤ]. However, it is not always clear whether an [ə] after an [e], [o] or [ø] is a weakened [r]. If not, the [e], [o] and [ø] should not be changed, to avoid that e.g., the relation between two different (dialect) pronunciations of *zee* ‘sea’, namely [ze:] and [ze:ə] (the latter would be changed to [zɪɹə]) is lost. However it is infeasible to determine the exact meaning of the large number of schwa’s in the large number of varieties. The other possibility is to replace each [ɪ], [ʊ] or [ɤ] before [r] or [ʀ] by respectively [e], [o] or [ø]. However, when applying these substitutions, problems arise since the *r* is deleted in some pronunciations. E.g. the relation between [pɪr̥rt], which is a dialect pronunciation of *paard* ‘horse’ (and which would be changed in [pe:rt]) and [pɪtjə], which is a dialect pronunciation of *paardje* ‘small horse’ will be lost. To overcome all the problems mentioned, we replaced simply each [ɪ], [ʊ] and [ɤ] everywhere by respectively [e], [o] and [ø], not only when they appear before [r], [ʀ] or [ə], but in all other contexts as well. On the one hand, in the IPA quadrilateral the substitutes are very close to the substituted vowels. On the other hand, some contrasts are lost. However, we prefer the loss of these contrasts to retaining contrasts that only reflect notation differences and no real differences in pronunciation.

9.1.4.2 Nasal + nasal

In volume 12 of the RND, we found that *bloemen* ‘flowers’ (sentence 2) was noted as [blo:m̥] by Entjes (dialect of Laren), and as [blum̥] by Hol (Spankeren). *Stenen* ‘stones’ (sentence 25) was noted as [ste:n̥] by Entjes (Groenlo), and as [ste:n̥] by Hol (Spankeren). *Brengen* ‘bring’ (sentence 39) is noted as [brɛŋ̥] by Entjes (Laren), and as [brɛŋ̥] by Hol (Spankeren). The examples show that transcribers do not note the Dutch ending *en* as pronounced in Low Saxon dialects in the same way. We found similar variation between and even within transcriptions of Flemish dialects. Although it is conceivable that some of the different transcriptions represent genuine pronunciation differences (and our procedures are equipped to deal with this), we preferred again to err on the side of caution.

In the introduction of volume 12, Entjes mentioned that he transcribed the word *kunnen* ‘can’ as [kyn̩], while Hol noted the same pronunciation as [kyn̩̩]. Entjes writes that he only heard one longer [n], and not two [n]’s as suggested by Hol. To make the data as consistent as possible we have to replace either the two-nasal notations by one-nasal notations, or the one-nasal notations by two-nasal notations. We prefer to use the two-nasal notations which are also suggested by Twilhaar (1990). Considering the one-nasal notations, we found for e.g. the Dutch word *spannen* ‘to put’ (in the context of: put a horse to a cart) the following transcriptions: 1) [span] (Nieuw Schoonebeek), 2) [span̩] (Oldemarkt), 3) [span̩:] (Blankenberge), 4) [span̩:] (Alveringem), and 5) [span̩] (Borger). Replacing the half-long, long and syllabic nasals as in the cases 2), 3), 4) and 5) can be done by an automatic procedure. Since the short nasal in case 1) represents probable de-gemination, we would like to let this short [n] also be replaced by a two-nasal notation especially to retain the relation with the half-long nasal. However, this cannot be done by an automatic procedure. Only nasals that correspond with the Dutch syllable *en* should be replaced. But not each short [n] corresponds with the Dutch syllable *en* as in case 1), so each short [n] in the data should be checked by hand. Therefore, we made a conversion in the other direction. We retained the notations where only one nasal is noted. This nasal may be noted as half long, long or syllabic. We replaced the two-nasal notations [mn̩], [nn̩] and [ɲn̩] by respectively [m̩], [n̩] and [ɲ̩] when they are found at the end of a word. If they are not found at the end of a word but rather are followed by a vowel, the substitutions are only made when the second nasal is noted as half-long, long or syllabic. If they are followed by another consonant, the substitutions are always made since in these contexts the second nasal can hardly be pronounced as non-syllabic. When replacing the two-nasal notation by a one-nasal notation, diacritics of the second nasal are left out. If the second nasal was respectively an [m], [n] or [ɲ], the same procedure was followed.

9.1.4.3 Plosive + nasal

In volume 12 of the RND, we found that *dopen* ‘baptize’ (sentence 35) was transcribed as [dø:p̩m̩] by Entjes (Groenlo) and as [dø:p̩n̩] by Hol. In volume 16 *Hebben* ‘have’ (sentence 106) was noted as [hɛb̩m̩] by A. Sassen (Bellingwolde). In volume 12 the same word was transcribed as [hɛb̩n̩] by Hol (Spankeren). In volume 12 Entjes noted *bakken* ‘bake’ (sentence 113) as [bak̩ɲ̩] (Wilp) while Hol noted this word as [bak̩n̩] (Hoenderlo). In part 15 G. van der Woude noted *geslagen* ‘hit’ (sentence 131) as [sl̩q̩g̩ɲ̩] (Kollum). The same author noted this word also as [sl̩q̩g̩n̩] (Dokkum).

In our opinion, an [n] after [p], [b], [k] and [g] is an unnatural pronunciation. On the other hand, an [m] after [p] or [b] and a [ɲ] after [k] or [g] may be pronounced easily. Therefore, we replaced [pn̩] by [p̩m̩], [bn̩] by [b̩m̩], [kn̩] by [k̩ɲ̩] and [gn̩] by [g̩ɲ̩] if the combinations were found at the end of a word. If they were

not found at the end of a word and followed by a vowel, the substitutions are only made when the nasal is noted as half-long, long or syllabic. If the nasal was followed by another consonant, the substitutions are always made since in these contexts the second nasal can hardly be pronounced as non-syllabic. Existing diacritics of either the plosive or the nasal are not changed.

9.1.4.4 Voiceless palatal plosive

In the feature table of Hoppenbrouwers and Hoppenbrouwers (2001, p. 40) the [c], [tj] and [tʲ] get the same definition. The [c] is only found in volume 16. We follow Hoppenbrouwers and Hoppenbrouwers by changing all [tj]’s and [tʲ]’s in [c]’s. For the [tj] the substitution is made regardless of whether the [t] or [j] or both are noted as extra-short.

9.1.4.5 Voiceless velar fricative

In the phonetic overview in volume 16 A. Sassen explicitly mentioned the [g] (RND notation) as a voiceless fricative. As a sample word Sassen gives the Dutch word *wasgoed* ‘wash’. We processed this simply as the IPA [x].

9.2 Distances

Using the Levenshtein distance, we find the distance between two pronunciations of the same word. The distance between two varieties is equal to the average of a series of Levenshtein distances computed from a series of word pairs. For 360 varieties, the average Levenshtein distance is calculated for each possible pair of varieties. The result is a 360×360 matrix. In Figure 9.4, the distances are geographically visualized. Strongly related varieties are connected by darker lines, while more distant varieties are connected by lighter ones. Where no lines are seen, the varieties are actually connected by white lines, indicating large distances.

In the picture, the strong relationships between the ‘pure’ Frisian varieties (Northwest) are clearly shown. When examining the picture, we should be aware of the fact that lighter lines, which represent the weaker relations between Frisian and town Frisian, are covered by the darker lines. Also the Groningen dialects (east of Frisian) form a group. Especially the most northern Groningen varieties are as close to each other as the Frisian varieties are. South of Groningen in Drenthe another small but close group is found. South of this group a large group is found in Overijssel. Especially the northern varieties are close. South of this group a sharp boundary is found, known as the boundary between Low Saxon (northeastern dialects) and Low Franconian (western, southwestern and southern dialects). In the rest of the map groups can also be found, although they are less distinct. However, when looking at the map from some distance, they



Figure 9.4: Average Levenshtein distances between 360 RND varieties. Darker lines connect close varieties, lighter lines more remote ones.

can be found. In the southwest, we find a French and West Flemish group. In the center of this area a group of strongly related dialects can be found. East of this area we find an East Flemish group, although the varieties are not so close. The same applies for the Zeeland varieties north of the East Flemish varieties. They are not close but emerge as a group since distances to other dialects are large, resulting in a lighter stroke around this set of varieties. East of the East Flemish group, we find an Antwerpen group (north) and a Brabant group (south). The two groups are connected in the East. East of these two groups, a lot of white area is seen, indicating the large distances that exist among the Limburg dialects. In the Dutch part of the Limburg area a small core area is found, where varieties are rather close. In the remaining part of the Dutch language area, it is hard to recognize groups on the basis of this map alone. Therefore, cluster analysis is described in Section 9.3. The result is an obvious division into groups. This division is compared to the division as shown in Daan's map.

Some borders suggested by the picture may not be real dialect borders. When looking at Figure 9.3, the border between Frisian and Groningen varieties, Frisian and Overijssel varieties and between Groningen and Overijssel varieties coincide with transcriber borders. This may be accidental, but we will keep track of it in the sections below.

9.3 Classification

9.3.1 Cluster analysis

On the basis of the distances between the 360 RND varieties we perform cluster analysis (see Section 6.1). The result is a large dendrogram in which all varieties are hierarchically ordered. In Figure 9.5 the dendrogram is displayed, showing only the 13 most significant groups. The scale distance shows percentages. The way in which percentages are found is described in the Sections 5.1.8 and 5.1.10. In the map in Figure 9.6, the 13 groups in Figure 9.5 are visualized by different colors. The colors are chosen by hand and inspired by the dialect map of Daan. When neighboring points belong to different groups, the exact border between the points is found on the basis of triangulation (see Section 6.1.5). To keep the picture simple, the dots and labels are given only for a restricted set of (in general) better-known locations. In the map, diamonds with and without labels can also be found. A diamond represents a language island, i.e. a variety which is only spoken in the location itself, and not in the area around the location.

We choose 13 groups since most of them correspond neatly with the groups which we found in the map in Figure 9.4. Some groups were not found in Figure 9.4. The Frisian mixed varieties were not found since they are geographically spread among the Frisian varieties. The Southwest Limburg group and the North-east Luik group were not found since they consist of only a few dialects and are

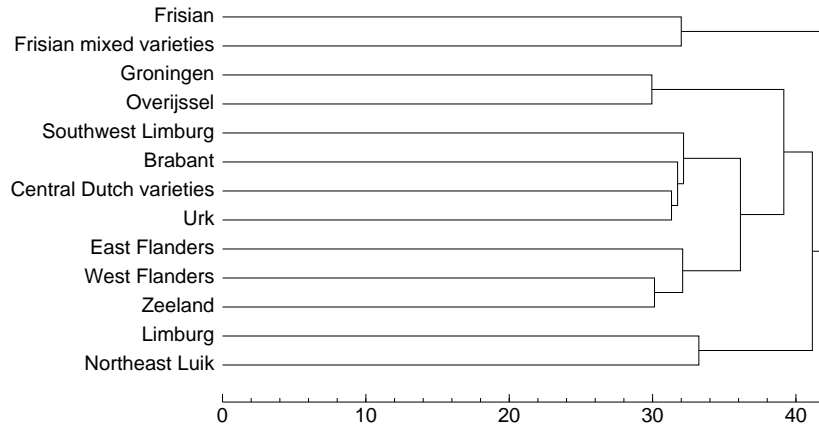


Figure 9.5: Dendrogram derived from the 360×360 matrix of Levenshtein distances showing the clustering of groups of Dutch dialects. UPGMA clustering is used (see Section 6.1.2). The scale distance shows percentages. Each of the 13 most significant groups is summed in one label and geographically visualized in Figure 9.6. The tree structure explains 70% of the variance.

considered a heterogenous area. Urk was not found as a ‘group’ since it consists of only one dialect. However, all of these groups are more significant than some other groups that were clearly found in Figure 9.4, e.g. the Zeeland group and the West Flanders group. When choosing more than 13 groups, we get groups that are not clearly recognized as groups in Figure 9.4 or groups with only a few, or even one dialect. Therefore, the number of groups in the main division was restricted to 13. In Section 9.4 each of the groups is discussed in more detail.

When considering the four main groups in the dendrogram in Figure 9.5, we get respectively Frisian (Frisian and Frisian mixed varieties), Low Saxon (Groningen and Overijssel), Low Franconian (Southwest Limburg . . . Zeeland) and Limburg varieties (Limburg and Northeast Luik). A difference between our division and the division of Hoppenbrouwers and Hoppenbrouwers (2001, p. 58) is that we find Frisian to be most distant from the central Dutch varieties, while in the division of Hoppenbrouwers and Hoppenbrouwers the Frisian varieties are more closely related to the dialects in Noord-Holland, Zuid-Holland and Brabant. This difference confirms our leading hypothesis that regarding words as linguistic units and considering the structure of a word is important. These two aspects are not processed in the methodology of Hoppenbrouwers and Hoppenbrouwers. Similar to the main division of (Hoppenbrouwers and Hoppenbrouwers, 2001, p. 58) we find that the Limburg varieties do not belong to the Low Franconian varieties, but form a separate group.

9.3.2 Area map

When comparing the map in Figure 9.6 with Daan's map in Figure 9.2, we found similarities and differences. Our Frisian group corresponds perfectly to group 27 in the map of Daan. Our group of Frisian mixed varieties includes group 28 in Daan's map, but also the northern part of group 22. Daan's group 27 contains town Frisian varieties (most diamonds on our map), Ameland island (the island north of Leeuwarden) and the dialect of *het Bildt* (the area northwest of Leeuwarden). Group 22 contains the *Stellingwerf* varieties (on our map northeast of Steenwijk and southwest of Assen). It is striking that Daan's group 27 and a part of group 22 are one group on our map. Possibly the speakers of the *Stellingwerf* area in Daan's study did not consider the town Frisian language islands, but focused mainly on the sharp contrast between the *Stellingwerf* continuum and the Frisian continuum in their judgments.

In three locations in the Frisian continuum, and in one variety in the *Stellingwerf* continuum, two varieties are spoken. In Tjalleberd (the higher diamond south of Grouw), a Tjalleberd variety and a Frisian variety are spoken. The lighter blue color in the diamond represents the Tjalleberd dialect island, and the darker blue around the diamond the Frisian variety. In Donkerbroek (the lighter blue diamond east of Grouw and west of Assen) both a Frisian and a *Stellingwerf* variety is spoken. The lighter blue color in the diamond represents the *Stellingwerf* dialect island and the darker blue color around the diamond the Frisian variety. In Appelscha (the darker blue diamond southwest of Assen) the same two varieties as in Donkerbroek are spoken. The darker blue color in the diamond represents the Frisian language island, the lighter blue color around the diamond the *Stellingwerf* variety. The three locations are discussed in more detail in the sections below.

The Groningen group corresponds with the groups 25, 26 and 24 of Daan's map. However, the south border on our map is found more southerly, probably as the result of transcriber differences (see Figure 9.3). The Overijssel group corresponds with the groups 23, 19, 20 and 21 of the map of Daan. The south border is the border between the Low Saxon area (north) and the Low Franconian area (south). The border does not coincide with the border between group 19 (north) and groups 18 and 13 (south) in Daan's map. The difference may be explained by transcriber differences that influenced our results. In Daan's map both the Groningen group and the Overijssel group are divided into smaller groups. In our results, a closer division is not found when regarding only the 13 most significant groups. Daan and Blok (1969, p. 28) writes that when dialect differences are small, misgivings may be justified as to whether non-linguistic criteria had greater influence than linguistic-criteria. For example differences in social and economic structure may influence the awareness of the speakers. Furthermore, we suppose that borders in heterogeneous areas may represent larger differences than in homogeneous areas. Our method does not reckon with non-linguistic

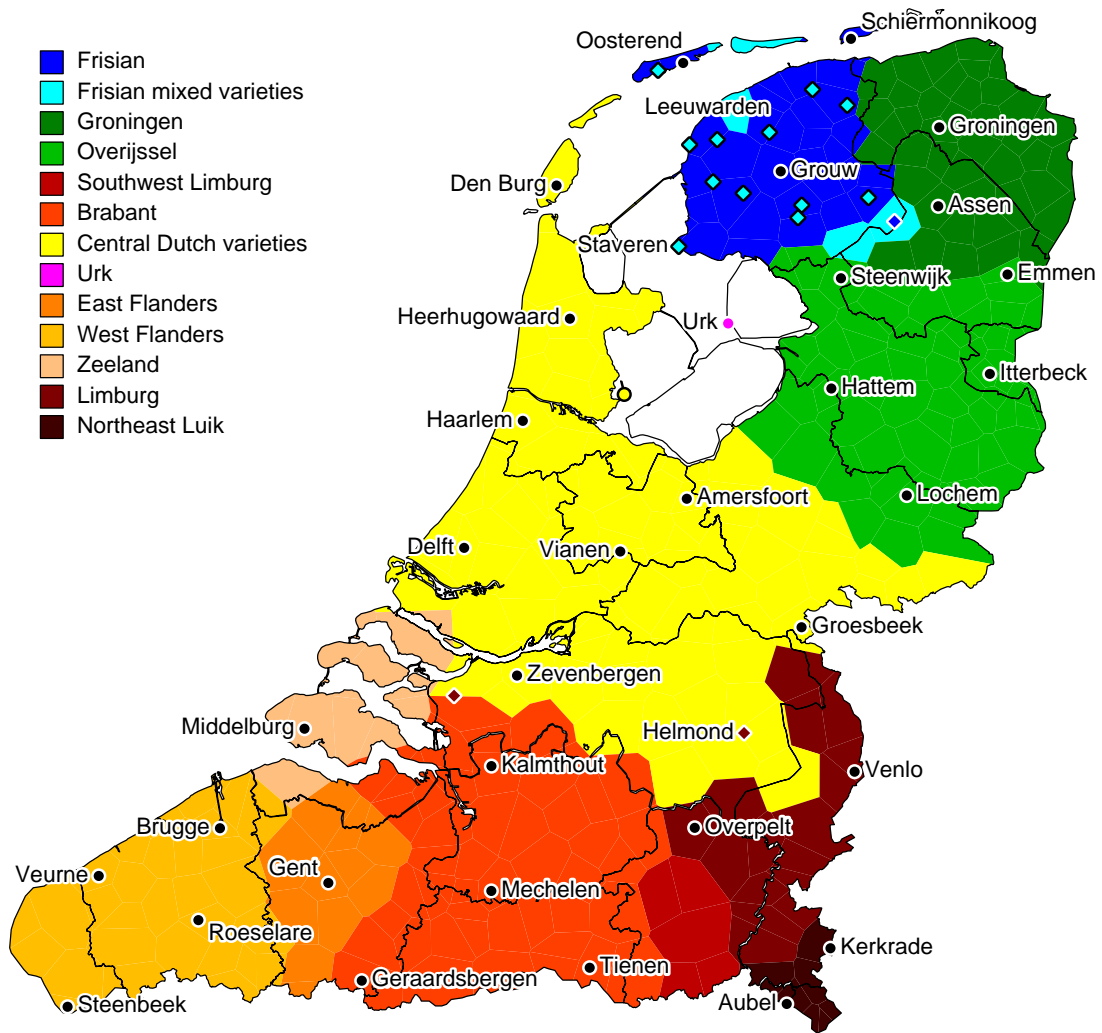


Figure 9.6: The 13 most significant groups as given in the dendrogram in Figure 9.5. UPGMA clustering is used (see Section 6.1.2). Diamonds represent language islands. Colors are chosen by hand and inspired by the dialect map of Daan.

factors on the one hand, and the degree of homogeneity on the other hand. This may explain why groups on our map are divided into several groups in Daan's map.

Our group of Central Dutch varieties corresponds with the groups 18, 6, 5, 2, 3, 4, 1, 13 and 14 in the map of Daan. The greater part of the south border corresponds rather well with the border in the map of Daan between the groups 1 and 14 (north) and the groups 7, 15 and 17 (south). Differences can be explained by transcriber differences. However, the northern part of the province of Limburg (the part north of Venlo and south of Groesbeek) is a part of group 14 in Daan's map, while it belongs to the Limburg group on our map. In the map of Te Winkel (1901) this area is separated from both group 14 and from the our Limburg group, while it is extended more to the north. We will discuss this in more detail in Section 9.4.12. Just as our Groningen and Overijssel group, the group of Central Dutch varieties is divided into a large number of groups in Daan's map. We mentioned some possible causes in the previous paragraph. However, for our Limburg group we found the opposite. Group 17 in the map of Daan divides into a southwest Limburg group, a Limburg group and a Northeast Luik group on our map. The fact that the borders did not exist in the awareness of the speakers may be explained by the fact that the Limburg area is a very heterogeneous area (Hoppenbrouwers and Hoppenbrouwers, 2001, p. 187). In an area with great dialect variation, it is more difficult to recognize groups (compare the Limburg area in the map in Figure 9.4). It is striking that the varieties of Steenbergen (the darker red diamond southwesterly of Zevenbergen) and Helmond were found among the Limburg dialects. Their relation with the Limburg varieties is discussed in Section 9.4.12.

Examining the remaining groups, we see that our Brabant group corresponds with groups 15 and 12 in Daan's map, our East Flanders group corresponds with group 11 of the map of Daan, our West Flanders group corresponds with groups 10 and 9 in Daan's map, and our Zeeland group corresponds with group 7 in the map of Daan. The border between the group of central Dutch varieties and the Brabant group is similar to the border between group 14 (north) and group 15 (south) in Daan's map, but is also influenced by transcriber differences. A striking difference between our map and the map of Daan is that the northeast part of group 11 forms a part of our Brabant group, and not a part of the East Flanders groups as may be expected on the basis of Daan's map. This is discussed in Section 9.4.6. Furthermore, the groups 12 and 11 form one group on our map as well as the groups 10 and 9. When examining these differences, we should be aware of the fact that the Belgian part of Daan's map is not based on the arrow method, but on information of language geographers who often belonged to dialect-speaking groups themselves. However, we did not know the exact criteria that were used by the language geographers, and the weightings of the criteria. Although the groups 12 and 10 can be found as separate groups on a deeper level (see Sections 9.4.9 and 9.4.10), their significance is not strong

enough for them to be recognized as groups in the map of Figure 9.4 or found among the 13 most significant groups.

Among the 13 groups, the dialect of Urk is also found as a separate group. In Daan's map, this dialect belongs to group 19. Possibly the dialect of Urk is most like the dialect of group 19, so speakers judge the dialect of Urk as nearly the same as the dialect of group 19, although strong differences may exist. In our results, Urk is closest to the group of central Dutch varieties, but nonetheless appears as one of the 13 most significant groups. For more details see Section 9.4.8.

9.3.3 Composite cluster map

On the basis of the dendrogram in Figure 9.5 (including the subtrees of the main groups) we create a composite cluster map. The map is shown in Figure 9.7. In this map the borders between the most significant groups are darker blue, blue and green. These borders can also be found in Figure 9.4 more or less, and distinguish the areas as shown in Figure 9.6. Less significant borders are lighter green, greenish yellow, yellow and lighter yellow. The least significant borders are white. The benefit of this picture compared to Figure 9.6 is that it shows both the main groups and further classifications per group. To keep a clear picture, dialect islands which belong to the Frisian mixed varieties, and the dialect islands Steenbergen and Helmond are excluded. The dialect of Urk belongs to one of the 13 main groups. However, no borders are drawn around this variety since this former island is isolated by an area in which only Standard Dutch was spoken.

Examining the 13 main groups, we find that the Frisian group appears as an homogeneous area. Only the dialect of Hindeloopen (southwest) appears to be rather deviant from the other varieties of this group. Frisian mixed varieties are spoken on the Ameland island (north of the mainland), in *het Bildt* (northwest on the mainland) and in the Stellingwerf area (along the southeast province border). The Groningen group is divided in a northern and a southern part. The border partly coincides with the province border between Groningen (north) and Drenthe (south). The Overijssel group is divided in a western and an eastern part. The western part appears to be very homogeneous. The eastern part is less homogeneous. On the border between the western and the eastern part, the dialect of Vriezenveen appears as a dialect island. This variety is discussed further in Section 9.4.4. In the map it can be seen that the large group of Central Dutch varieties is divided in a western and a southeastern part. The Limburg area is divided a Limburg group, a southwest Limburg group and a southeast Luik group. The southeast Luik group is divided in a western and a eastern group. The division in these two groups reflects the Benratherlinie (see Section 9.4.13). In the Brabant group we find in the furthest Southeast a small but rather deviant group. In this group we find a west/east division. In the two western varieties (Diest in the north and Tienen in the south) the uvular [ʀ] is used, just as in the varieties of the Limburg group and the Northeast Luik group. This may

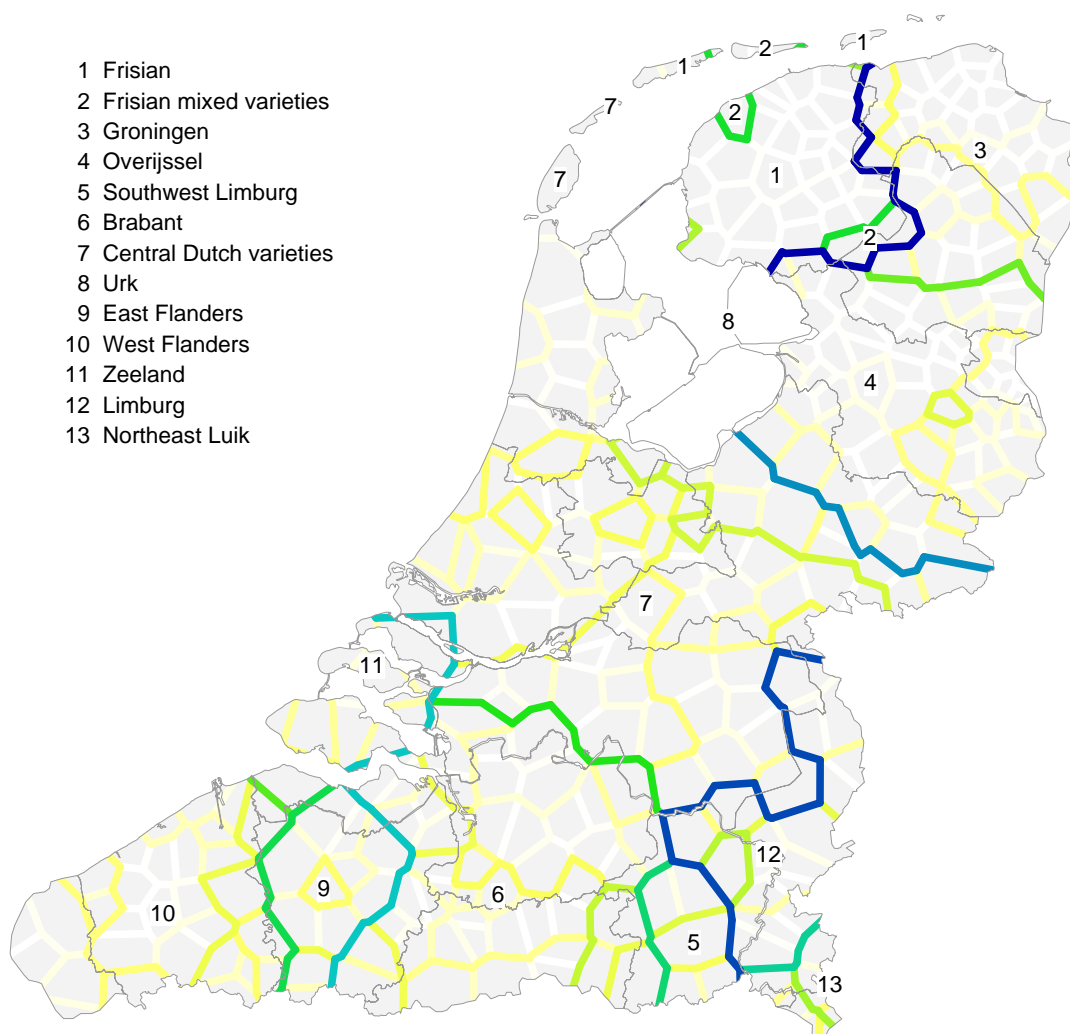


Figure 9.7: Composite cluster map on the basis of the dendrogram in Figure 9.5. UPGMA clustering is used (see Section 6.1.2). The most significant borders are represented by darker blue, blue and green lines, less significant borders by lighter green, greenish yellow, yellow and lighter yellow lines. The least significant borders are white. Dialect islands (the diamonds in Figure 9.6) are excluded.

distinguish them from the other Brabant varieties. In Daan's map this group belongs to the Limburg varieties, and not to the Brabant varieties. However in the western variety (Velm) the alveolar [r] is used. Examining the Brabant group further we find a rather separate group in the Northwest. This group is discussed in Section 9.4.6. Furthermore we find a division between Anwerpen varieties (north) and Brabant varieties (south). In the center of the East Flanders group the dialect of Gent is found as a dialect island. Phonologically the dialect of Gent differs strongly from the surrounding varieties. E.g. all vowels in the variety of Gent are longer than in the varieties around this city. In the West Flanders group the central part appears to be rather homogeneous. The varieties along the eastern province border form a separate group which possibly may be seen as a transition zone between East Flanders and West Flanders. In the furthest Southwest we find the French Flanders varieties as an separate group. They are separated from the other varieties of the West Flanders group by a transition zone. Finally we find some borders in the Zeeland group. However these borders are not so sharp.

In Section 9.4 the classification of each of the 13 main groups is discussed in more detail.

9.4 Classification per subgroup

In this section each of the 13 groups as found in the dendrogram in Figure 9.5 and displayed in the map in Figure 9.6 is discussed in more detail. In Sections 9.4.1 through 9.4.13 a dendrogram will be given per group. From a dendrogram a closer division per group can be derived. The smaller clusters correspond with smaller areas within the larger area. These areas are displayed in Figure 9.8. The grey diamonds represent varieties that do not actually belong to the group in which they are geographically found. The white dots represent locations where two varieties are spoken. The one belongs to the local group, the other is a dialect island that does not belong to the group. When considering this map one should be aware that the significance of the groups is different. However, the main goal of this map is to find the varieties in a cluster of a dendrogram more quickly in the map. In Sections 9.4.1 through 9.4.13 different parts of the map are shown in more detail.

9.4.1 Frisian

In the map of Daan, group 28 represents the 'pure' Frisian varieties. This group is not divided further. However, a closer division is given in Hof's map (1933, p. 14a). This map is based on isoglosses. Furthermore, a closer division of



Figure 9.8: Closer division of the Dutch language area on the basis clusters as found within each of the 13 main groups. Provincial borders are represented by thinner lines and dialect cluster borders by thicker ones. The grey diamonds represent varieties that do not actually belong to the group in which they are geographically found. The white dots represent locations where two varieties are spoken. The one belongs to the group, the other is a dialect island that does not belong to the group.

mainland-rural Frisian is given by Van der Veen (1986, 1994). Van der Veen obtained his division on the basis of computational processing of isoglosses. These isoglosses are based on high-frequency words gathered from different sources (e.g. the RND).

Our division is given in Figure 9.9. The locations of the varieties can be found in the map in Figure 9.10. Most distinct within the Frisian group are Schiermonnikoog, Oosterend, West-Terschelling and Hindeloopen. The first three varieties are found on islands, and Hindeloopen is known as an isolated place inhabited by fishermen with an archaic dialect (Hoppenbrouwers and Hoppenbrouwers, 2001, p. 99). Apart from these four varieties, we find a division in a northern and a southern group. When examining the map of transcriber borders in Figure 9.3 this division perfectly reflects the division between the two transcribers who made the recordings in the Frisian area. However, within each of these groups more interesting results can be found.

In the ‘northern’ group the cluster Hallum . . . Holwerd belongs to the Kleifries varieties, and the cluster Rottevalle . . . Bakkeveen to the Woudfries varieties. Looking at the ‘southern’ group, we find the variety of the small city of IJlst as most distinct, and next, a cluster containing the varieties of Koudum and Lemmer. This cluster represents the Zuidhoeks (Frisian: Súdhoeksk) varieties. Going one level deeper, we find two groups. The cluster Workum . . . Appelscha 2 may be considered as belonging to Woudfries (Frisian: Wâldfrysk), although Workum and Tjalleberd 1 are unexpected members of this group. For Workum we cannot explain this. For Tjalleberd this may have to do with the fact that in this place not only a Frisian, but a Low Saxon variety is spoken as well. The cluster, containing the varieties Oudeschoot . . . Spannum represents a part of the Kleifries (Frisian: Klaaifrysk) area.

When combining the ‘northern’ Woudfries varieties with the ‘southern’ Woudfries varieties, and the ‘northern’ Kleifries varieties with the ‘southern’ Kleifries varieties, we obtain a division in three groups (apart from Hindeloopen, the island varieties and IJlst): Woudfries, Kleifries and Zuidhoeks varieties. These groups correspond rather well with the groups as given in Hof’s map. However, the exact course of the border between Woudfries and Kleifries in our results is different from the course of the same border in Hof’s map, and is nearly equal to the course of this border as suggested in the map of Van der Veen (1994, p. 7). We have the most confidence in the map of Van der Veen because of his well-considered choice and weighting of the isoglosses.

9.4.2 Frisian mixed varieties

In Daan’s map group 27 includes the dialects of *het Bildt*, the Frisian cities, Midsland, and Ameland Island. In addition, group 22 is a transitional zone, consisting of the *Stellingwerf* varieties. In this section, the Frisian mixed varieties encompass group 27 and the northern part of group 22. We have not found a

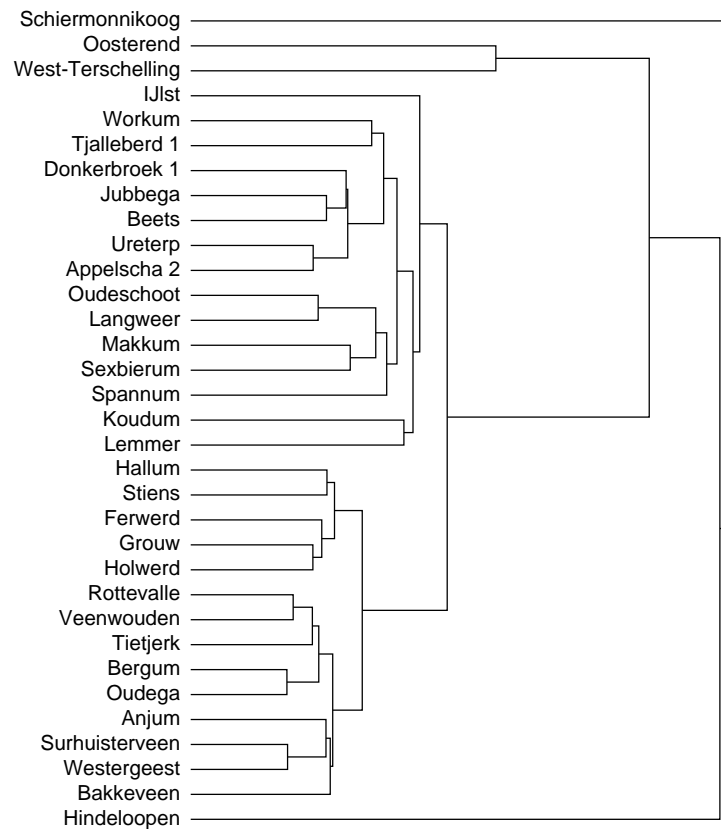


Figure 9.9: Subtree of the dendrogram in Figure 9.5, representing the Frisian group.



Figure 9.10: The northwestern part of the Dutch language area. Province borders are represented by thinner lines and dialect cluster borders by thicker ones. The grey diamonds represent varieties that do not actually belong to the group in which they are found geographically. The white dots represent locations where two varieties are spoken. The one belongs to the local group, the other is a dialect island which does not belong to this group.

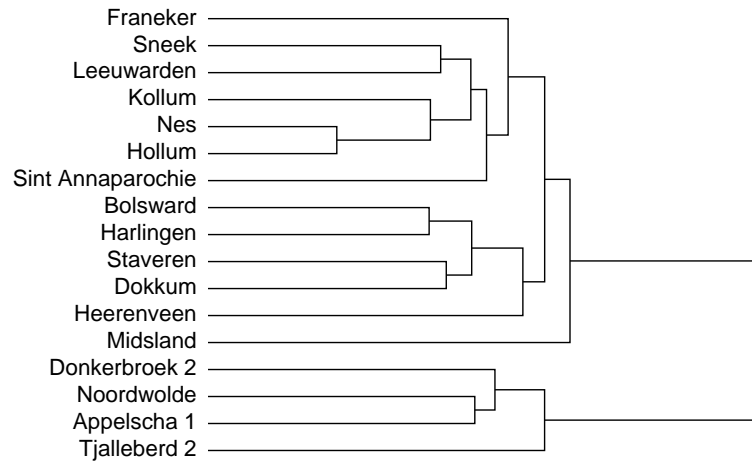


Figure 9.11: Subtree of the dendrogram in Figure 9.5, representing the Frisian mixed varieties group. The group contains town Frisian varieties and the *Stellingwerf* varieties. The dialect of *het Bildt* (represented by Sint Annaparochie) is found among the town Frisian varieties.

detailed discussion on the division of the groups 27 and 22 in the literature yet. In Daan's map, group 25 is also a transitional area, known as the *Westerkwartier*. This area is discussed in Section 9.4.3.

In Figure 9.11, a dendrogram is given that shows the division of the Frisian mixed varieties. The locations of the varieties can be found on the map in Figure 9.10. The dendrogram gives a clear division between *het Bildt* and town Frisian varieties on the one hand (cluster Franeker . . . Midsland), and the *Stellingwerf* varieties on the other hand (cluster Donkerbroek 2 . . . Tjalleberd 2).

Considering the cluster Franeker . . . Midsland, and ignoring Midsland we again find a northern and a southern group. Unfortunately, the border between these two groups represents the transcriber border (see Figure 9.3). Nonetheless, some conclusions can be drawn. First, the variety of Kollum clearly belongs to the town Frisian varieties, although it is found at the border of the Low Saxon area Kollumerlands (group 25) in Daan's map. The same finding was also found by Hoppenbrouwers and Hoppenbrouwers (2001, p. 96). Furthermore, the dialect of Sint Annaparochie (representing the dialects of *het Bildt*) does not appear as an outlier. Therefore, it is correct that in the map of Daan the dialect of *het Bildt* is considered as a town Frisian dialect. While the town Frisian varieties originated in making Frisian more Dutch, the dialect of *het Bildt* originated in making Dutch more Frisian. Our division suggested that these developments resulted in similar varieties.

Looking at the cluster Donkerbroek 2 . . . Tjalleberd 2, we found that the Low Saxon variety of Tjalleberd is clustered with the *Stellingwerf* varieties, although

it appears as the most deviant variety within this group. On the map of Daan, the *Stellingwerf* area is colored green, suggesting that this area belongs to the Low Saxon varieties. However, in our results the *Stellingwerf* varieties of Donkerbroek, Noordwolde and Appelscha are not found among the Low Saxon varieties. They form a cluster (together with Tjalleberd) which is clustered with the cluster of the dialects of the Frisian cities (and other related dialects). Hoppenbrouwers and Hoppenbrouwers (2001) who obtained a similar classification, cited Sassen (1953, p. 305) who described the *Stellingwerf* dialect as a Drenthe dialect which became more like Frisian. On the other hand, we found some other *Stellingwerf* varieties clustered among the Low Saxon varieties (see Sections 9.4.3 and 9.4.4). The fact that the *Stellingwerf* varieties in our data set are found among both Frisian and Low Saxon varieties may partly be explained by transcriber differences, and meanwhile the question cannot be answered whether *Stellingwerf* varieties are more related to Frisian than to Low Saxon, or the other way round.

9.4.3 Groningen

We labeled the group in this section as Groningen varieties since the greater part is found in this province of Groningen. However, the northern part of Drenthe also belongs to this group. The group encloses the groups 25, 26, 24 and a small northern part of group 23 of Daan's map. In the map of Reker (1993, p. x) the province of Groningen is divided in West-Groningen, North-Groningen, *Oldambt*, *Westerwolde*, *Veenkoloniën*, and the city of Groningen. The map is based on isoglosses. A map of Heeroma (1963) suggests that the varieties in the northern part of Drenthe are more related to the Groningen varieties than to the varieties in the southern part of Drenthe. In Daan's map Groningen and northern Drenthe are clearly one group.

The division of the Groningen and northern Drenthe varieties is given in Figure 9.12. The location of the varieties can be found on the map in Figure 9.13. At the highest level in the dendrogram we find two groups. The cluster Marum . . . Zoutkamp represents the *Westerkwartier* varieties, found as group 25 (Kollumerlands) on the map of Daan, and as the West-Groningen group in the map of Reker (1993). Going one level deeper, we get a division into Groningen dialects on the one hand (cluster Niekerk . . . Groningen), and dialects mainly found in the northern part of Drenthe on the other hand (cluster Onstwedde . . . Dwingelo). According to the map of Daan, the dialect of Zoutkamp should not belong to the *Westerkwartier* cluster, but to the cluster of Groningen varieties. The fact that this variety is classified with the *Westerkwartier* varieties in our results may be explained by the fact that Zoutkamp is a borderline case on the one hand, and transcribed by the same transcriber as the other 'real' *Westerkwartier* varieties on the other hand.

Examining the Groningen dialects, the cluster Niekerk . . . Adorp represents a northern group. This cluster corresponds with the North-Groningen group in

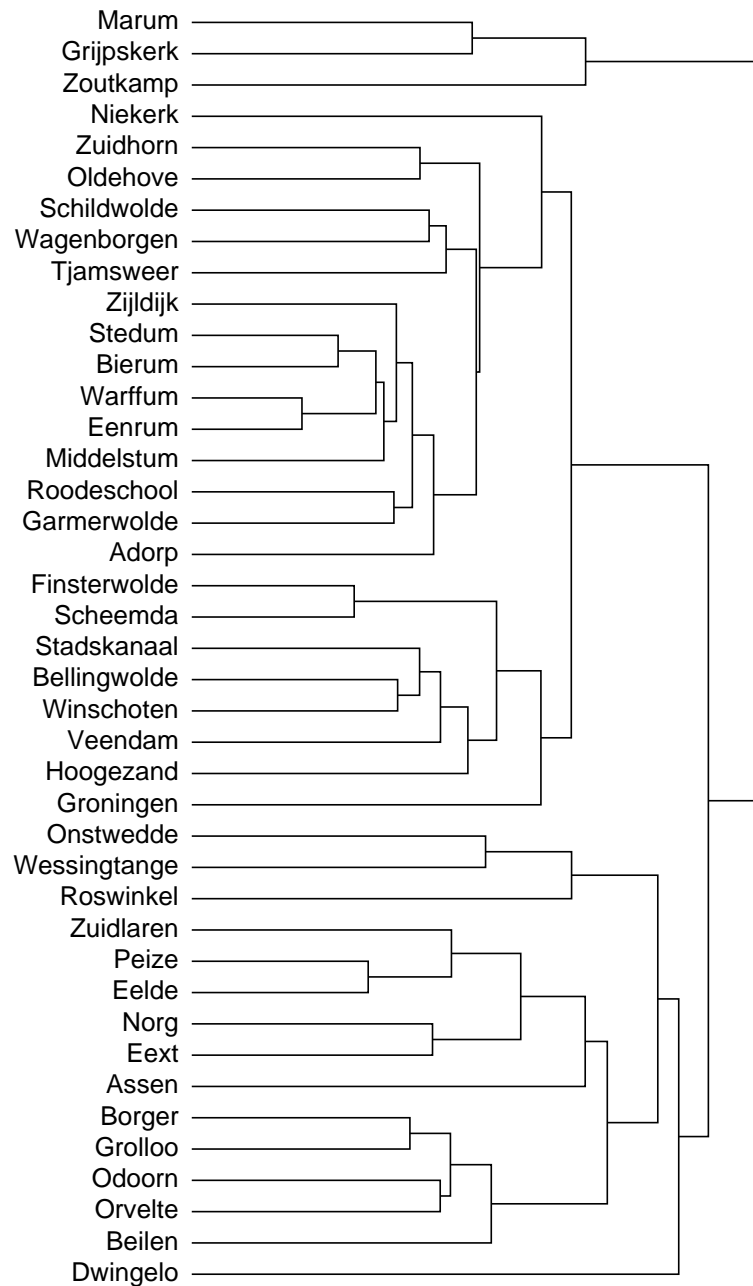


Figure 9.12: Subtree of the dendrogram in Figure 9.5, representing the Groningen group. The group contains varieties in Groningen and the northern part of Drenthe.

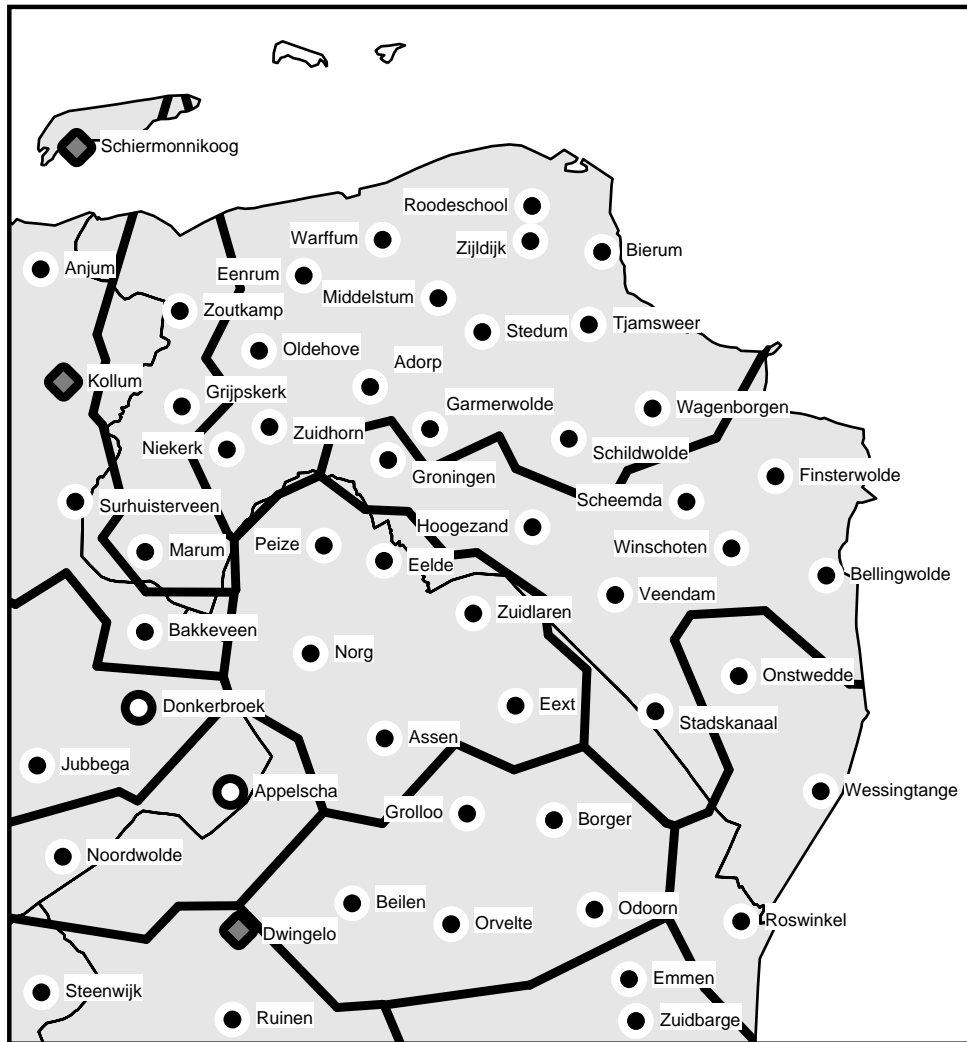


Figure 9.13: The northern half of the northeastern part of the Dutch language area. Provincial borders are represented by thinner lines and dialect cluster borders by thicker ones. The grey diamonds represent varieties that do not actually belong to the group in which they are found geographically.

the map of Reker. However, the cluster also encloses the northern part of the *Oldambt* area as given in the map of Reker. The cluster Finsterwolde . . . Groningen represents an eastern group. The cluster corresponds with the *Oldambt*, the *Veenkoloniën* and the city of Groningen. A clear division between *Oldambt* and the *Veenkoloniën* is not reflected by this cluster. The dialect of the city of Groningen is most distinct in this cluster.

Looking at the other cluster that contains mostly varieties in Drenthe, we find that Dwingelo is not clustered with any of the other varieties in this group. In the map of Daan, this dialect belongs to the *Stellingwerf* varieties (group 22). Unfortunately, Dwingelo was not found in the *Stellingwerf* group in the dendrogram in Figure 9.11. The map in Figure 9.3 makes it clear that transcriber differences caused this separation. Going one level deeper we find that the cluster Onstwedde . . . Roswinkel corresponds roughly with the *Westerwolde* area in the map of Reker. The other cluster contains varieties which are only found in Drenthe. The cluster Zuidlaren . . . Eext is the group that is suggested to be strongly related to the Groningen varieties in the map of Heeroma and which simply belongs to the Groningen group on the map of Daan. The cluster Borger . . . Beilen corresponds with the northern part of group of southern Drenthe varieties, found as group 23 on the map of Daan. However, the group halts exactly at a transcriber border.

Assen is clustered with the northern Drenthe group, although it appears as the most deviant variety in this cluster. In the map of Daan, it belongs to the central Drenthe varieties, mentioned as group 24. However, according to Daan's map, Grolloo also belongs to this group. In our dendrogram, Grolloo belongs to another cluster, namely the cluster of the southern Drenthe varieties. The transcription of Assen as given in the RND suggests that this variety is strongly influenced by Standard Dutch, which may explain why Assen is not too close to any of the other Drenthe varieties. Furthermore, the position of Grolloo does suggest that this variety fits perfectly in the southern Drenthe group rather than that it belongs to another group, namely the central Drenthe group. We found no explanation for this.

We may conclude that our results partly are in accordance with the map of Reker (1993). However, the border between the North-Groningen varieties and the *Oldambt* varieties is found further south in our results. In addition, we found no clear division between the *Oldambt* varieties and the *Veenkoloniën* varieties. To explain this difference, we should check whether the isoglosses used by Reker exist in the RND data. Furthermore, we clearly found a *Westerwolde* area, a Drenthe group and a southern Drenthe group. The varieties of the central Dutch group were either strongly influenced by Standard Dutch (Assen) or classified in the southern Drenthe group (Grolloo).

9.4.4 Overijssel

The Overijssel group encloses the southern part of Drenthe, Overijssel and the northern part of Gelderland. Since most varieties are found in Overijssel, we label the group simply the Overijssel group. The group corresponds with the southern part of the groups 22 and 23, the northern part of group 19, and with group 20 and 21 in Daan's map.

The dendrogram is given in Figure 9.14. The locations of the varieties are displayed in the map in Figure 9.15. In the dendrogram the deviant position of Vriezenveen immediately catches the eye. This agrees with Daan's map, in which the dialect is encircled, indicating that the dialect in this location is in strong contrast with its surrounding. Vriezenveen is an old settlement. The settlers came from the western coast area (Holland) (Entjes, 1970, pp. 2–15). The dialect has Westphalian influences.

Apart from Vriezenveen, we find two main clusters. The cluster Usselo ... Eibergen contains varieties in Bentheim and Twente while the cluster Nunspeet ... Bronkhorst contains the varieties around Bentheim and Twente.

The Bentheim/Twente cluster corresponds with group 21 in the map of Daan. In this cluster the cluster Rijssen/Eibergen and the dialect of Wierden are found to be rather apart from the other varieties in the cluster. The cluster Usselo ... Emlichheim represents varieties in or very close to German, where especially the cluster Langeveen ... Emlichheim contains varieties in and around Bentheim. Heeringa et al. (2000) report an investigation into the Dutch-German contact in and around Bentheim. The research was performed on the basis of the RND transcriptions and transcriptions of new recordings made in 1999. It appeared that the Dutch dialects shifted more towards Standard Dutch while all German dialects shifted towards Standard German. Finally the cluster Tubbergen ... Oldenzaal contains the core Twente varieties.

In the cluster with varieties around Bentheim and Twente, the dialect of Nunspeet, the cluster Vaassen/Bronkhorst and the dialect of Kuinder are found to be rather distinct. The deviant position of Nunspeet, Vaassen and Bronkhorst may be explained by the fact that they are found in a transition zone between the Low Saxon area and the Low Franconian area. According to Daan's map, the dialect of Kuinder belongs to group 22, i.e. the *Stellingwerf* varieties. Therefore, Kuinder would fit better in the *Stellingwerf* group in the dendrogram in Figure 9.11. The fact that this is not the case may be explained by transcriber differences. Apart from these four special cases, we find a cluster Emmen ... Nieuw Schoonebeek in the utmost southeast part of Drenthe, and a cluster Lochem ... Kampen containing varieties west of the Bentheim/Twente group. The southeastern Drenthe group partly covers group 23 and partly group 19 in Daan's map. In the group west of Bentheim/Twente the cluster Lochem ... Wijhe form a southern cluster partly corresponding with group 20 in the map of Daan, and the cluster Hattem ... Kampen form a northern cluster, covering the utmost

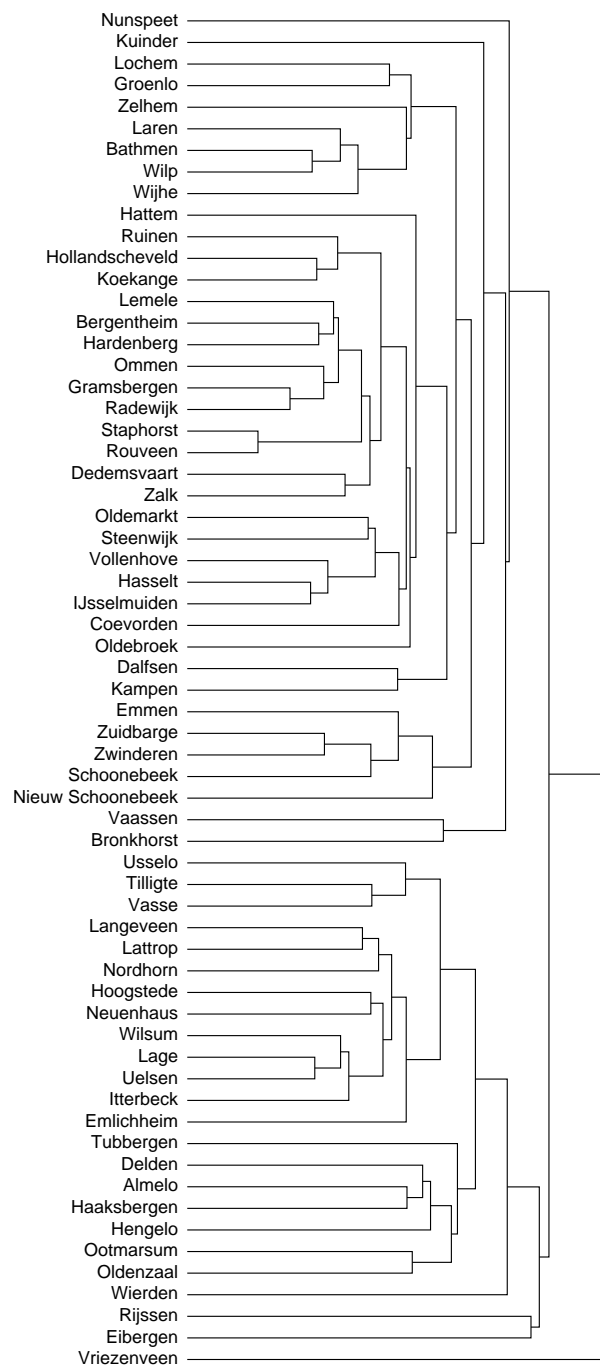


Figure 9.14: Subtree of the dendrogram in Figure 9.5, representing the Overijssel group. The group contains varieties in the southern part of Drenthe, Overijssel and the northern part of Gelderland.



Figure 9.15: The southern half of the northeastern part of the Dutch language area. The state border and the provincial borders are represented by thinner lines and dialect cluster borders by thicker ones. The grey diamonds represent varieties that do not actually belong to the group in which they are found geographically.

southern part of group 23, and the larger northern part of group 19 on the map of Daan. In the latter cluster, it is striking that the cluster Dalfsen/Kampen is rather apart. Characteristic for these varieties is the use of the uvular [ʀ] which is uncommon for Low Saxon varieties. This uvular [ʀ] is also found in Zwolle, Deventer and Zutphen, but these varieties are not included in our data set. More about this phenomenon can be found in Hoppenbrouwers and Hoppenbrouwers (2001, p. 92).

In the large cluster Hattem . . . Kampen, we find a small cluster containing the varieties of Ruinen, Hollandscheveld and Koekange. On the map of Daan, they belong to group 23, i.e. the southern Drenthe varieties. We expected them to be clustered with the southern Drenthe group in the dendrogram in Figure 9.12. The fact that they are absent there and found instead among the Overijssel varieties may be explained by transcriber differences. Furthermore, the varieties of Oldemarkt, Steenwijk and Vollenhove belong to group 22 in the map of Daan, i.e. the *Stellingwerf* varieties. Are they misclassified due to transcriber differences as well? We are not sure since they are clearly separated from Kuinder, which is also a *Stellingwerf* variety. Since Oldemarkt, Steenwijk and Vollenhove fit perfectly among the other Overijssel varieties while Kuinder is rather apart, we suspect that Kuinder is still a real *Stellingwerf* variety, while the other varieties are much more strongly related to the Overijssel varieties.

In the dendrogram discussed in this section, we found some differences with Daan's map, which cannot be explained by transcriber differences. The map of Daan is based on material from 1939 while the RND transcriptions are based on recordings made in 1974–1975 (south Drenthe and north Overijssel) and 1950–1970 (south Overijssel). Indeed, we found no explanation other than the fact that the situation has changed since 1939. In the dendrogram, we found a small southeast Drenthe group. This group suggested that the southeastern border of group 23 is shifted to the south. Furthermore, we expect that the southeast border of group 22 was shifted more to the west, since most varieties, which should belong to the *Stellingwerf* group in the map of Daan, do not appear as deviant to the other Drenthe and Overijssel varieties, with the exception of Kuinder.

9.4.5 Southwest Limburg

The southwest Limburg area is a small part of group 17 in Daan's map. Group 17 covers a large area, labeled as the dialect of Limburg. It is striking that our southwest Limburg area is not found as a separate group in the map of Daan. Nonetheless, it may be not surprising that a part of group 17 emerges as a separate group. It is known that the situation is complex in this area, and the varieties do not form a homogeneous group (Hoppenbrouwers and Hoppenbrouwers, 2001, p. 187). The dendrogram for our southwest Limburg area is given in Figure 9.16. The locations of the varieties can be found in the map in Figure 9.17.

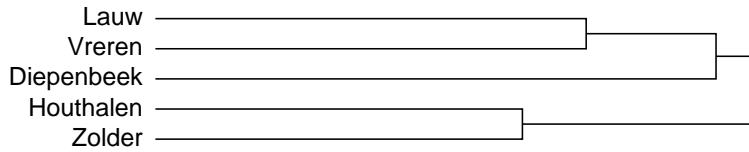


Figure 9.16: Subtree of the dendrogram in Figure 9.5, representing the Southwest Limburg group.

Although the borders of the southwest Limburg area partly coincide with transcriber borders, we think that this is not the only explanation why this set of varieties appears as one of the 13 most significant groups. The varieties distinguish themselves from the varieties in our larger Limburg group and smaller Northeast Luik group through the fact that the *r* is pronounced as an alveolar [r], and not as an uvular [ʀ]. The division as given by the dendrogram is not surprising: it follows the geography.

9.4.6 Brabant

The Brabant varieties roughly match group 15 in the map of Daan. However, a small number of varieties belongs to group 11, and the dialect of Geraardsbergen belongs to group 12. The division is shown in Figure 9.18 while the localities can be found in Figure 9.19.

The main division consists of the cluster Velm . . . Tienen and the cluster Lot . . . Geraardsbergen. All varieties in the first cluster are borderline cases. Velm just belongs to group 17 while Tienen and Diest marginally belong to group 15. On the original map of Daan, intermediate between group 15 and group 17 we find group 16: the dialect of the region between Brabant and Limburg. Since none of the varieties in our set of dialects belongs to this area, this area is not found in the map in Figure 9.3. However, if we assume that the borders in the map of Daan are drawn too narrowly or that the dialect area has been expanded since 1939, the dialects of Tienen, Diest and Velm represent a part of this dialect region.

In the second cluster we find a subcluster Lamswaarde . . . Geraardsbergen and a subcluster Lot . . . Arendonk. Most varieties in the subcluster Lamswaarde . . . Geraardsbergen are found in the northeast of group 11. It is striking that these varieties are not clustered with the other varieties of group 11, which are found in the dendrogram in Figure 9.23 (see Section 9.4.9). Obviously, this cannot be explained by transcriber differences. Furthermore, in this cluster we find the varieties of Geraardsbergen and Helderger. Geraardsbergen belongs to group 12 in the map of Daan, and Helderger is a borderline case. The dialects are geographically isolated from the other varieties in this cluster and found further south, Geraardsbergen so much so that is close to the Flemish/Walloon border.

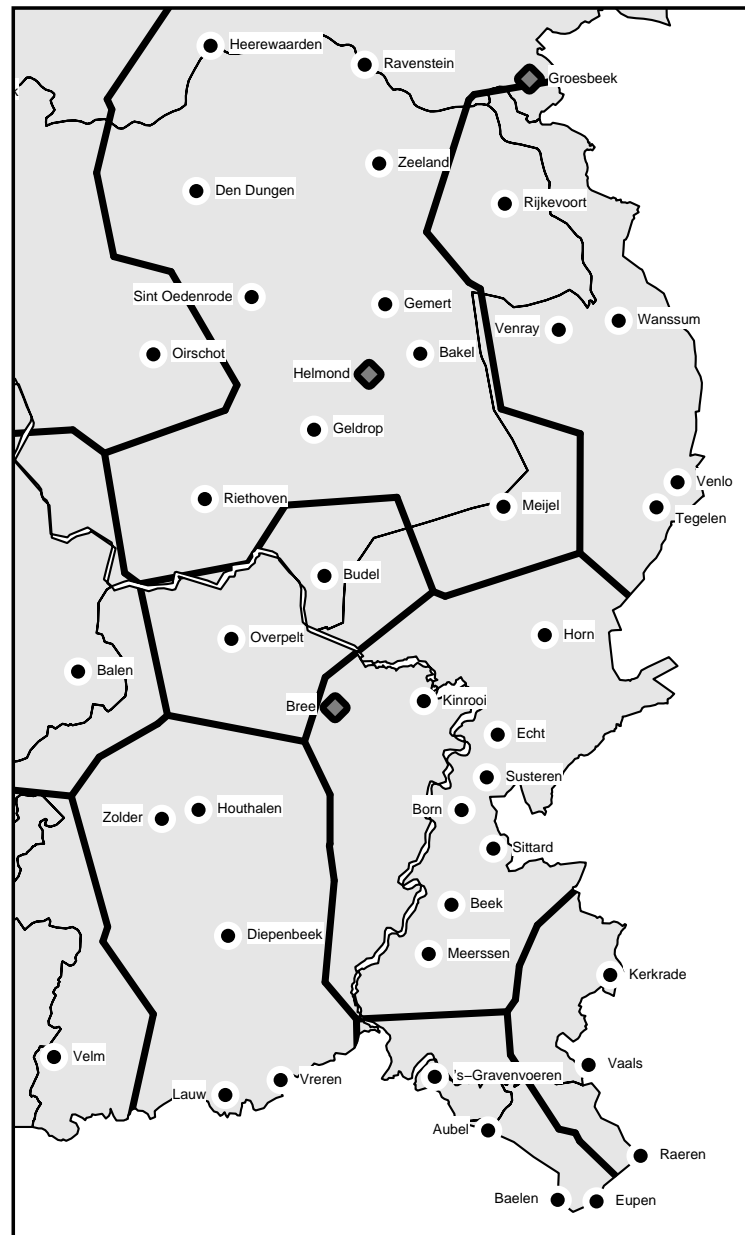


Figure 9.17: The southern half of the southeastern part of the Dutch language area. State borders and provincial borders are represented by thinner lines and dialect cluster borders by thicker ones. The grey diamonds represent varieties which, in our measurements, do not actually belong to the group in which they are found geographically.

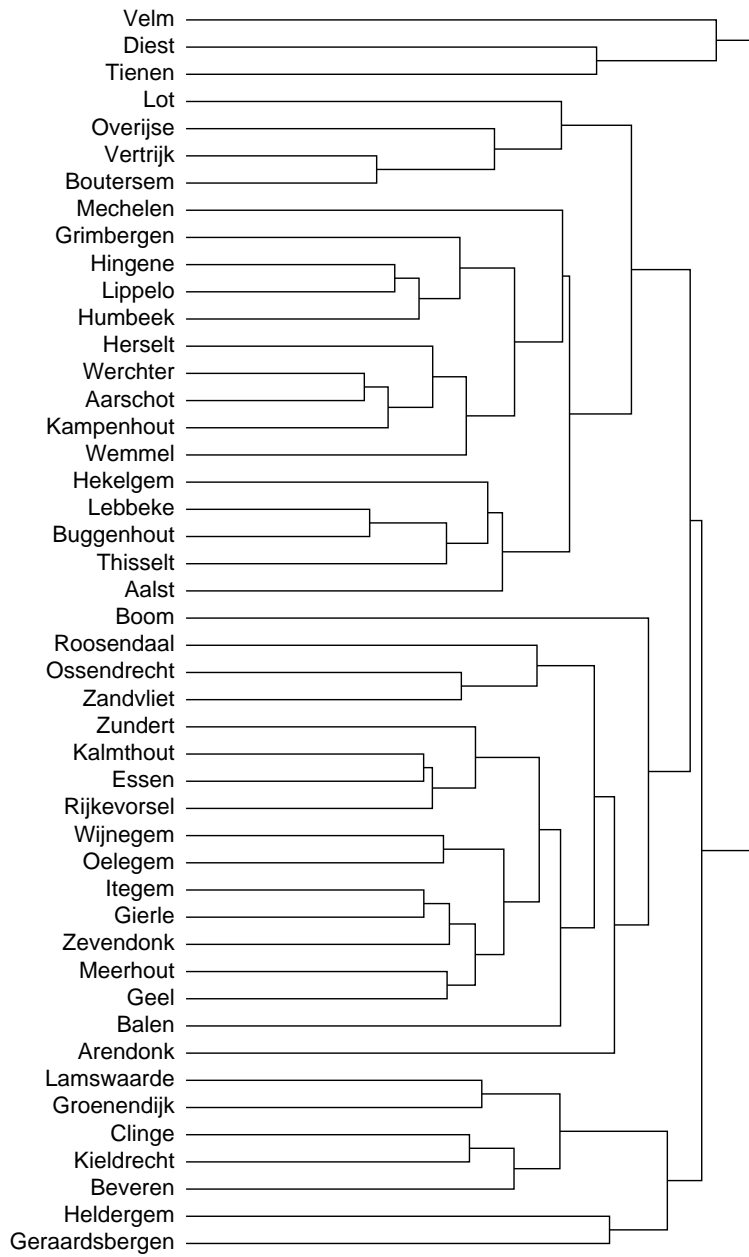


Figure 9.18: Subtree of the dendrogram in Figure 9.5, representing the Brabant group. The group contains varieties in Antwerpen and Brabant.

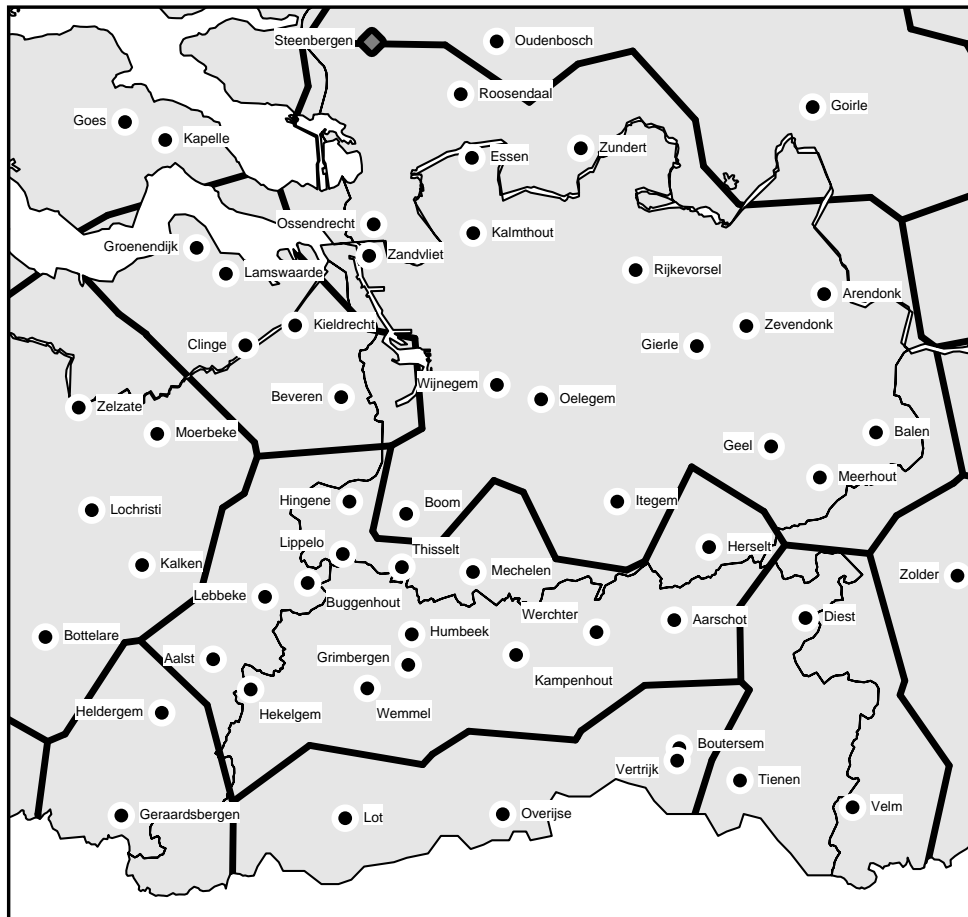


Figure 9.19: The mid-southern part of the Dutch language area. The state border and province borders are represented by thinner lines and dialect cluster borders by thicker ones. The grey diamond represents a variety that, in our measurements, does not actually belong to the group in which it is found geographically.

In the map of Daan, group 12 forms actually a small strip from the North to the South. Since this part of the map is not based on the arrow method, but on knowledge of language geographers, errors can be made because of the unexpected course of the borders. We therefore, suspect that group 12 should be expanded further to the north, covering the northeastern part of group 11.

In the subcluster Lot ... Arendonk we find a group containing southern varieties (Lot ... Aalst) and northern varieties (Boom ... Arendonk). In the group Lot ... Aalst, we in turn find a southern subgroup (Lot ... Boutersem) and a northern subgroup (Mechelen ... Aalst). These two subgroups roughly cover the province of Brabant. The group Boom ... Arendonk approximately matches the province of Antwerpen. Although the northern border roughly matches with the border as given in Daan's map, it matches perfectly with the transcriber border (see Figure 9.3).

9.4.7 Central Dutch varieties

The area of the central Dutch varieties is more or less found in the center of the Dutch language area, intermediate between the Frisian, Groningen and Overijssel varieties (northern) and the Zeeland, Flemish and Limburg varieties (southern). In our results this large central area is divided into a western and an eastern part only when examining the 28 most significant groups. The group encloses the southern part of the groups 18 and 19 (the central part of Gelderland) in the map of Daan, group 13 (the southern part of Gelderland) and group 14 (Noord-Brabant), group 1 (Zuid-Holland), group 6 (Utrecht) and the groups 5, 2, 3 and 4 (Noord-Holland). A division of the central Dutch varieties is given in Figure 9.20. The locations of the varieties are found in the maps in the Figures 9.21 (the eastern parts of Utrecht and Noord-Brabant and the southern part of Gelderland), 9.22 (Zuid-Holland and the western parts of Utrecht and Noord-Brabant) and 9.10 (Noord-Holland). When discussing the groups below, we will refer to the relevant map for the main groups.

Examining the dendrogram, the position of Huizen strikes the eye. This variety is most deviant from the other dialects. In Daan's map the special position of Huizen cannot be found. However, in Te Winkel's map (1901) the dialect of Huizen and its surroundings are suggested as a separate group, labeled as 'Gooisch'. Te Winkel's map is based on data obtained from questionnaires. Data was obtained for 383 regions and places. An exact account was not given for the division (Daan and Blok, 1969, p. 18).

When going one level deeper, the cluster Aalten ... Woudenberg represents a central Gelderland group (see Figure 9.21) while the cluster Groesbeek ... Utrecht contains other central Dutch varieties. In the main division in Hoppenbrouwers and Hoppenbrouwers (2001, p. 58), the central Gelderland group is clustered close to the Frisian and Frisian mixed varieties. In our results, this group is much more distant from Frisian varieties, and closer to the central Dutch varieties.

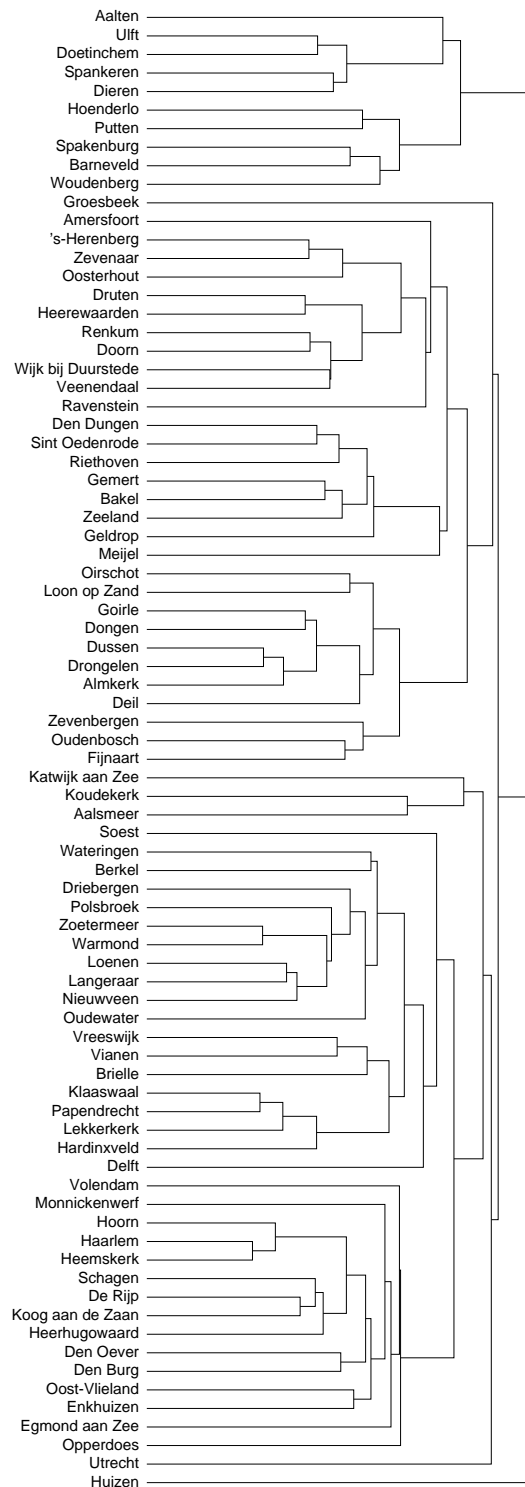


Figure 9.20: Subtree of the dendrogram in Figure 9.5, representing the Central Dutch varieties group. The group encloses varieties in south Gelderland, Noord-Brabant, Zuid-Holland, Utrecht and Noord-Holland.

The group covers the southern parts of the groups 18 and 19 of Daan's map. Unfortunately, the northern border of this group matches exactly the transcriber border (see Figure 9.3). The southern border, however, matches mainly with the border found intermediate between the groups 18 and 19 in the North and the groups 6 and 13 in the South in Daan's map (see Figure 9.2). However, our border runs west of Woudenberg, while in the map of Daan the border runs east of this location. So for this variety, the influence of transcriber differences is noticeable.

When going one level deeper again, we find a cluster Groesbeek ... Fijnaart which corresponds with southern Gelderland en Noord-Brabant, and a cluster Katwijk aan Zee ... Utrecht corresponding with Noord-Holland, Zuid-Holland and the greater part of Utrecht. In the cluster with the south Gelderland and Noord-Brabant varieties, Groesbeek is rather distinct from the other varieties. Geographically, Groesbeek is close to our Limburg group (see Figure 9.17 and Section 9.4.12). The dialect is probably an intermediate variant between the Gelderland varieties and the Limburg varieties. Apart from Groesbeek, we find a cluster Amersfoort ... Meijel, containing varieties in south Gelderland and the eastern part of Noord-Brabant (see Figure 9.21), and a cluster Oirschot ... Fijnaart containing mainly varieties in the western part of Noord-Brabant (see Figure 9.22). The south Gelderland/east Noord-Brabant cluster covers the eastern parts of the groups 6, 13 and 14 in Daan's map. The western part of the southern border matches with the the western part of the border between group 14 (north) and group 17 (south) in the map of Daan. The east Noord-Brabant cluster covers the eastern parts of the groups 13 and 14. The east/west division of the group 6, 13 and 14 in our results is clearly the result of transcriber differences. On the other hand, the east/west division of group 14 is also suggested in the map of Te Winkel (1901).

In the cluster with the varieties in Noord-Holland, Zuid-Holland and Utrecht, we found the dialect of Utrecht to be rather distinct (see Figure 9.22). The dialect of Utrecht is probably a typical town variety, contrasting with the surrounding rural dialects. Going one level deeper, we find the varieties of Katwijk aan Zee, Koudekerk and Aalsmeer as a separate cluster (see Figure 9.22). On the one hand, this is probably the result of transcriber differences (see Figure 9.3). On the other hand, Katwijk aan Zee is known to be an isolated place inhabited by fishermen, where an archaic dialect is spoken (see Hoppenbrouwers and Hoppenbrouwers, 2001, p. 152). In Hoppenbrouwers and Hoppenbrouwers (2001, p. 150) the dialect is even clustered with the Zeeland varieties, while our dendrogram suggests a stronger relation to the Holland and Utrecht varieties.

Going one level deeper again, the cluster Soest ... Delft contains varieties in Zuid-Holland and Utrecht (see Figure 9.22), while the cluster Volendam ... Opperdoes contains mainly varieties in Noord-Holland (see Figure 9.10). The cluster that contains varieties in Zuid-Holland and Utrecht, corresponds with the groups 1 (Zuid-Holland) and 6 (Utrecht) of the map of Daan. However, we found no border between these two groups. Rather we found a north/south

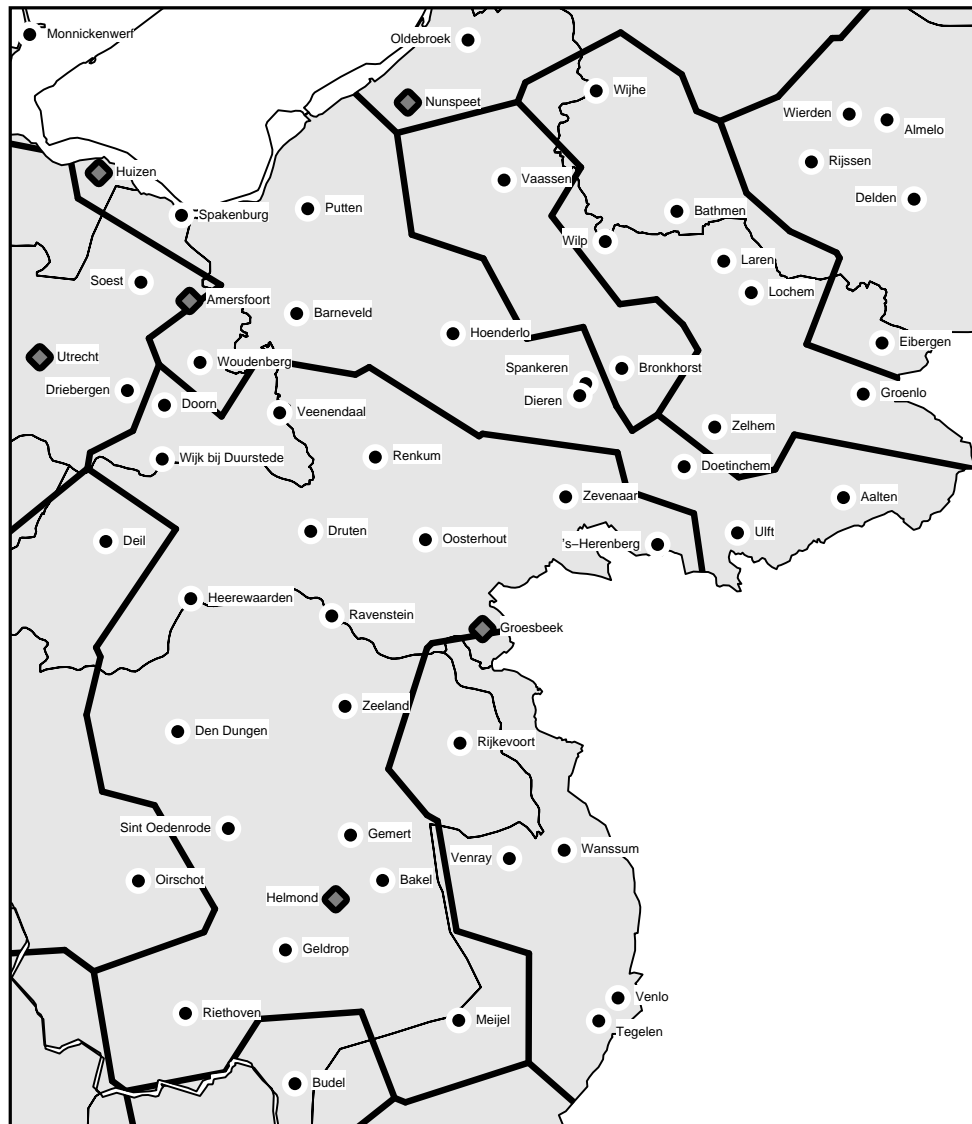


Figure 9.21: The northern half of the southeastern part of the Dutch language area. The state border and provincial borders are represented by thinner lines and dialect cluster borders by thicker ones. The grey diamonds represent varieties that do not actually belong to the group in which they are found geographically.

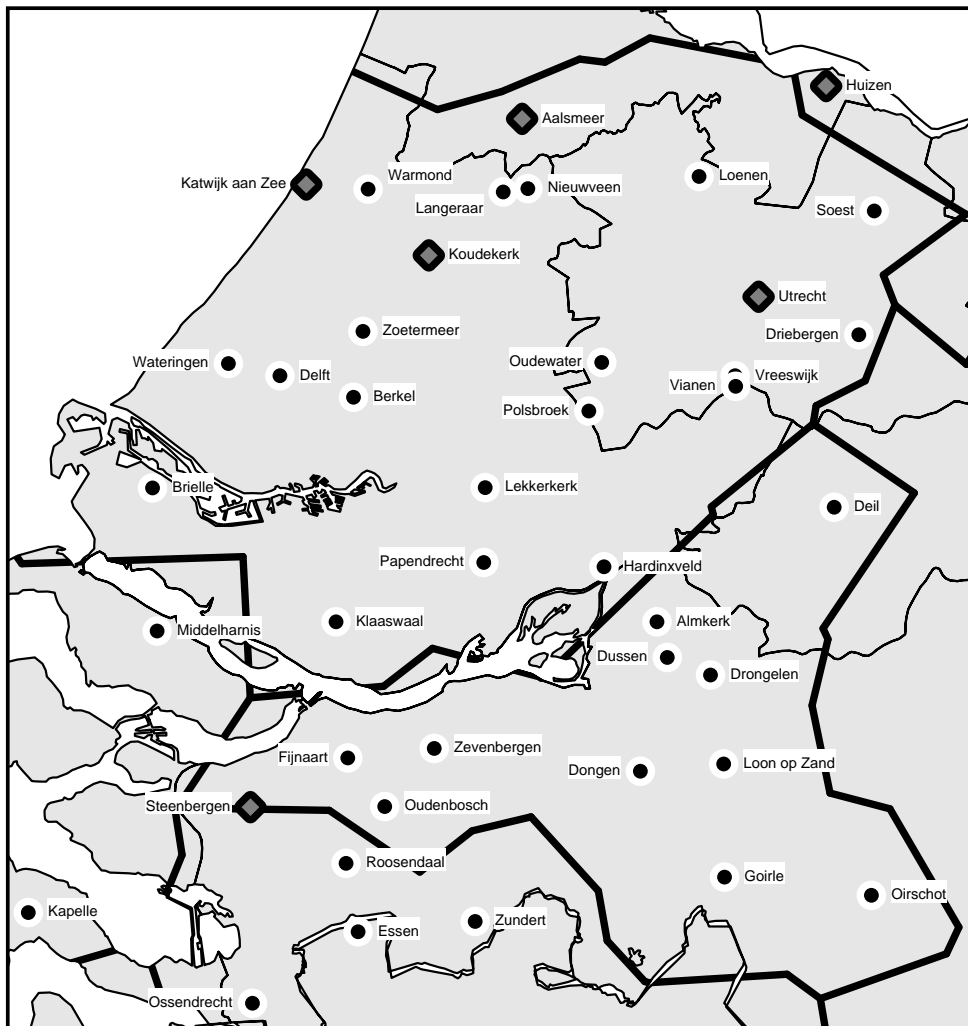


Figure 9.22: The mid-western part of the Dutch language area. The state border and province borders are represented by thinner lines and dialect cluster borders by thicker ones. The grey diamonds represent varieties that, in our measurements, do not actually belong to the group in which they are found geographically.

division across the both provinces. Apart from Delft and Soest, we find a cluster Wateringen . . . Oudewater representing the northern part, and a cluster Vreeswijk . . . Hardinxsveld representing the southern part. Daan and Blok (1969, p. 31) write that it was problematic to find the right border between Zuid-Holland and Utrecht, although they intuitively thought that there must be a border. Finally, they found that the surname *Bartels* was pronounced as [bartəls] in Zuid-Holland, and as [bartəls] in Utrecht. However, the α/a -isogloss could not be found on the basis of the data obtained from the questionnaires, but was finally found on the basis of tape recordings. We doubt whether this Zuid-Holland/Utrecht border still reflects the language awareness of the dialect speakers, all the more since we did not find this border either.

The cluster containing the varieties in Noord-Holland corresponds with the groups 2, 3, 4 and 5 and a small northern part of group 1 in Daan's map. These different groups in the map of Daan cannot be clearly identified in our dendrogram. At a deeper level, we find two close clusters: Hoorn . . . Heemskerk and Schagen . . . Heerhugowaard, but neither the one nor the other corresponds clearly with one of the Noord-Holland groups in Daan's map. Possibly the situation in the time that the recordings were made (1950–1962) has changed compared to the situation at the time that the data for Daan's map was gathered (1939). On the other hand, the dialects of Egmond aan Zee, Volendam and the island Marken (represented by Monnickenwerf in our data set) are known to be independent varieties (Daan, 1956).

We observe that the border between our Zuid-Holland/Utrecht group and the Noord-Holland group exactly matches with the transcriber border (see Figure 9.3). The difference concerns Haarlem only. In our results, this dialect is clustered with the Noord-Holland varieties. In the map of Daan, this variety belongs to group 1 (Zuid-Holland). However, in Daan (1956) the conglomerate of Haarlem is mentioned as a separate group among the varieties of Noord-Holland.

Summarizing, we found five main clusters: the central Gelderland cluster Aalten . . . Woudenberg (see Figure 9.21), the south Gelderland/east Utrecht/east Noord-Brabant cluster Groesbeek . . . Meijel (see Figure 9.21), the west Noord-Brabant cluster Oirschot . . . Fijnaart (see Figure 9.22), the Zuid-Holland/Utrecht cluster Katwijk aan Zee . . . Delft (see Figure 9.22), and the Noord-Holland cluster Hoorn . . . Opperdoes (see Figure 9.10). The southern part of the border between the Zuid-Holland/Utrecht cluster and the west Noord-Brabant cluster matches the border in the map of Daan between group 1 (north) and the groups 7 and 14 (south). The Zuid-Holland/Utrecht cluster and the central-Gelderland cluster just miss bordering on each other. They are separated by the Amersfoort dialect. According to Daan's map, Amersfoort belongs to group 6 and thus was expected to be clustered among the varieties in our Zuid-Holland/Utrecht cluster. However, it is clustered with the south Gelderland/west Utrecht/west Noord-Brabant cluster, probably as the result of transcriber differences.

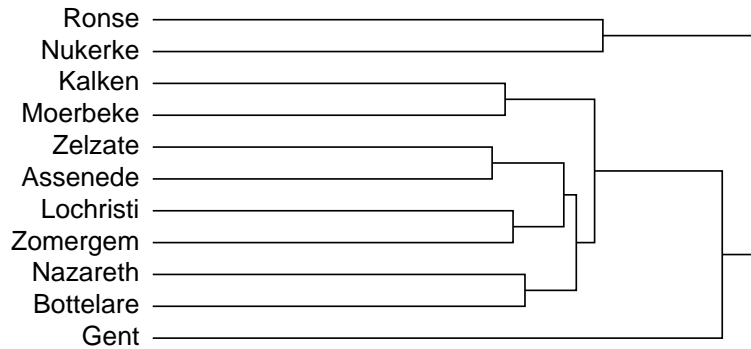


Figure 9.23: Subtree of the dendrogram in Figure 9.5, representing the East Flanders group.

9.4.8 Urk

It is striking that Urk belongs to none of the other groups. This can also be seen in the map in Figure 9.4, where Urk has no strong connection with other varieties. The explanation is found in the fact that Urk was an island in the past, until the Noordoostpolder was impoldered in 1942. In the map of Daan, the variety belongs to group 19. The Urk dialect has a regular vowel system where the duration of vowels is relevant. Rounded and close vowels are used more frequently. For the consonants we note *inter alia* that [sk] is pronounced where [sx] is pronounced in Standard Dutch. More about the dialect of Urk can be found in Daan (1990).

9.4.9 East Flanders

The group of East Flanders varieties corresponds with the groups 10 and 11 in Daan's map. The dendrogram is given in Figure 9.23. The locations of the varieties is shown in the map in Figure 9.24.

In the dendrogram, the cluster with the dialects of Ronse and Nukerke corresponds with group 10 in Daan's map. The group forms a region between the West and East Flanders dialects. The cluster Kalken . . . Gent corresponds with group 11. However, in the northeast in Daan's map the border runs more easterly than it does on our map. As explained in Section 9.4.6, in our results these northeastern varieties are clustered with the varieties that belong to group 12 in the map of Daan (see Figure 9.18). Furthermore, the dendrogram shows that the town dialect of Gent differs rather strongly from the varieties in its surroundings.

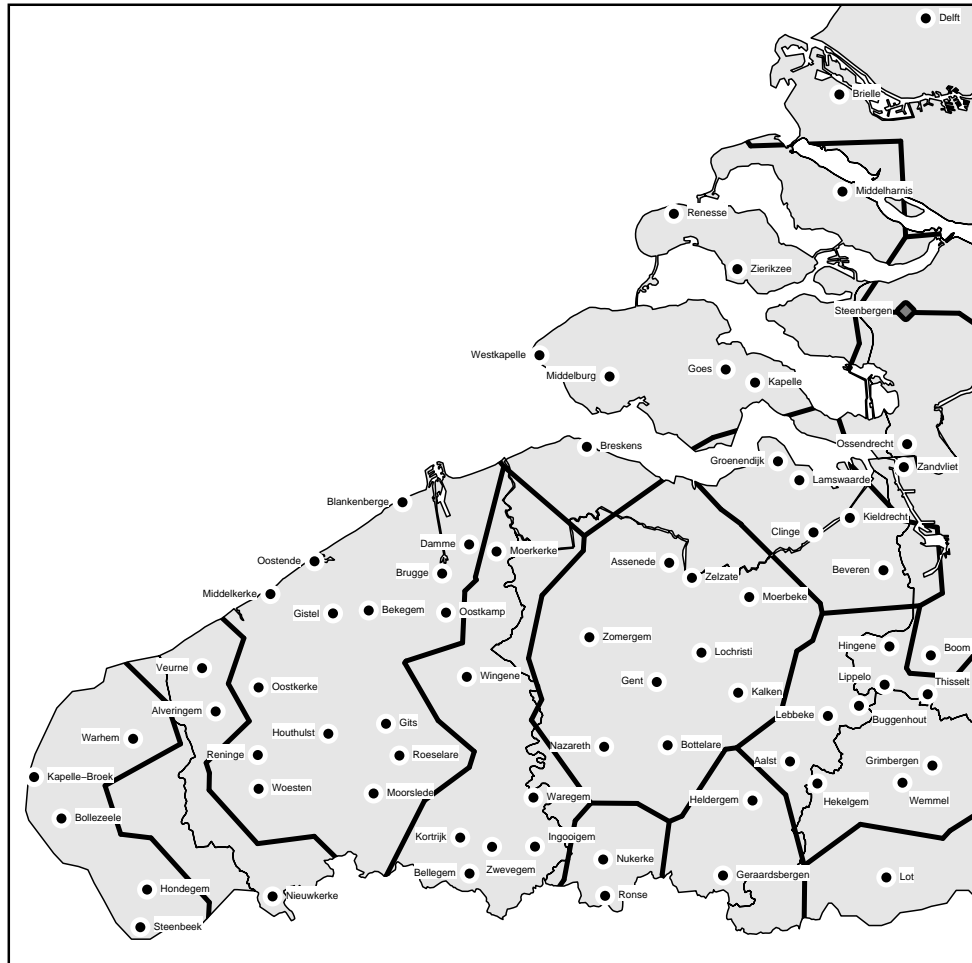


Figure 9.24: The southwestern part of the Dutch language area. National borders and provincial borders are represented by thinner lines and dialect cluster borders by thicker ones. The grey diamond represents a variety that does not actually belong to the group in which it is found geographically.

9.4.10 West Flanders

The varieties of West Flanders correspond with group 9 and 10 in the map of Daan. The dendrogram is given in Figure 9.25, and the locations of the dialects in the map in Figure 9.24.

On the highest level, we find a cluster Moerkerke . . . Waregem containing varieties along the eastern provincial border of West Flanders, and a cluster Nieuwkerke . . . Hondegem containing varieties in the remaining western part. The southern varieties in the eastern group are in or very near group 10 of Daan's map. On our version of Daan's map as given in Figure 9.2 this southern part is visualized. On the original map of Daan, a northern part of group 12 is also found, which is geographically not connected to the southern part. This northern part is not given on our map, since we have no sample sites that fall within this northern part. However, when examining the eastern group in our dendrogram, it is striking to see that it contains varieties both in the North and in the South. Especially the northern varieties of Moerkerke and Wingene are found west of the West Flanders/East Flanders provincial border, while on the map of Daan the northern part of group 10 is only found east of the provincial border. Our results suggest that the Flemish language geographers judge the borders of group 10 too narrowly.

The group of remaining western varieties corresponds with group 9 in Daan's map. A closer division of this area is not given. In our dendrogram the cluster Nieuwkerke . . . Alveringem contains varieties easterly along the French Flanders/West Flanders border. Going one level deeper, the cluster Brugge . . . Houthulst contains the other West Flanders varieties, while the cluster Bollezeele . . . Hondegem contains the varieties in French Flanders. On the original map of Daan the area of French Flanders is shaded, which suggested some contrast to West Flanders. Our dendrogram show that the varieties on both sides of the political border were rather strongly related when the RND recordings were made.

The cluster Nieuwkerke . . . Alveringem appears as a transition zone between the Belgian and the French varieties. At first glance the special position of Veurne and Alveringem may be explained by transcriber differences (see Figure 9.3), but since these varieties form one cluster with Nieuwkerke, we rather think that this cluster represents a transition area.

9.4.11 Zeeland

The Zeeland varieties belong for the greater part to group 7 in the map of Daan. The dendrogram is shown in Figure 9.26. The locations of the varieties can be found in the map in Figure 9.24.

In the dendrogram, Breskens is most distinct from the other varieties. On the map of Daan, this variety belongs to group 9, i.e. the dialect of West Flanders and Zeeuws-Vlaanderen. However, it may be possible that the dialect of Breskens

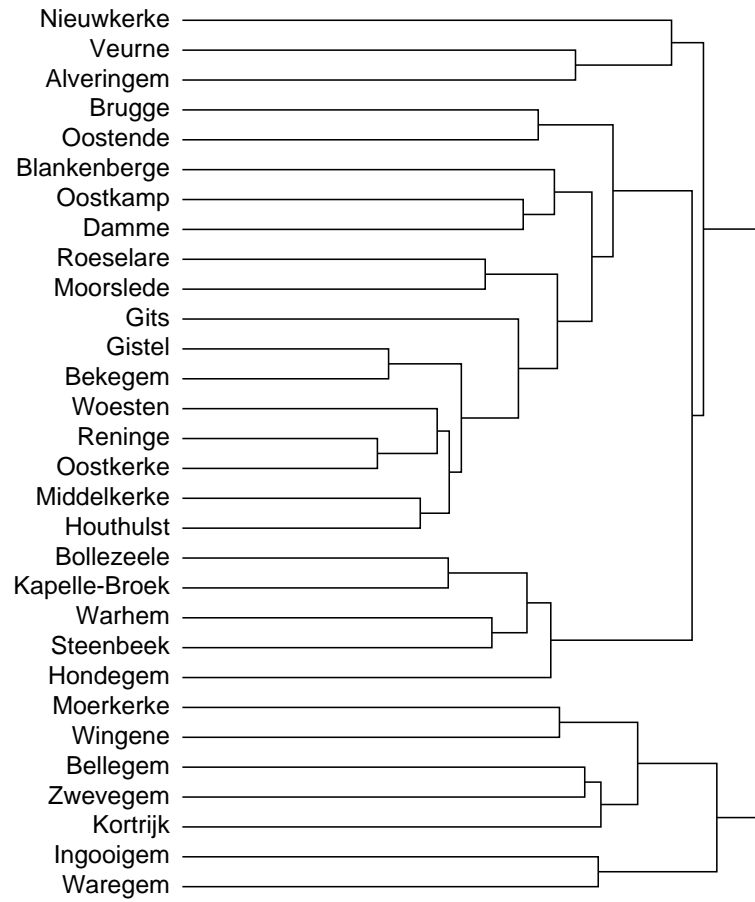


Figure 9.25: Subtree of the dendrogram in Figure 9.5, representing the West Flanders group.



Figure 9.26: Subtree of the dendrogram in Figure 9.5, representing the Zeeland group.

was much more strongly influenced by the Zeeland varieties than other dialects in the same region. This can be explained by the regular ferry lines between Breskens and Vlissingen which started in 1828 and ended only in 2003 when a tunnel connection was completed. Apart from Breskens, the varieties of Goes and Zierikzee form the core in the dendrogram while the other varieties are clustered with them one by one. No clear division per peninsula or island can be found in the dendrogram.

9.4.12 Limburg

The varieties in our Limburg group are found in the western part of group 14 and in the northern and eastern part of group 17 in Daan's map. The division of the Limburg varieties is found in the dendrogram in Figure 9.27. The locations of the varieties are shown in the map in Figure 9.17.

In the dendrogram we first find a cluster that excludes the varieties of Steenberg and Helmond. Steenberg is found in Noord-Brabant and close to Zeeland (see the map in Figure 9.22). In the two dialects, the uvular [ʀ] is mainly used, but the varieties are geographically located among varieties in which the alveolar [r] is used. In the two varieties, the uvular [ʀ] is mainly used. In the direct surroundings of Helmond the alveolar [r] is used, but this dialect is geographically rather close to the Limburg area, where the use of the uvular [ʀ] is common. In the direct surroundings of Steenberg the alveolar [r] is used as well, but also in geographically rather distant varieties the alveolar [r] is still used. However, the pronunciation of the /r/ cannot be a sufficient explanation. The [ʀ] is also used in the varieties of Amersfoort and Ravensbergen, which are located among varieties in which the [r] is used (see Figure 9.20). These two varieties do not deviate so strongly from their geographic neighbours in the dendrogram. Therefore, we cannot explain why the Steenberg dialect is found in the Limburg group, but conjecture that it has to do with migration.

Going one level deeper, the position of Bree is striking. This cannot be explained by transcriber differences with certainty since in that case a stronger relation with Budel was expected (compare Figure 9.3). Apart from Bree, we find a northern cluster Budel . . . Tegelen and a southern cluster Horn . . . Meerssen.

In the northern cluster, we find a western cluster with the varieties of Budel and Overpelt, and a northeastern cluster Rijkevoort . . . Tegelen. Considering the varieties of Budel and Overpelt, we see that Budel is found at the Dutch side of the state boundary, and Overpelt at the Belgian side. The cluster of the two varieties forms the northwestern part of group 17 on the map of Daan. The border between these two varieties and the south Gelderland/east Utrecht/east Noord-Brabant cluster in the group of central Dutch varieties (see Section 9.4.7) corresponds with both the western part of the border between group 13 and 14 in Daan's map and the transcriber border.

It is striking that our northeastern cluster does not belong to group 17 (the Limburg dialects) but to group 14 (dialects of Noord-Brabant and northern Limburg) in Daan's map (see Figure 9.2). In the map of Daan this northeastern cluster is not found as a separate area, it is only a part of group 14. However, in the map of Te Winkel (1901) the eastern part of group 14 is suggested to be a separate dialect area, labeled as 'Saksisch-Oostfrankisch' (Saxon-East Franconian). The southern part of this area corresponds with our northeastern group, although the southern border of our cluster is found more south. Our more southern border coincides with a transcriber border. The western and northern border of the northeastern cluster also coincides with a transcriber border, with the exception of Meijel which is found in the south Gelderland/east Utrecht/east Noord-Brabant cluster in the group of central Dutch varieties. In contrast to the dialects in the dendrogram of the Limburg group, in the dialect of Meijel mainly the alveolar [r] is used rather than the uvular [ʀ]. The western border is also found in the map of Te Winkel (1901).

The southern cluster is a part of group 17 of the map of Daan. The southern part of the southern cluster is bounded on the west side by the state boundary which coincides almost with the Maas river. On the west side of this boundary, our Southwest Limburg group is found (see Section 9.4.5). However, this border is not found on the map of Daan. On the one hand, this is a transcriber border, but on the other hand, this boundary is also found in the map of Te Winkel (1901). As mentioned in Section 9.4.5 the border coincides with an [r]/[ʀ]-isogloss.

9.4.13 Northeast Luik

The northeast Luik group covers the furthest southeastern part of group 17 in Daan's map. The dendrogram is given in Figure 9.28. The locations of the varieties can be found in the map in Figure 9.17. The varieties of Aubel, Baelen, Eupen and Raeren are actually found south of group 17 and belong to the province of Luik. Aubel and Baelen belong to the French language area, and Eupen and Raeren belong to the German language area.

When considering group 17 in the map of Daan it may be unexpected that the varieties in this group do not form one group with the varieties in the Limburg group. However, it is known that the situation in Limburg is complex, and the varieties do not form a homogeneous group (Hoppenbrouwers and Hoppenbrouwers, 2001, p. 187). Most varieties in our Northeast Luik group are found east of the isogloss that represents the opposition between *zeggen* 'to say' (west, the [ɛ] is used) and *sagen* (east, the [a] is used, see the map in Goossens (1977) on p. 21 and 60). In the dendrogram, we find a western cluster 's-Gravenvoeren ... Eupen, and an eastern cluster Raeren ... Kerkrade. The division in these two clusters perfectly reflects the Benratherlinie. This isogloss represents the opposition between Dutch/Low German *maken* 'to make' (west, the [k] is used) and High German *machen* (east, the [x] is used). The dialects west of the Benrather-

linie belong to the East Limburg-Ripuarisch transition zone and the dialects east of the Benratherlinie belong to Ripuarisch, a dialect area around the German city of Cologne (see the maps in Goossens (1977) on p. 21 and 60 again).

9.5 Continuum

9.5.1 Multidimensional scaling

In addition to cluster analysis, we also applied multidimensional scaling to the distances between the 360 RND varieties (see Section 6.2). It appears that one dimension explains 36% of the variance, two dimensions 52%, three dimensions 88%, four dimensions 92%, five dimensions 95%, six dimensions 96% and seven dimensions 97%. These percentages make clear that a rather good representation is already obtained when using three dimensions. When using more dimensions only a rather small improvement of the variance is obtained. Therefore, we examine the three-dimensional solution.

Examining the three dimensions, we found that the first dimension distinguishes between Frisian and Low Saxon varieties on the one hand (low values), and Low Franconian varieties on the other hand (high values). The second dimension distinguishes between Frisian on the one hand (low values), and Low Franconian (high values) and Low Saxon (even higher values) on the other hand. The town Frisian varieties and the dialect of *het Bildt* are intermediate between Frisian and Low Franconian, but they are closer to Low Franconian than to Frisian in this second dimension. The *Stellingwerf* varieties are intermediate between Frisian and Low Saxon, but they are closer to Low Saxon than to Frisian. The third dimension divides the Low Franconian varieties in three groups. The first group (low values) contains the Dutch Limburg varieties. The second group (mean values) contains varieties in Belgian Limburg, Brabant, Antwerpen and the greater part of the Netherlandic Low Franconian area. The third group (high values) contains varieties in East Flanders, West Flanders and also some Zeeland varieties. The third dimension does not divide either the Frisian or the Low Saxon varieties, all of them belong to the second group.

We investigate which phenomena are especially responsible for each dimension. For this purpose we calculate distances between varieties per dimension. When e.g. two varieties have respectively the values x and x' in a dimension, the distance is equal to $|x - x'|$. In this way, for each pair of varieties the distance for one dimension is found. Having 360 varieties, we get $(360 \times 359)/2$ distances. In Section 5.1 we explained how we applied Levenshtein distance using transcriptions. Using Levenshtein distance, a distance matrix can be obtained, containing Levenshtein distances between the different pronunciations of one particular word. This matrix also contains $(360 \times 359)/2$ distances.

When we have calculated multidimensional scaling distances per dimension on the one hand, and Levenshtein distances per word on the other hand, the two sorts of distances can be correlated. The stronger the Levenshtein distances correlate with the distances of one dimension, the more the variation of the corresponding word contributed to the values of that dimension. For each of the 125 words we calculated the Levenshtein distances between the 360 varieties. This results in 125 matrices. Subsequently, each of the matrices was correlated with the distances derived from the first, second and third dimension, respectively.

It appears that the distances in the first dimension correlate strongest with distances obtained on the basis of equivalents for *waren* ‘were’ ($r = 0.70$). The variation of this word is shown in the map in Figure 9.29. In the north we find forms like [va:r̥n̥] or [va:d̥n̥] and in the south forms like [va:r̥ə], [wɔ:r̥ə] or [wa:r̥ə]. In the furthest southwest, we find forms like [wɔ:r̥n̥] or [wɔ:r̥n̥]. So the forms in the north and in the furthest southwest end on [n̥] or [n̥] and the forms in the south on [ə]. In some other strongly correlating words, the same phenomenon was found. We conjecture therefore that the treatment of the weak syllable /ən/ is the single most significant dialect marker in Dutch. The two types of endings corresponds with the division in Frisian and Low Saxon varieties on the one hand (north), and Low Franconian varieties on the other hand (south), as represented by the first dimension. However, the stronger relation between the Low Saxon varieties and the Low Franconian varieties in the furthest southwest was not found in the first dimension, which is possibly the main explanation for the fact that a perfect correlation was not found.

Distances in the second dimension correlate strongest with distances obtained on the basis of equivalents for *vader* ‘father’ ($r = 0.64$). The variation of this word is shown in the map in Figure 9.30. In the northwest we find the Frisian forms like [hɛ:t̥] and [hɑ:t̥], and the town Frisian form [vɑ:ɔdər]. In the remaining part we find forms like [va:dər], [vɔ:dər] and [vɑ:ɔdər]. The two different lexical forms clearly correspond with the division between Frisian and non-Frisian varieties as represented by the second dimension. Other strongly correlating words represent both lexical and phonological differences.

Distances in the third dimension correlate strongest with distances obtained on the basis of equivalents for *breder* ‘broader’ ($r = 0.56$). The variation of this word is shown in the map in Figure 9.31. At first glance the word seems to divide the area in north (using [d]) and south (the [d] is substituted by the [j] or even deleted). However, the correlation with the distances in the first dimension was much lower ($r = 0.27$). Another phenomenon that catches the eye is that in most varieties the alveolar [r] is used in forms like [brɛ:dər], [brɛ:jər] or [bri:ɔr], while the uvular [ʀ] was used in the Limburg varieties (southeast) in forms as [brɛ:ʀər]. From Figure 4.7 we may conclude that the difference between [r] and [ʀ] weight



Figure 9.29: Variation of the equivalents for *waren* ‘waren’. The transcriptions correspond with the labels in Figures 9.32 and 9.33. Extra-short sounds are noted in superscript. Distances among 360 Dutch varieties as found on the basis of the first dimension of a three-dimensional MDS solution correlate most strongly with distances obtained on the basis of pronunciations of this word ($r = 0.70$).

more heavily than the difference between [d] and [j]. The [r]/[R] distinction divides only the southern part of the Dutch language area into two parts: an eastern and a western part. This accords with the third dimension that represents an east-west dimension as well. However, the third dimension distinguishes between an east, central and west group, while the [r]/[R] difference distinguishes only between an east and west area. This difference can be explained by the fact that the third dimension is based not only on the [r]/[R] difference, but rather on the aggregate of several phenomena. Nonetheless, the [r]/[R] difference is one of the most important phenomena as appears from the strong correlation. In other strongly correlating words the same phenomenon was found as well.

9.5.2 Continuum map

As described in Section 6.2.4 on the basis of the three dimensions of the three-dimensional multidimensional scaling solution, each variety can be represented by a color. The three dimensions determine the intensities of red, green and blue. We used this approach to create a map in which each variety gets its own unique color. We assign the colors to the three dimensions so that the color scheme of our map approaches the color scheme in Daan's map maximally. The first dimension represents the intensity of red, the second dimension inversely the intensity of blue, and the third dimension inversely the intensity of green.

In Figure 9.32 a color map based on three MDS dimensions is shown. On this map dialect points are blown up to small areas until they border each other (see Section 6.2.4). However, dialect islands are not blown up, but represented by diamonds, just as in the map in Figure 9.6. For an explanation about the dialect islands see Section 9.3. To keep a clear picture, the same restricted set of labels of (in general) better-known locations is printed in the map as in the map in Figure 9.6.

On the map, the Frisian area (northwest) is represented by bright blue, the Low Saxon area (northeast) is green. The Netherlandic Low Franconian area is colored by different grey shades. Note that the town Frisian varieties (most diamonds and the island Ameland north of Leeuwarden) have a color intermediate between grey and blue. The same color is found for *het Bildt* (northwest of Leeuwarden). The Frisian mixed varieties of Tjalleberd (the higher diamond south of Grouw) and Donkerbroek (the diamond intermediate between Grouw and Assen) are more greenish. The 'pure' Frisian variety of Appelscha (west of Assen) is colored blue, just as are the other 'pure' Frisian varieties. Turning now to the southern part of the Low Franconian area, the West Flanders varieties (in the furthest southwest) are colored darker red, the East Flanders varieties

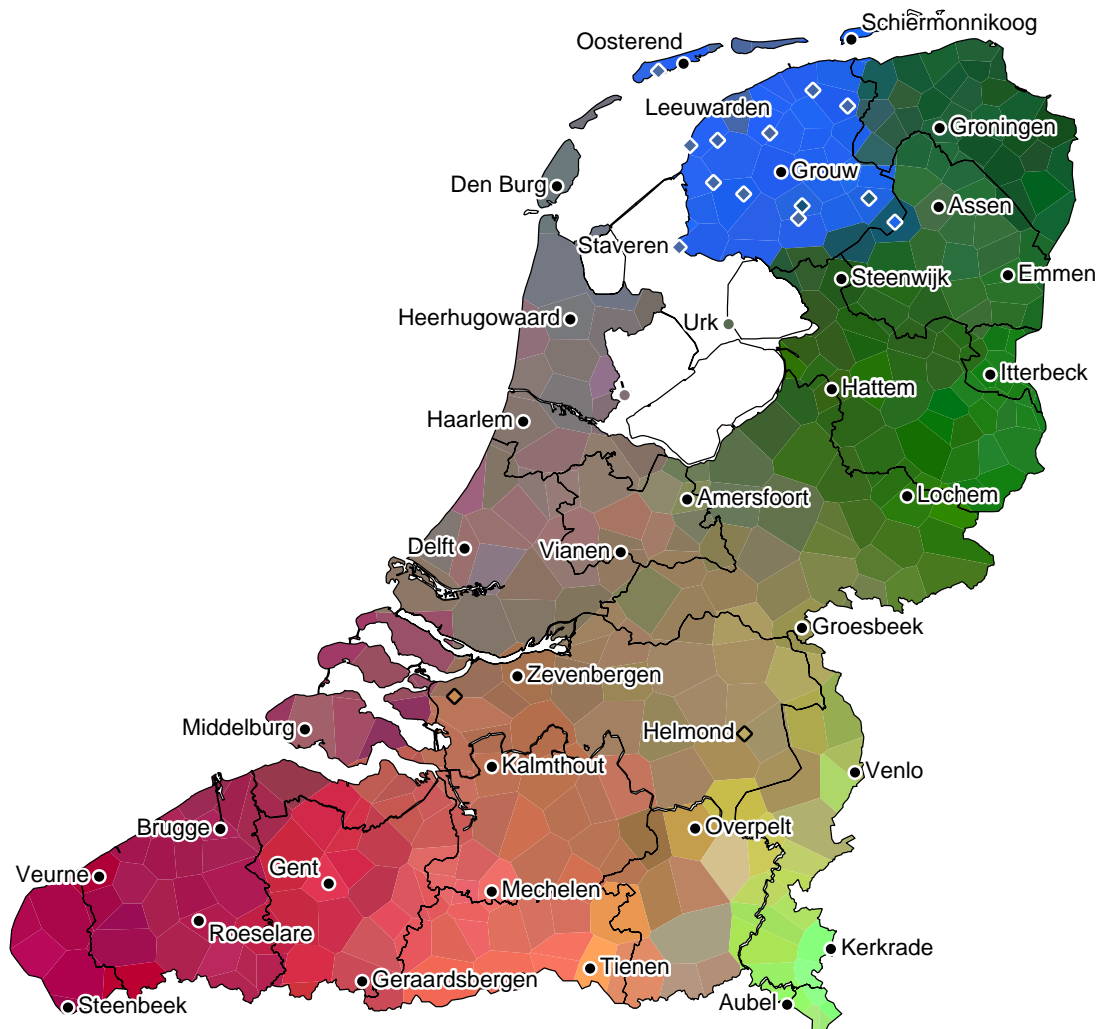


Figure 9.32: Dialect variation represented by color. The first MDS dimension is mapped to red, the second inversely to blue and the third inversely to green. Kruskal's Non-metric MDS is used (see Section 6.2.2). Each of the dialect points is blown up to a small area, except the dialect islands that are marked with diamonds.

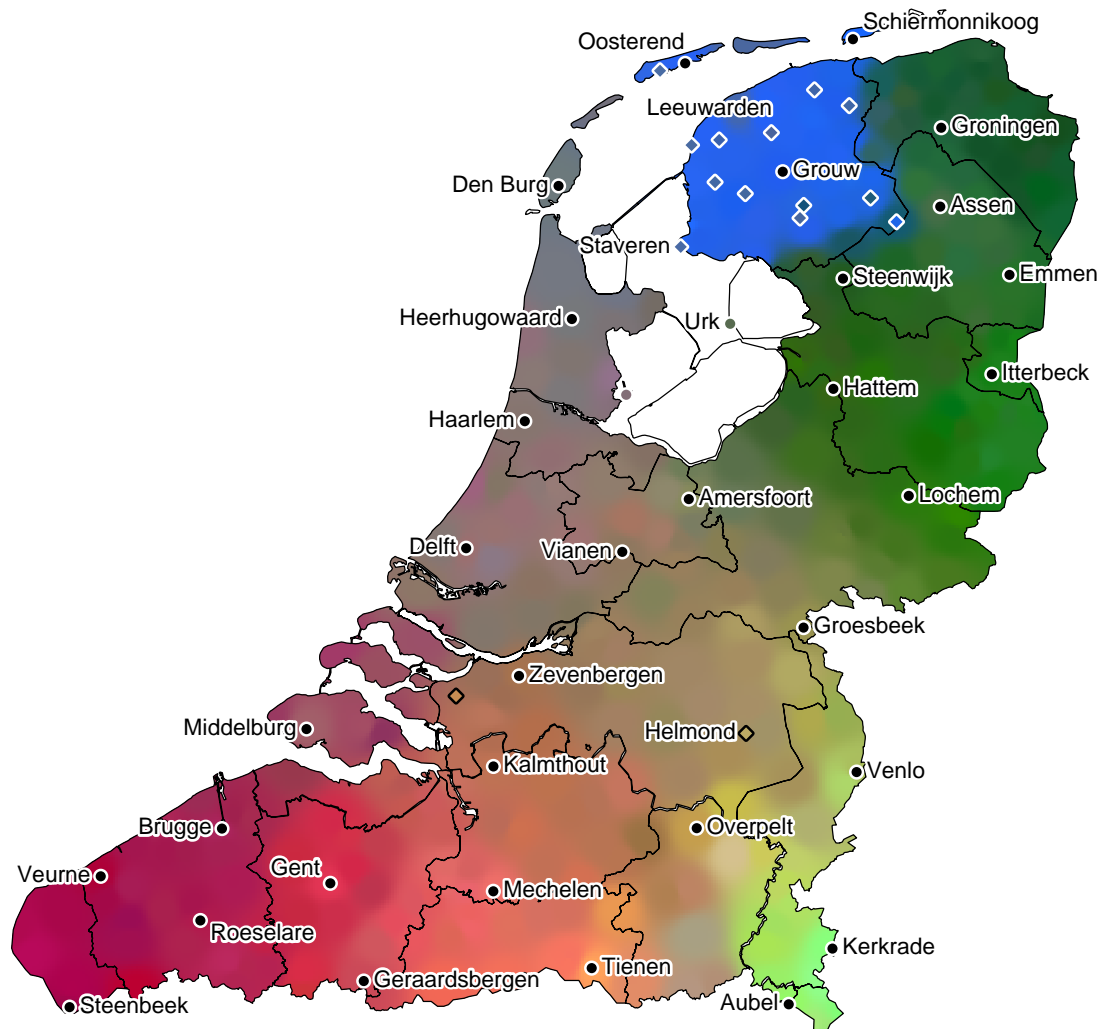


Figure 9.33: Dialect variation represented by color. The first MDS dimension is mapped to red, the second inversely to blue and the third inversely to green. Kruskal's Non-metric MDS is used (see Section 6.2.2). The color of intermediate points is determined by interpolation using Inverse Distance Weighting. Dialect islands are marked with a diamond. They are not involved in the interpolation process.

(around Gent) bright red. The Antwerpen area (around and south of Kalmthout) is greyish red, the Belgian Brabant area lighter red. The varieties in the southern part of Limburg are colored lighter green. In the map the colors of the diamonds of Steenberg (west of Zevenbergen) and Helmond are a little bit more like the colors of the Limburg varieties. However, they do not contrast as strongly with their surroundings as might be expected on the basis of the findings in Section 9.4.12. Although results of cluster analysis and multidimensional scaling are in accordance with each other in general, this example shows that minor differences can be found. Therefore, it is useful to show results obtained with both techniques. Broadly speaking our color map is similar to the map of Daan.

In Figure 9.33 a map is given that is related to the map in Figure 9.32. However, in this map the space between the points is colored on the basis of MDS values which are found by interpolation using Inverse Distance Weighting (see Section 6.2.4 again). This map consequently does justice to the idea that the dialect landscape may be regarded as a continuum. Dialect islands are of course not involved in the interpolation process.

9.6 Relation to Standard Dutch

In the same way as distances are calculated between dialects, distances between dialects and a standard language can be calculated. For the RND material we calculated distances with respect to Standard Dutch. The results are shown in Figure 9.34. In the map the dialects are colored according to the rainbow. The most similar dialects are red, followed by orange, yellow, green and lighter blue. The dark blue dialects are most distant.

Exactly in accordance with the general opinion, Haarlem is closest to Standard Dutch. The dialect has a distance of 14.7%. The way in which percentages are found is explained in the Sections 5.1.8 and 5.1.10. Haarlem was followed by Brielle (16.8%, south of Delft), Hoorn (16.9%, east of Heerhugowaard), Warmond (17.0%, south of Haarlem), Heemskerk (17.2%, north of Haarlem) and Vianen (17.5%).

Most distant was the dialect on the island Schiermonnikoog, north of the border between Friesland and Groningen. The variety has a distance of 44.9%. Although the Schiermonnikoog variety is Frisian, we found in Section 9.4.1 that it has a distinct position among the other Frisian varieties. However, Schiermonnikoog is followed by the other ('pure') Frisian varieties. The 'pure' Frisian variety most like Standard Dutch was West-Terschelling (40.9%, west of Oosterend). This may be explained by the fact that the island Terschelling belonged to the province of Noord-Holland until 1942. The high percentages for the Frisian varieties may justify the fact that Frisian is recognized as a second official language

in the Netherlands. The dialects of *het Bildt*, the Frisian cities, Midsland, and Ameland Island are clearly less distant from Standard Dutch. Most distant was Franeker (33.0%, west of Leeuwarden), most similar was Hollum (29.5%, on the island Ameland north of Leeuwarden).

Besides Frisian, four Limburg dialects were also very distant to Standard Dutch: Vaals (43.2%, south of Kerkrade), Kerkrade (42.2%), Raeren (41.8%, south of Vaals) and Aubel (41.6%). Also in the West-Flemish dialect area we found distant dialects: Alveringem (42.3%, south of Veurne), Warhem (41.0%, southwest of Veurne), Reninge (40.9%, southeast of Alveringem) and Veurne (40.7%). In the province of Overijssel, we found the dialects of Vriezenveen (41.8%), Rijssen (40.8%) and Wierden (40.4%, the three dialects are found southwest of Itterbeck) to be rather distant. The special position of Vriezenveen was already found in Section 9.4.4. Although the Groningen dialects belong to the more distant dialects (indicated by lighter blue), they are not as distinct as for example the Frisian varieties and the four Limburg varieties that we mentioned above. Most distant were Finsterwolde (41.0%, east of the city of Groningen) and Onstwedde (40.4%, east of Assen).

In Hoppenbrouwers and Hoppenbrouwers (2001) the authors also discuss a ranking with respect to Standard Dutch (pp. 124–131). Distances are obtained by means of the feature frequency method (see Section 2.3.2). The scores in the rank order list are divided into 23 intervals. Frisian is found in interval 5 and 6, which suggest that Frisian is rather close to Standard Dutch. Town Frisian varieties are not obviously closer to Standard Dutch than ‘pure’ Frisian varieties, both are found in interval 5 and 6. This is clearly different from our results, where the ‘pure’ Frisian varieties form a group of the most distant varieties, while the town Frisian varieties are obviously more related to Standard Dutch (see Figure 9.34: darker blue versus green). In the results of Hoppenbrouwers and Hoppenbrouwers the Groningen dialects are found in the intervals 13, 15 and 16, suggesting that they are much more distant than the Frisian varieties. In our results, both, Frisian and Groningen dialects are rather distant, but Frisian is more distant (Figure 9.34: darker blue versus lighter blue). Our results agree with those of Hoppenbrouwers and Hoppenbrouwers that the dialect of Kerkrade and neighboring varieties belong to the most distant ones.

The comparison of our results with those of Hoppenbrouwers and Hoppenbrouwers shows that the use of Levenshtein distance gives different results than the feature frequency method. Especially when looking at the Frisian dialects, our results are much more in accordance with the prevailing opinion and especially the opinion of the Frisians themselves. Using the Levenshtein distance, words are regarded as linguistic units while the order of segments in a word is also considered. The example of the Frisian underscores the importance of these two aspects.

9.7 Conclusions

We calculated distances between 360 Dutch varieties. Data was taken from the RND. We compared the results with Daan's map (see Figure 9.2). However, since the RND was compiled by 16 different transcribers, we also kept track of transcriber borders (see Figure 9.3).

Comparing our division of the main 13 dialect groups with the map of Daan, we found similarities and differences. We found that especially our Groningen and Overijssel group and our group of central Dutch varieties were divided into different groups in the map of Daan. The opposite result was also found. The Limburg group in the map of Daan was divided into three different groups on our map. We explain the differences by the different weighting of dialect contrasts by dialect speakers, whose judgements formed the basis of Daan's divisions. Social and economic structures may influence judgements of dialect similarity. Furthermore, in a homogeneous area differences can be more easily identified than in a heterogeneous area. This may explain why e.g., the Overijssel area is divided into several groups in the map of Daan while the Limburg area is found to be one group.

Both when examining the 13 groups and when examining each of the groups in more detail, we find real dialect borders and transcriber borders. Sometimes, dialect borders and transcriber borders nearly coincide. For these borders no obvious conclusion can be drawn. Although transcriber differences were normalized to some extent, we did not succeed in eliminating them fully, as appears in the results. Normalizing transcriber differences is risky, since it is not always clear whether differences in notation reflect transcriber differences or dialect differences. Advisable future work would be to calculate distances between Dutch varieties again on the basis of data from the *Fonologische Atlas Nederlandse Dialecten* (FAND) (Goeman and Taeldeman, 1996; Goossens et al., 1998, 2000). This material is known to be of higher quality.

We also calculated distances with respect to Standard Dutch. Just as in Hoppenbrouwers and Hoppenbrouwers (2001) we found the dialect of Haarlem to be most similar to Standard Dutch. In our results, the Frisian varieties were most distant, while in the results of Hoppenbrouwers and Hoppenbrouwers (2001) the Frisian varieties were relatively related to Standard Dutch. We think that our results are most in accordance with linguistic reality and with general opinion. This difference shows that it is important to regard words as linguistic units and to consider the order of segments in word pronunciation. With the Levenshtein distance used in our research these two aspects are taken into account.

