Determining Dutch dialect phylogeny using bayesian inference

Peter Dekker
Bachelor's thesis (7.5 ECTS)
BSc Artificial Intelligence, Universiteit Utrecht

Supervisors: Alexis Dimitriadis and Martin Everaert

July 31, 2014

Abstract

In this thesis, bayesian inference is used to determine the phylogeny of Dutch dialects. Bayesian inference is a computational method that can be used to calculate which phylogenetic tree has the highest probability, given the data. Dialect data from the Reeks Nederlandse Dialectatlassen, a corpus of words in several Dutch dialects, serves as input for the bayesian algorithm. The data was aligned and converted to phonological features. The trees generated by bayesian inference were evaluated by comparing them with an existing dialect map by Daan and Blok.

Contents

1 Introduction

How are languages related? Languages are genetically related if they share a single ancestor from which they derive (Campbell, 1998). To prove a common ancestor, an array of methods can be applied. The phylogeny, or evolutionary relationship, of languages can be viewed as a tree, where a branching shows that two languages B and C derived from an ancestor A.

1.1 Bayesian inference

In recent years, computational methods have seen their advent in historical linguistics. One of them is bayesian phylogenetic inference. This method is inspired by Bayes' law: the probability of a hypothesis H_x for a certain phenomenon E can be given using the probability of the phenomenon given the hypothesis.

$$P(H_x|E) = \frac{P(H_x) \cdot P(E|Hx)}{\sum_{k=1}^{n} P(H_k) \cdot P(E|H_k)}$$

In our case, a phylogenetic tree is the hypothesis. A tree is a tuple $\omega = (\tau, \upsilon, \phi)$ with (Larget and Simon, 1999):

- a tree topology τ
- a vector of branch lengths υ associated with topology τ. Each branch length in υ represents the distance between two adjacent nodes in τ.
- a substitution model ϕ, which determines the probability that a certain element of a language changes into a certain other element.

The tree topology τ is defined as:

- a set of vertices V
- a set of edges $E \in V \times V$
- the graph that E describes on V is strongly connected
- there are no cycles.

Branch lengths show how much distance there is between languages: a longer branch means that more substitutions have been made between the input strings. The substitution model determines how likely the change from a character in the input string to a certain other character is.

The question is: what is the most probable tree to describe the linguistic data X? Our application of Bayes' law becomes (Ronquist and Huelsenbeck, 2003):

$$f(\omega|X) = \frac{f(\omega) \cdot f(X|\omega)}{f(X)}$$

$f(\omega)$ is the prior distribution, containing the a priori probabilities of the different trees. $f(X|\omega)$ is the likelihood function, which returns the probability that the data has been generated by a tree. $f(X)$ is the total probability of the data. $f(\omega|X)$ is the posterior distribution, containing the probabilities of all the trees.

Assumptions have to be made about the prior probabilities of trees, because they are generally unknown. Calculating the posterior probability means summing over all of the trees, whose posterior probabilities have not been calculated yet, and integrating over all the possible combinations of τ, υ and ϕ. The posterior probability distribution cannot be calculated directly (Huelsenbeck et al., 2001).

To address this issue, bayesian algorithms use a technique called Markov Chain Monte Carlo (MCMC) sampling. MCMC is an approximation of Bayes' law, with a number of simplifications. The prior probability distribution is determined by a Dirichlet distribution, inferring the prior probability of a tree from the data itself (Ronquist et al., 2011). The algorithm starts off with a random tree. Every generation, a small random change of the parameter values is proposed. It is accepted or rejected with a probability given by the Metropolis-Hastings algorithm (Larget and Simon, 1999). Every s generations (where s is the sample frequency), the accepted tree of the current generation is saved to the posterior sample. After a number of generations, the posterior sample should approximate the real posterior probability distribution (Huelsenbeck et al., 2001). From this distribution, a best tree can be drawn, according to the desired criteria.

In principle, the input for the bayesian method could be any kind of linguistic data. When applied to languages, it is common to use Swadesh lists, a list of words which are unlikely to be borrowed. A linguist manually classifies each word in a dialect to be in a certain cognate class. Crucial in the cognate classification is the *Neogrammarian hypothesis*, which states that sound changes are regular. Regularity means that if a sound in a word changed into another sound, it will do so in every other word. Two words can only be cognate, if the common ancestor can be reached from both cognate candidates by a number of known sound changes (Campbell, 1998). Generally, in phylogenetic linguistics, a string of the cognate classifications of every word in a language serves as the input for the bayesian algorithm. For dialects, words in the different dialects are likely to be cognates (Dunn, 2008). If cognacy is assumed, manual classification can be omitted and an objective measure of distance between the phonological forms in different dialects can be used.

1.2 Earlier research

In earlier research, the bayesian method has been applied to Bulgarian dialects (Prokić et al., 2011). The focus in the research was on vowel change. The consonants were dropped and the vowels were classified into a limited number of classes to reduce computational cost. A broader approach will be used in this thesis, because consonants may also amount to important distinctions between dialect groups.

1.3 Applying bayesian inference to the Dutch dialects

I would like to evaluate the bayesian method when used on the Dutch dialects. My research question is: how well is bayesian inference suited to determine the phylogeny of Dutch dialects?

2 Method

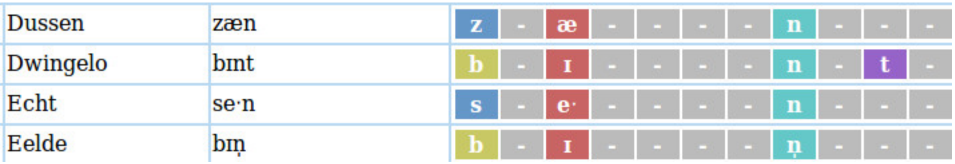
The input for the bayesian inference is a string for every dialect, which uniquely describes that dialect. A corpus of words and their translations into different dialects were used as the basis for the input string. Each word was aligned with its counterparts in different dialects. All the aligned words for a dialect were concatenated. The concatenated strings were converted to phonological features. The resulting feature strings were used as input for the bayesian inference.

2.1 Data

The dialect data was taken from the Reeks Nederlandse Dialectatlassen (RND). This is a corpus of transcribed speech in Dutch and Frisian dialects in the whole Dutch language area: The Netherlands, a neighbouring area in Germany, Flanders and the north of France. The corpus was recorded between 1925 and 1982. A selection of 166 words and 363 dialects has been made from this corpus and digititalized (Heeringa, 2001). An interesting addition to the digital version is the Plautdietsch dialect from Protasovo, Siberia. This dialect descended from 16th century Mennonites who migrated via Eastern Europe to Siberia. It maintained its Dutch character in Slavic surroundings (Nieuweboer, 1998).

The data was written in X-SAMPA, an ASCII version of the International Phonetic Alphabet (IPA). The data was converted to IPA, represented in Unicode, using the cxs2ipa script (Theiling, 2008)[1]. The dialect data was split into two subsets: one set of 269 Netherlandic (and neighbouring German) dialects and one set of 94 Belgian (and neighbouring French) dialects. It is interesting to use the parameters from the Netherlandic data set on the Belgian data, to see whether the setting of the parameters is generally applicable to different data sets.

2.2 Alignment



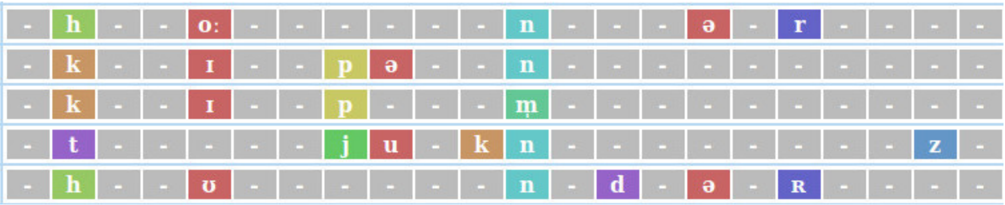
Dussen	zæn	z	-	æ	-	-	-	-	n	-	-	-
Dwingelo	bɪnt	b	-	ɪ	-	-	-	-	n	-	t	-
Echt	se·n	s	-	e·	-	-	-	-	n	-	-	-
Eelde	bɪn̩	b	-	ɪ	-	-	-	-	n̩	-	-	-

Figure 1: Alignment of translations of the lemma *are*. The sound classes (colors of the phones) enable comparable sounds to be matched.

In order to compare which sounds differ in the translation of a lemma in different dialects, the words need to be aligned. Comparable sounds are put in the same

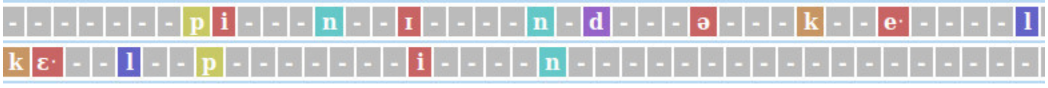
[1] The script was modified in order to convert the æ properly as well.

column (Figure 1). Making an alignment assumes that the words in different dialects are cognates. For most, but not all, words in the RND, this is the case. For example, the lemma *chickens* has entries which look like Dutch *kippen* and entries that look like German *Hühner*. Aligning these with each other is less informative (Figure 2). For some dialects, there was more than one translation for a certain lemma. In these cases, the first one was chosen as the only translation. There are also lemmas where the alignment may have been distracted by morphological rather than phonological differences. For example, the lemma *sore throat* has items which look like *keelpijn* (throat-sore) and other entries which look like *pijnindekeel* (sore-in-the-throat). The sounds are roughly the same, only the order of stems is different. This is not fully reflected in the alignment (Figure 3).



-	h	-	-	oː	-	-	-	-	-	-	n	-	-	-	ə	-	r	-	-	-	-
-	k	-	-	ɪ	-	-	p	ə	-	-	n	-	-	-	-	-	-	-	-	-	-
-	k	-	-	ɪ	-	-	p	-	-	-	m̩	-	-	-	-	-	-	-	-	-	-
-	t	-	-	-	-	-	j	u	-	k	n	-	-	-	-	-	-	-	-	z	-
-	h	-	-	ʊ	-	-	-	-	-	-	n	-	d	-	ə	-	ʀ	-	-	-	-

Figure 2: Alignment of translations of the lemma *chickens*. The entries are not cognate, there are entries which look like *kippen* and entries that look like *Hühner*. Still, they are aligned based on their phonological characteristics.



-	-	-	-	-	-	-	p	i	-	-	-	n	-	-	ɪ	-	-	-	-	n	-	d	-	-	-	ə	-	-	-	k	-	-	e·	-	-	-	-	l
k	ɛ·	-	-	l	-	-	p	-	-	-	-	-	-	-	i	-	-	-	-	n	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

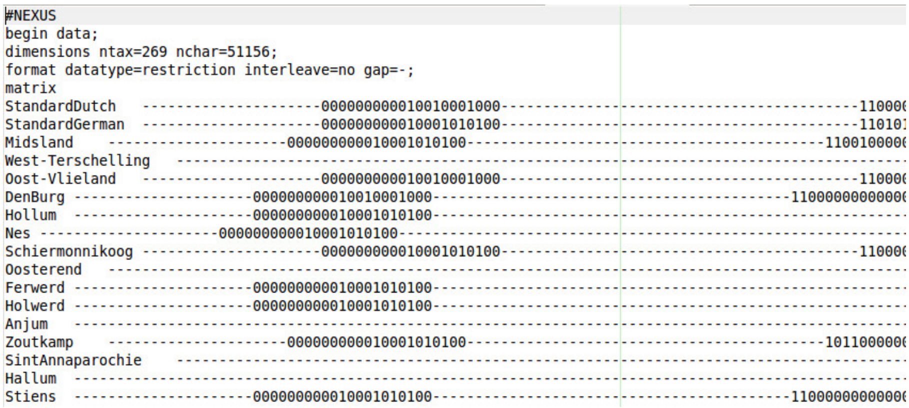
Figure 3: Alignment of translations of the lemma *sore throat*. The phonological similarity is not fully reflected in the alignment, because of the morphological difference in order. The part *pin* in both words is however aligned.

The alignment is done using LingPy (List and Moran, 2013). This is a program for multiple sequence alignment, which means that all words are aligned with each other at the same time. LingPy matches phones, by classifying them into a number of sound classes (List, 2012). Phones in the same sound class have the highest probability of matching.

Before the alignment, the data was tokenized: phones were grouped with diacritic signs to form one token. The list of possible tokens was based on Hoppenbrouwers and Hoppenbrouwers (1988). It supplies a list of IPA tokens and maps those to phonological features. The list is based on the RND data. Still, some combinations of vowels/consonants and diacritics that were used in my RND data, were missing in the list. These tokens were omitted, because this means there is no phonological mapping for these tokens as well. The result is that the omitted diacritic signs are shown as a *?* in the alignment, which means it can be aligned with a random phone. It seems this has not decreased the quality of the alignment heavily.

Standard Dutch	k	ɪ	p	-	ə	m	ɛ	i	n	b	l	u	m	ə	-
Standard German	h	y	-	n	ə	m	ɑ	i	n	b	l	uː	m	ə	n
Midsland	h	ɛ	-	n̩ː	-	m	i	-	-	b	l	u	m	ə	n

Figure 4: Concatenation of the lemmas *chickens*, *my* and *flowers* for three dialects. The real concatenated strings are far longer, they contain 166 lemmas.



```
#NEXUS
begin data;
dimensions ntax=269 nchar=51156;
format datatype=restriction interleave=no gap=-;
matrix
StandardDutch   ---------------------000000000010010001000------------------------------------------11000
StandardGerman  ---------------------000000000010001010100------------------------------------------11010
Midsland    ---------------------000000000010001010100------------------------------------------1100100000
West-Terschelling   --------------------------------------------------------------------------------------
Oost-Vlieland   ---------------------000000000010010001000------------------------------------------11000
DenBurg ---------------------000000000010010001000------------------------------------------11000000000000
Hollum  ---------------------000000000010001010100--------------------------------------------------------
Nes ---------------------000000000010001010100------------------------------------------------------------
Schiermonnikoog ---------------------000000000010001010100------------------------------------------11000
Oosterend   ----------------------------------------------------------------------------------------------
Ferwerd ---------------------000000000010001010100--------------------------------------------------------
Holwerd ---------------------000000000010001010100--------------------------------------------------------
Anjum   --------------------------------------------------------------------------------------------------
Zoutkamp    ---------------------000000000010001010100------------------------------------------1011000000
SintAnnaparochie    --------------------------------------------------------------------------------------
Hallum  --------------------------------------------------------------------------------------------------
Stiens  ---------------------000000000010001010100------------------------------------------11000000000000
```

Figure 5: The feature strings which serve as input for the bayesian inference. This is the result of converting the concatenations from figure 4 to phonological features.

After the alignment, all aligned words for a dialect were concatenated with each other, resulting in a long string of all words for that dialect (Figure 4).

2.3 Phonological mapping

A possiblity would be to directly use the concatenated string of all aligned words as the input string for a dialect. The positions in the alignment, the phones, would then be the features, on the basis of which a dialect can be compared with other dialects. However, the symbol alphabet of all phones used in the RND is too big to be computationally feasible. Furthermore, it would be nice if the algorithm also takes into account that phones that are phonologically close to each other can change more easily than phones that are further from each other. For these two reasons, the aligned phone strings were converted to an array of phonological features. Hoppenbrouwers and Hoppenbrouwers (1988) provide a mapping from each character to 21 binary phonological features, which was used. The result is a long string of 0's and 1's for every dialect (Figure 5).

2.4 Bayesian inference

The program used to execute the bayesian inference was MrBayes (Ronquist and Huelsenbeck, 2003). As described in the introduction, an MCMC analysis starts from a randomly chosen tree and proposes small random changes to this tree. MrBayes runs two different MCMC analyses at the same time, starting from two different randomly chosen trees. By calculating the convergence between the two analyses, it is possible to get an indication whether a stable posterior probability distribution has been reached. The algorithm was run for 1,000,000 generations. The sample frequency was set to 20, which means that every 20 generations, the most probable tree is saved.

The likelihood is determined by two parameters: the substitution model and the rate variation model. Together they provide the probability of the data, given a certain tree. The substitution model determines the chance that a certain character changes into another character. We have only two characters (0 and 1), so the only state changes are $0 \rightarrow 1$ and $1 \rightarrow 0$. A substitution model with equal probability for every state change was used. The rate variation model determines the chance that a certain feature changes state. For example, a rate variation model could state that letters at the end of a word have a higher chance of changing than letters in the middle (Prokić et al., 2011).

Two rate variation models were tried: an equal rate variation model and a gamma-distributed rate variation model. In an equal rate variation model, every feature has the same chance of changing. In a gamma-distributed rate variation model, the bayesian algorithm infers from the data which features change more often than others. It categorizes the features in rate classes of higher or lower probability of change, according to a gamma distribution.

2.5 Evaluation

A dialect map by Daan and Blok (1969) is used as the gold standard to evaluate the results of the bayesian analysis. The map is based on the perception of speakers. In a questionnaire, people from villages in the Dutch language area were asked which dialects from other villages were (almost) the same. Arrows could be drawn between villages with roughly the same dialect. Daan and Blok's map is based on this arrow method, combined with some linguistic knowledge, in cases where the arrow method did not match the known insights (Daan and Blok, 1969).

Finally, the bayesian algorithm with the same settings was applied to the Belgian dialects. The results were also compared with Daan and Blok's map.

3 Results

The output of the bayesian inference is a set of trees, each with their own probability. The consensus tree is a tree that tries to reconcile all trees. If the branching is contradictory between trees, the consensus tree places the branch at a lower level (Dunn, 2008). The trees were shown graphically using the FigTree (Rambaut, 2013) program.

3.1 Netherlandic dialects

3.1.1 Equal rate variation model

The consensus tree that was outputted correctly shows groups of dialects that are connected locally, but does not generally show higher-order grouping between the local groups. This probably happens because the consensus tree could not decide between two specific branchings and places dialects at a lower level. The MCMC analysis has also not converged optimally, even after 1,000,000 generations.

Hardly any false groupings are made. Dialects that are grouped in the tree, are generally also in one group on Daan's map. Sometimes dialects are linked with a dialect that is just across the border of a different group on the map, but still geographically close. This is visible in the grouping of the dialects of Zeeuws-Vlaanderen with some neighbouring dialects from Noord-Brabant (Figure 6). No strange groupings between different parts of the country are made. The price for this accuracy is that a lot of dialects remain ungrouped or are only connected with their direct neighbours (Figure 7).

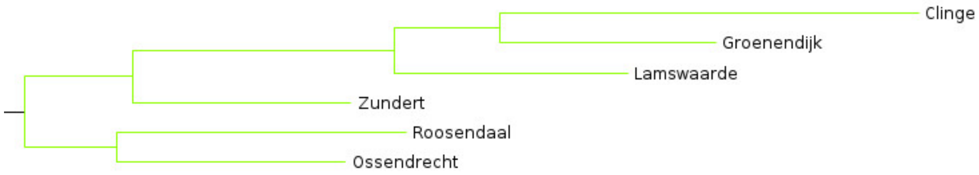


Figure 6: Equal model. The dialects Clinge, Lamswaarde and Groenendijk from Zeeuws-Vlaanderen have been grouped together. They all belong to the Zeeuws group on the map. The geographically close Zundert, Roosendaal and Ossendrecht dialects have been grouped together in the tree, although they belong to the different Noord-Brabant group on the map.

Some distinguishing groups can be seen, which correspond with groups on Daan and Blok's map. The dialects of Groningen (Figure 8) and southern Dutch Limburg (Figure 9) form groups which correspond with Daan and Blok's map. The dialects of northern Noord-Holland and the islands of Texel and Vlieland form one group, as the map would predict (Figure 10). There is a branch of the tree that splits into Frisian dialects and Frisian city dialects (Figure 11). It is good to see this clear division but still close connection between Frisian dialects and Frisian city dialects. The Frisian city dialects are dialects which originate from Frisian, but have been influenced by the dialects from Holland in the 16th century (Jansen, 2002).

It is clear that Daan's Utrecht-Alblasserwaard group is not well-visible in the tree. Many dialects are clustered with other groups. Utrecht and Amersfoort are unresolved (Figure 7).

Dialects from eastern Noord-Brabant have been connected, but are not connected with dialects from the west of Noord-Brabant, which from one group on Daan's map. The dialects are however closely connected with two dialects from Zuid-Gelderland, a related, but different group on the map (Figure 12).

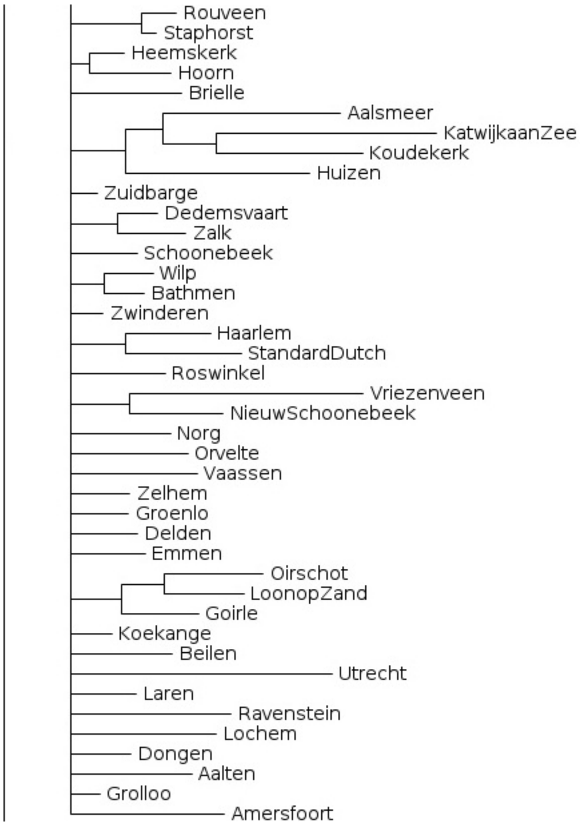


Figure 7: Equal model. A lot of dialects have not been grouped: dialects from the Utrecht-Alblasserwaard group like Utrecht and Amersfoort are on the same level as eastern dialects like Beilen and Emmen. Other dialects have been clustered into small groups: for example Goirle, Oirschot and Loon op Zand.

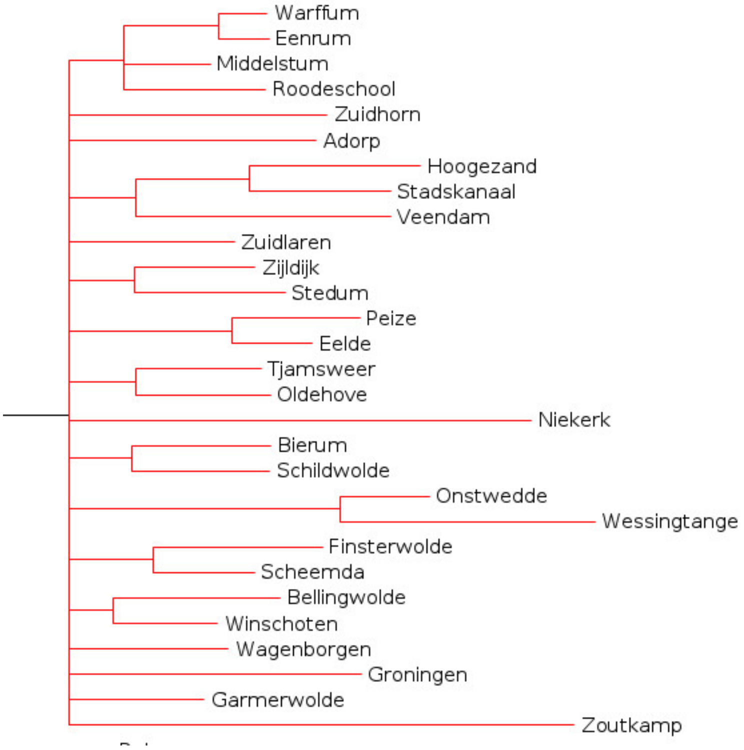


Figure 8: Equal model. The dialects of Groningen have been grouped according to Daan's map.

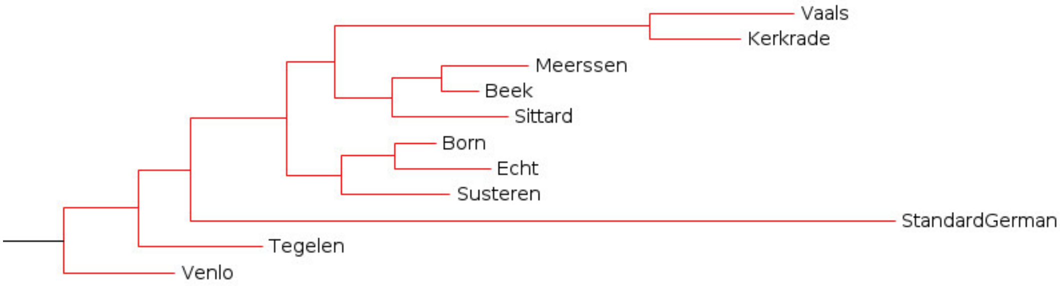


Figure 9: Equal model. The dialects of southern Dutch Limburg form a well-divided group that is coherent with the map.

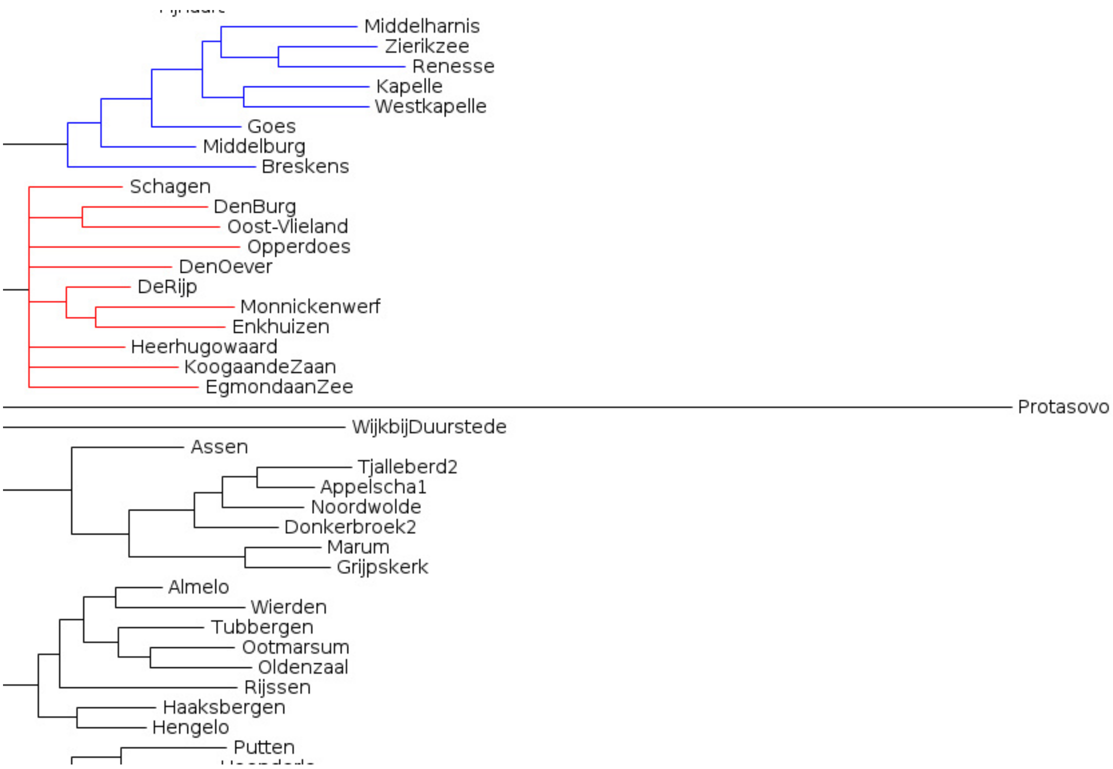


Figure 10: Equal model. The dialects of northern Noord-Holland are grouped with the dialects from the islands of Texel and Vlieland, as Daan predicts. The group is on the same level with a totally different, but also coherent group, that of Zeeland. The long branch length of Protasovo is remarkable and signifies a large distance compared to the other dialects.

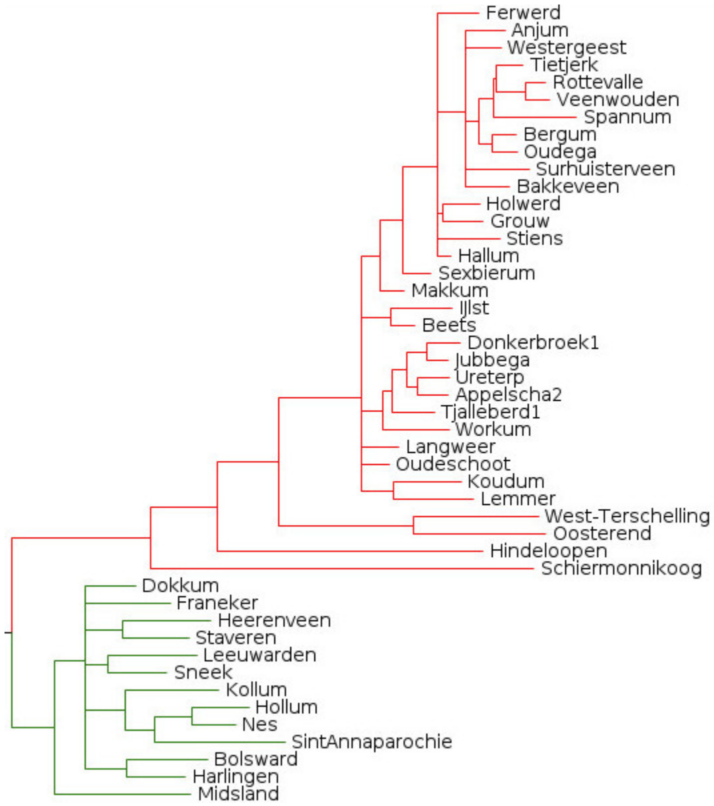


Figure 11: Equal model. The dialects of Friesland. The Frisian dialects (red) and Frisian city dialects (green) are related, but it is clear that there is a division.

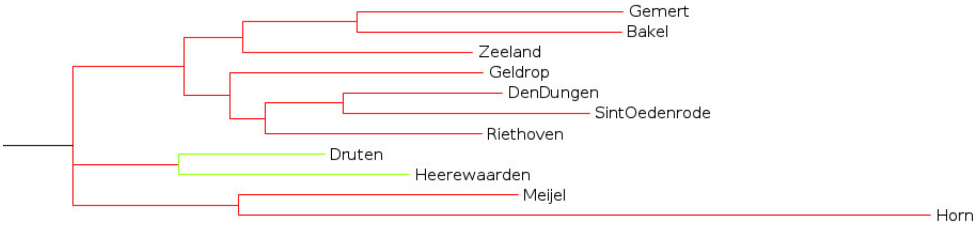


Figure 12: Equal model. Dialects from the eastern side of the Noord-Brabant group (red) have been grouped with dialects from the river region (green). Although these groups are related, it is remarkable that the Noord-Brabant dialects match with dialects from a different group, whereas they do not match with dialects from the western part of the same Noord-Brabant group.

The tree lacks some higher-order grouping. The dialects of the southern Netherlands are shown as a family in Daan's map using red shades. The Low-Saxon dialects of the eastern and northern Netherlands are shown as a family using green shades. These higher-order groupings are however not visible in the tree (Figure 13).

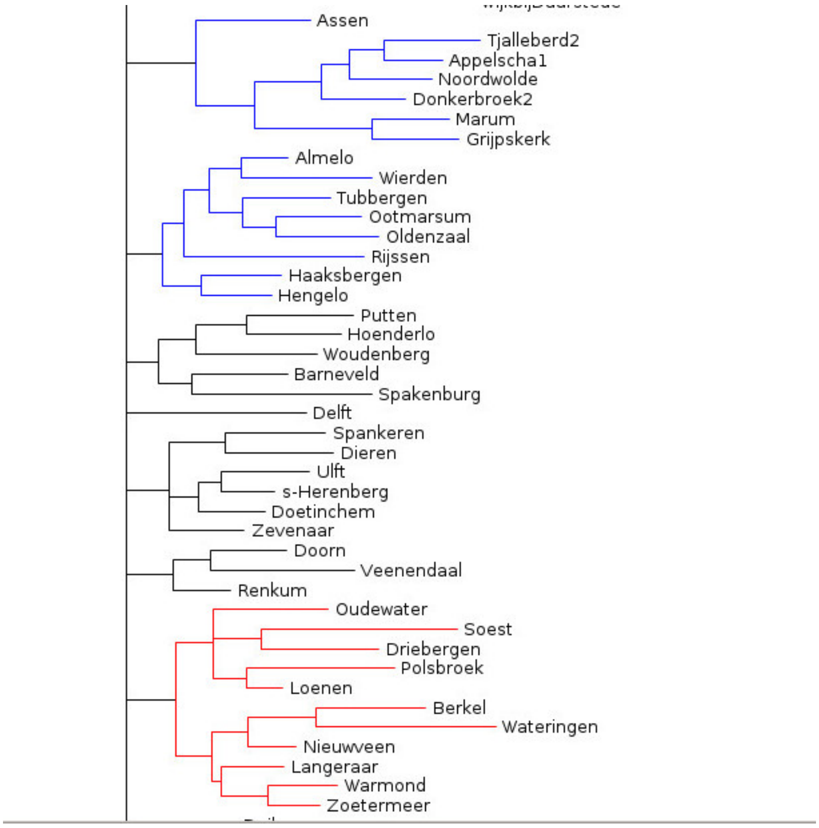


Figure 13: Equal model. The red group is a mix of dialects from the Utrecht-Alblasserwaard group (Oudewater, Soest, Driebergen, Polsbroek) and Zuid-Holland (Berkel, Wateringen, Nieuwveen, Langeraar, Warmond, Zoetermeer). Maybe the border between these groups is not really clear-cut, as Daan and Blok (1969) state. A second observation is that there is no sufficient higher-order grouping. The red group of western-central dialects is at the same level as the two blue groups of northeastern (Low-Saxon) dialects. These two blue groups would be expected to be on a different level, together with other Low-Saxon dialects.

The Protasovo (Plautdietsch) dialect has a very long branch, which shows that it differs a lot from the other dialects (Figure 10). This seems reasonable, given that it is a form of Dutch that has not been in contact with other Dutch dialects for centuries.

Concludingly, the dialects that have been grouped together form groups that are coherent with Daan and Blok. The groups have however not been grouped in higher-order groups that show relations between dialect regions. This makes the explanatory power of the tree smaller.

3.1.2 Gamma-distributed rate variation model

The consensus tree of the gamma-distributed rate variation model shows the same pattern as the consensus tree of the equal rate variation model. There are some differences in the groupings, sometimes these are improvements, sometimes these are degradations. It is not really clear whether these small differences are caused by different rate variation models. Differenes across different executions of the same rate variation model also occurred.

The groups of southern Dutch Limburg and Groningen are also salient in this consensus tree. Some groupings from Daan's map are better under the gamma model. The dialects of Twente have been grouped together under this model, whereas they were spread across different groups in the equal model (Figure 14). The dialects groups of Noord-Holland and Zuid-Holland are related, this is shown to a greater extent in this tree (Figure 15).

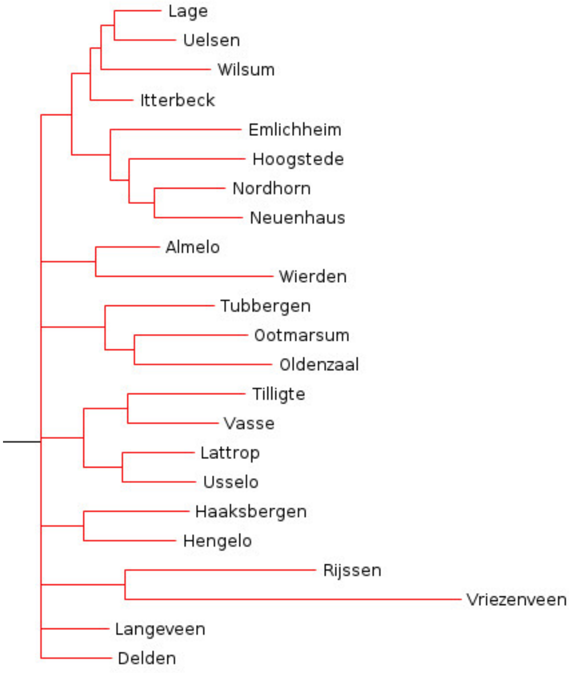


Figure 14: Gamma model. The dialects of Twente (and the directly neighbouring places in Germany) form one group under the gamma model, whereas they were spread across several groups in the equal model.

The distinction between Frisian dialects and Frisian city dialects is still shown, but this time the Frisian dialects are shown as a subgroup of the Frisian city dialects (Figure 16). This is not correct from a historical point of view, since the Frisian city dialects split off the Frisian dialects. However, from a distance point of view, it is less remarkable. The Frisian city dialects are closer to the

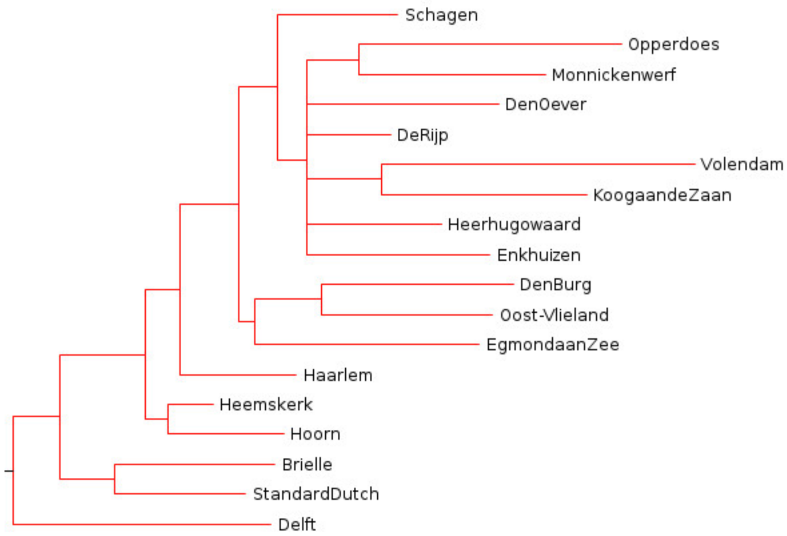


Figure 15: Gamma model. Dialects of Noord-Holland and Zuid-Holland have been combined as one group under the gamma model.

other Dutch dialects at the root of the tree, because they have been influenced by the dialects from Holland.

An interesting result is that the dialect of Katwijk aan Zee, a coastal place in Zuid-Holland, is grouped with the dialects of Zeeland, a different group further to the south (Figure 17). Apparently there are some shared characteristics between these coastal areas.

There are also dialects that were grouped in the consensus tree of the equal model, but are unresolved under the gamma model. Examples are the places Oldemarkt and Steenwijk.

It is hard to say whether the gamma or the equal model is better. The gamma model shows a few interesting groups that the equal model does not show, but it also leaves dialects ungrouped which the equal model grouped. Furthermore, some differences can occur across different executions of the same model and are not caused by the model choice.

3.2 Comparison: Belgian dialects

The Belgian data was kept apart to see whether the method works for a different data set as well. The Belgian data was processed in the same way as the Netherlandic data and the bayesian algorithm was run with the same parameters (1,000,000 generations, sample frequency 20).

Again, an equal rate variation model and a gamma-distributed rate variation were tried.

In the equal rate variation model, three important groups are seen. Only a few small groups are available and few dialects remain unresolved. This could mean

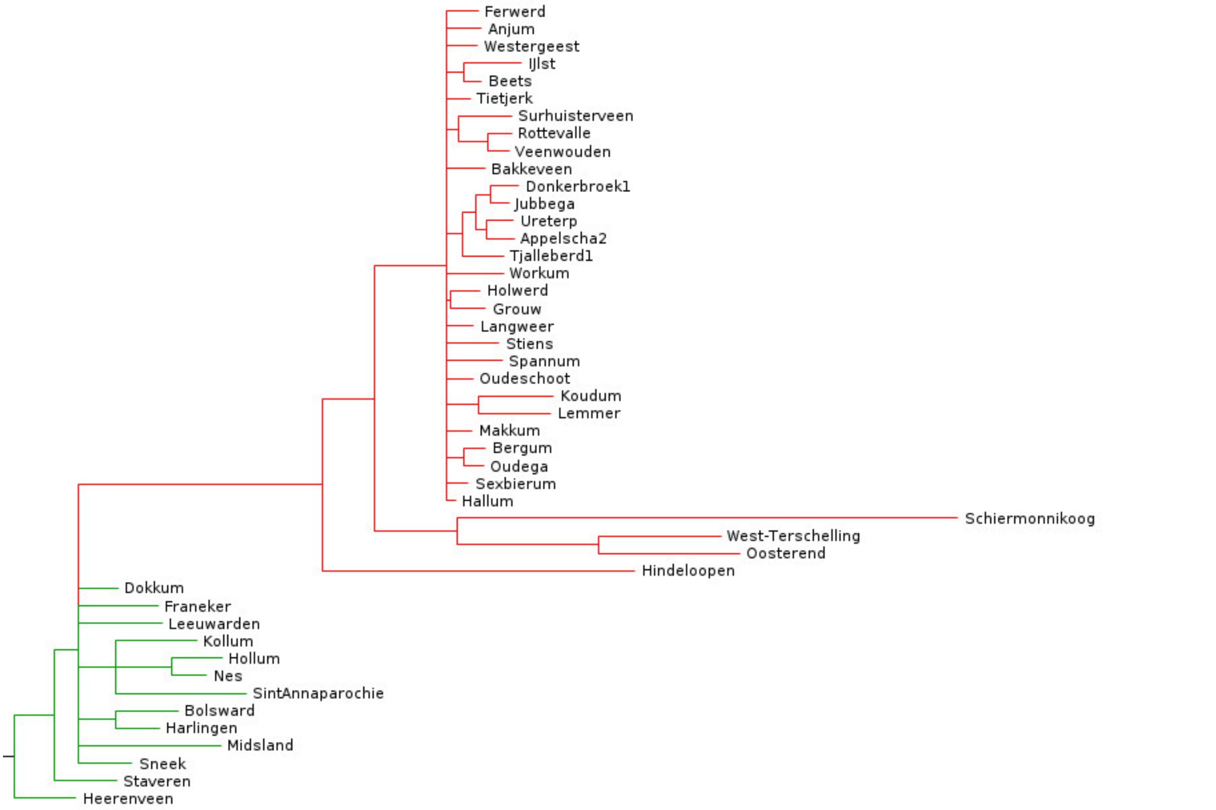


Figure 16: Gamma model. The Frisian dialects are shown as a subgroup of the Frisian dialects.

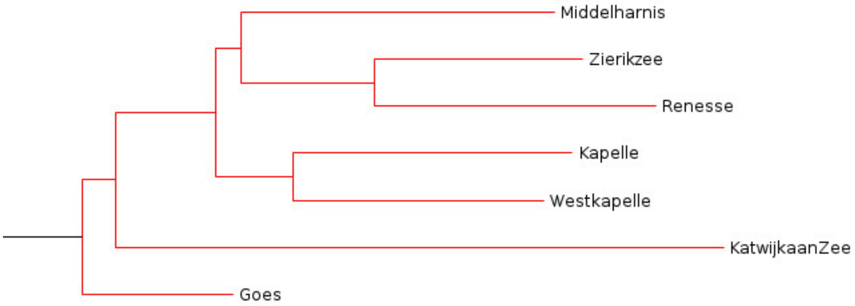


Figure 17: Gamma model. Katwijk aan Zee, a coastal place in the Zuid Holland dialect region, is grouped with dialects from the Zeeland group, further to the south.

there is less contradiction between the different trees than in the results of the Netherlandic data set. Also, the convergence between the runs is better.

The first group contains places from the east of Belgium (Figure 18). Most dialects belong to the Limburg group on Daan's map, two belong to the Brabant group and one belongs to the group of dialects between Brabant and Limburg. This group in the tree seems to be a coherent group of Limburg dialects with some other dialects which are geographically very close.

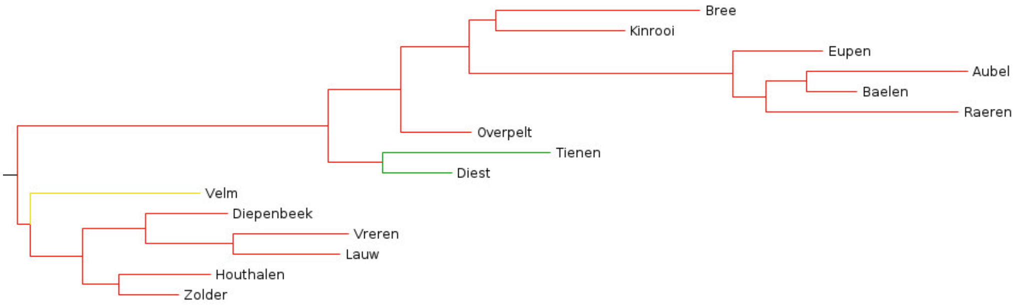


Figure 18: Equal model. This subtree contains dialects from the east of Belgium. The red dialects belong to the Limburg group, the green dialects belong to the Brabant group, the yellow dialect belongs to the group of dialects between Brabant and Limburg.

The second big group consists solely of Brabant dialects (Figure 19). There is a division into subgroups that are geographically close to each other. There are only two small groups of Brabant dialects that are not included in this big group and are connected separately in the tree. The grouping is stronger than in the Netherlandic tree.

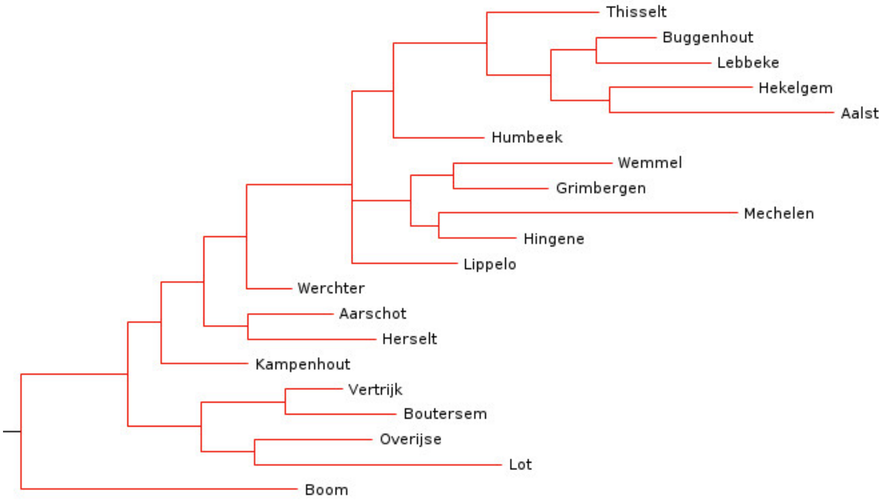


Figure 19: Equal model. This subtree consists of the Brabant dialects.

The third group consists of dialects from the west of Belgium and northern

France. The dialects are roughly in three groups from Daan's map: Western Flemish, Eastern Flemish and dialects between Western and Eastern Flemish. As can be seen in figure 20 the tree is nicely subdivided into these three groups.

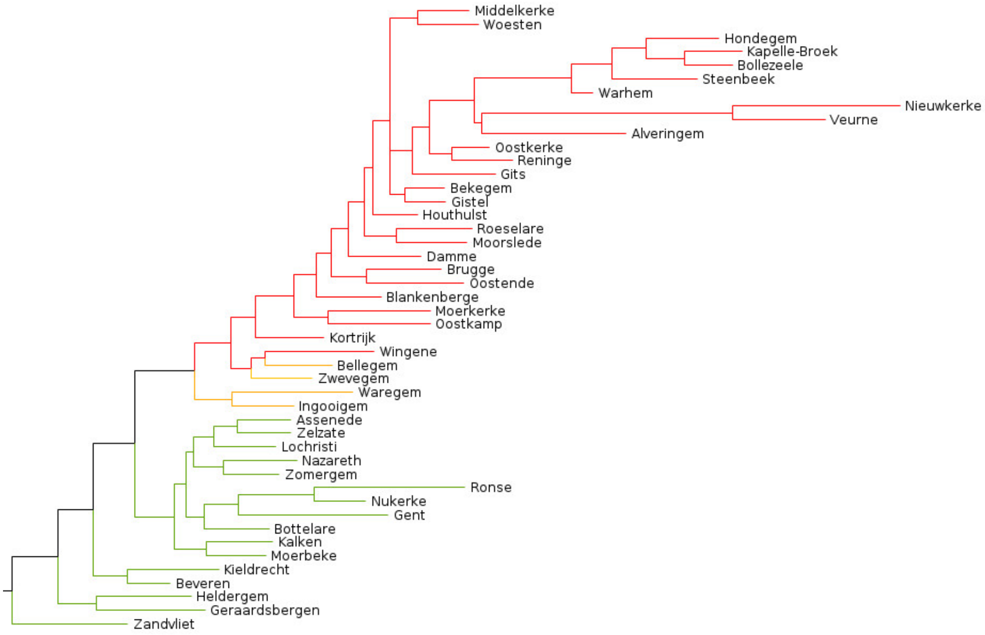


Figure 20: Equal model. This subtree contains dialects from the west of Belgium. The red dialects belong to the Western Flemish group, the green dialects belong to the Eastern Flemish group, the yellow dialects belong to the group of dialects between the Western and Eastern Flemish dialects.

The consensus tree under the gamma model shows the same three main groups. The only difference is that some dialects have split from the bigger groups and formed a smaller group.

4 Discussion

The data from the RND which was used seems to have given a reliable set of basic words. However, there is no guarantee that the words are used as often in one area as in another area. The closest approximation to a list of words that is used in every area would be a Swadesh list.

The alignment has been done using a system of sound classes, which gives good results. The quality of the alignments could possibliy become even higher. To focus on phonology and filter out morphological effects, the stems of composed words could all be put in the same order. Furthermore, lemmas which contain words from different cognate sets, could be split in several lemmas: one for every cognate sets. Finally, more combinations of phones and their diacritics could be added to the token list. The diacritics that were not listed in (Hoppenbrouwers and Hoppenbrouwers, 1988) were not taken into account in the alignment now.

As a summary of the tree sample, consensus trees were used. A characteristic of the consensus tree is that it is not guaranteed to be a real tree from the sample, but a reconciliation of the trees. Contradicting branchings are solved by placing a branch at a lower level. Other tree summaries are the the maximum probability tree (Nichols and Warnow, 2008) and the maximum clade credibility tree (Dunn, 2008). Both methods pick a tree which exists in the tree sample: the tree with the highest probability or the tree with the highest sum of probabilities of the branchings respectively. The phylogenetic program (MrBayes) that was used, was not accustomed to the creation of these trees. It was possible to fetch a maximum probability tree topology, but without branch lengths. Furthermore, it was possible to create a maximum clade credibility tree with an external program, but this did not succeed. For these reasons, only consensus trees were used in my analysis.

Although bayesian inference is a quantitative method, which draws conclusions from large amounts of data, the evaluation of the method in this thesis was done qualitatively. Ideally, an objective measure of distance between a bayesian inference tree and Daan's dialect map would be used. Both representations would then have to be converted to the same format. Zhang and Shasha (1989) proposes an algorithm for edit distance between trees. Implementing this algorithm and processing the tree data in such a way that it could be read by the algorithm could be a direction for future research. It would also have to be assessed whether the edit distance between language trees coincides with a linguistic feeling of similarity between language trees.

In earlier quantative dialect research (Heeringa and Nerbonne, 2006), the evaluation of the tree was also done qualitatively. However, new methods are being applied to visualize the data in such a way that it is easier to do a human comparison with the gold standard. Nerbonne et al. (2011) present the Gabmap package, which has, among other features, the possibility to project a dialect tree onto a map.

Gamma-distributed and equal rate variation models were evaluated. The differences in the resulting trees of the models were not very large. It seems that for the current input format, long strings of phonological features, the choice of the rate variation model is not of utmost importance.

The results for the bayesian inference on the Belgian dialects were better than the results on the Netherlandic dialects. The bayesian analyses for the Belgian dialects had better convergence rates than the analyses for the Netherlandic dialects. It must be noted that the number of Belgian dialects was smaller than the number of Netherlandic dialects (94 vs. 269 dialects), but they ran the same number of generations. It may be that the Netherlandic analyses should have run for more generations, to compensate for the large number of dialects. This is however made inattractive by the long running times of the algorithm. There could also be other reasons for the better performance on the Belgian data set. For example, it could be the case that the Belgian data set had clearer divisions between the dialects, making it easier to generate a tree.

5 Conclusion

The application of bayesian inference on the Netherlandic dialects performed well on local groups. Distinctive groups from common linguistic theory were visible. The grouping was very accurate, hardly any groupings were made that were not coherent with the dialect map by Daan and Blok. However, many dialects remained ungrouped. Also, local groups were not grouped with other groups in order to get higher-order families. This limited the explanatory power of the results.

The bayesian inference for the Belgian dialects gave suprisingly good results. Almost all dialects were grouped and there was higher-order grouping apparent. There were a few large groups, which contained dialects from a bounded geographical area, eg. the east of Belgium. These large groups were divided into smaller groups, which mostly followed the dialect groups from the dialect map by Daan and Blok.

All in all, bayesian inference seems to be a good addition to the tools used to determine the phylogeny of dialects. The performance of the method is not constant enough to use it as the only method to create a dialect tree. In this thesis, the method performed better on the Belgian dialects than on the Netherlandic dialects. However, once the results have been validated using a dialect map for the researched area, insights from bayesian inference can be used to get a full image of dialect kinship. For example, even if no higher-order groupings are returned in a bayesian inference tree, local groupings (as in Figure 17) can give interesting clues about relationships between dialects.

6 Literature

Campbell, L. (1998). *Historical linguistics: An introduction.* MIT press.

Daan, J. and Blok, D. (1969). *Van Randstad tot Landrand: Toelichting bij de kaart: Dialecten en Naamkunde. Bijdragen en mededelingen der Dialectenkommissie van de Koninklijke Nederlandse Akademie van Wetenschappen.* Noord-Hollandsche Uitgevers Maatschappij.

Dunn, M. (2008). Language phylogenies (in press). *Routledge handbook of historical linguistics.* http://pubman.mpdl.mpg.de/pubman/item/escidoc:1851319:5/component/escidoc:1851318/dunn-phylogenetic-approaches.pdf.

Heeringa, W. (2001). De selectie en digitalisatie van dialecten en woorden uit de reeks nederlandse dialectatlassen. *TABU, Bulletin voor Taalwetenschap*, 31, number 1/2:61–103.

Heeringa, W. and Nerbonne, J. (2006). De analyse van taalvariatie in het nederlandse dialectgebied: methoden en resultaten op basis van lexicon en uitspraak. *Nederlandse Taalkunde*, 11(3):18–257.

Hoppenbrouwers, C. and Hoppenbrouwers, G. (1988). De featurefrequentiemethode en de classificatie van nederlandse dialecten. *TABU, Bulletin voor Taalwetenschap*, Jaargang 18, nummer 2, 1988. Retrieved from http://urd.let.rug.nl/nerbonne/papers/inferring-sound-changes-Prokic-et-al-2011-Diachronica.pdf on 15-06-2014.

Huelsenbeck, J., Ronquist, F., Nielsen, R., and Bollback, J. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, 294(5550):2310–2314.

Jansen, M. (2002). De dialecten van ameland en midsland in vergelijking met het stadsfries. *Us Wurk. Tydskrift foar frisistyk*, 51:128–152. Retrieved from http://depot.knaw.nl/9683 on 25-06-2014.

Larget, B. and Simon, D. (1999). Markov chain monte carlo algorithms for the bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution*, 16:750–759.

List, J.-M. (2012). Multiple sequence alignment in historical linguistics. a sound class based approach. In *Proceedings of ConSOLE XIX*, pages 241–260.

List, J.-M. and Moran, S. (2013). An open source toolkit for quantitative historical linguistics. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations, August 4-9, Sofia, Bulgaria.*, pages 13–18.

Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P., and Leinonen, T. ((2011)). Gabmap – a web application for dialectology. *Dialectologia*, Special Issue II:65–89.

Nichols, J. and Warnow, T. (2008). Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass*, 2.5:760–820.

Nieuweboer, R. (1998). The altai dialect of plautdiitsh (west-siberian mennonite low german). Master's thesis, University of Groningen.

Prokić, J., Gray, R., and Nerbonne, J. (2011). Inferring sound changes using bayesian mcmc. *Submitted to Diachronica, 1/2011*. Retrieved from http://urd.let.rug.nl/nerbonne/papers/inferring-sound-changes-Prokic-et-al-2011-Diachronica.pdf on 15-06-2014.

Rambaut, A. (2013). Figtree 1.4.1. tree figure drawing tool.

Ronquist, F. and Huelsenbeck, J. (2003). Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19:15721574.

Ronquist, F., Huelsenbeck, J., and Teslenko, M. (2011). Draft mrbayes version 3.2 manual: Tutorials and model summaries. Retrieved from http://mrbayes.sourceforge.net/mb3.2_manual.pdf on 25-06-2014.

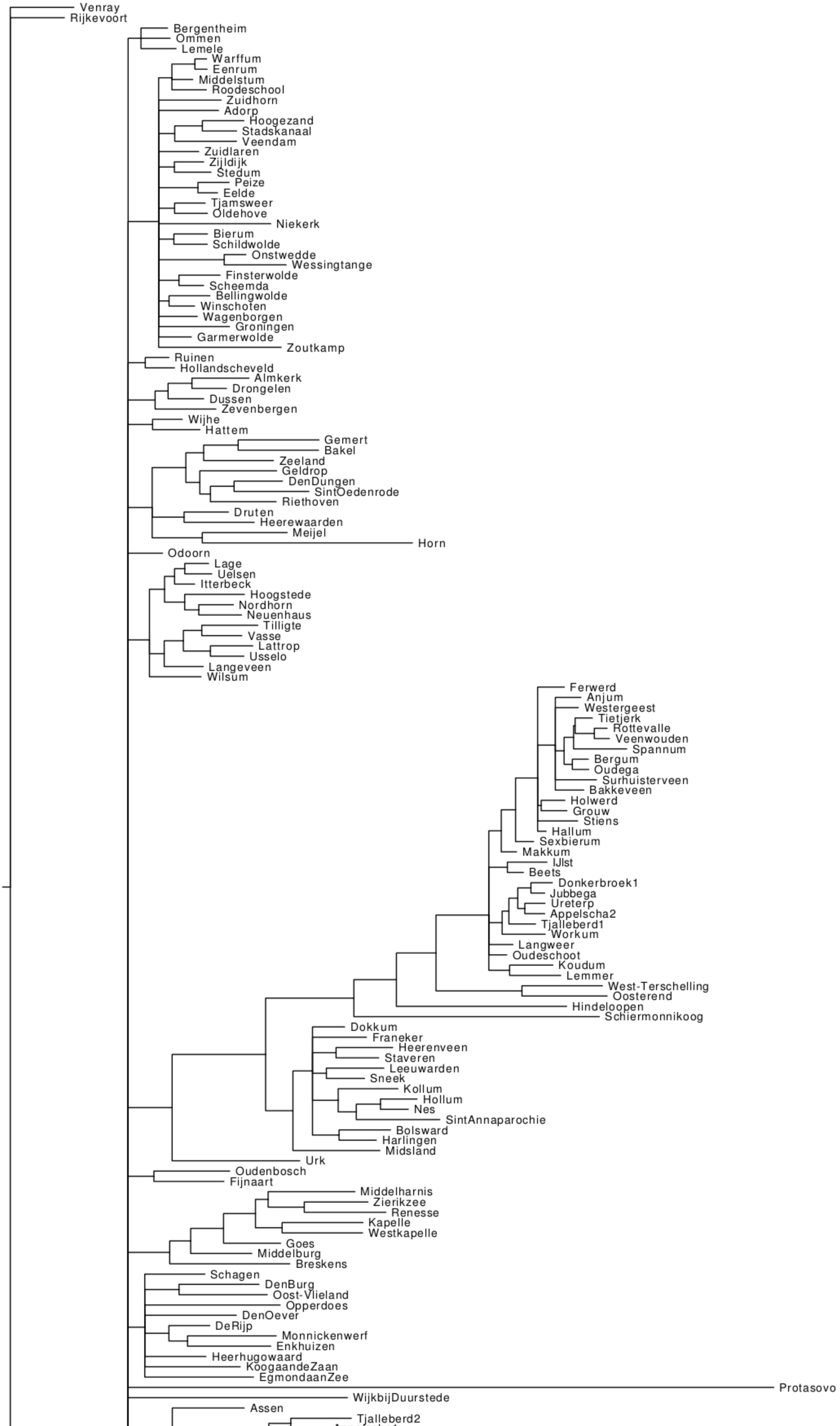
Theiling, H. (2008). cxs2ipa. an x-sampa to ipa converter. Retrieved from http://www.theiling.de/ipa/ on 16-06-2014.

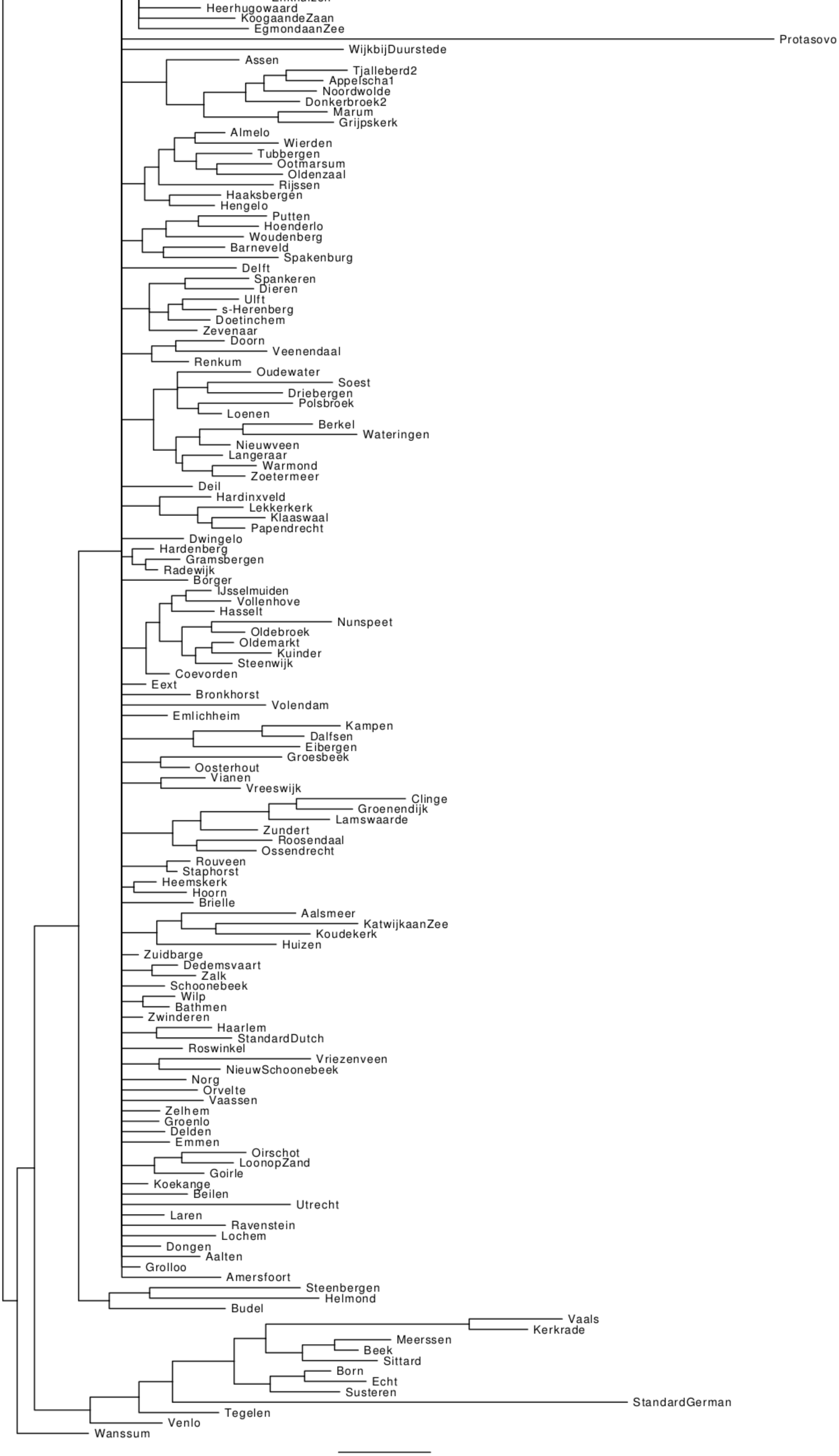
Zhang, K. and Shasha, D. (1989). Simple fast algorithms for the editing distance between trees and related problems. *SIAM journal on computing*, 18(6):1245–1262.

7 Appendix

7.1 Tree of Netherlandic dialects – equal rate variation

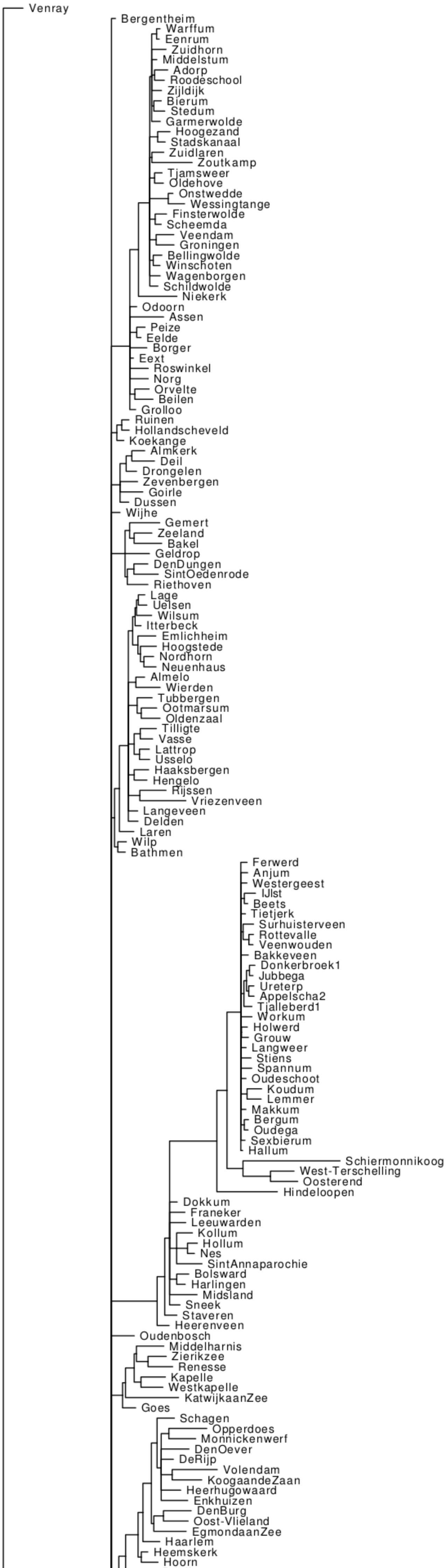
The tree spans two pages.

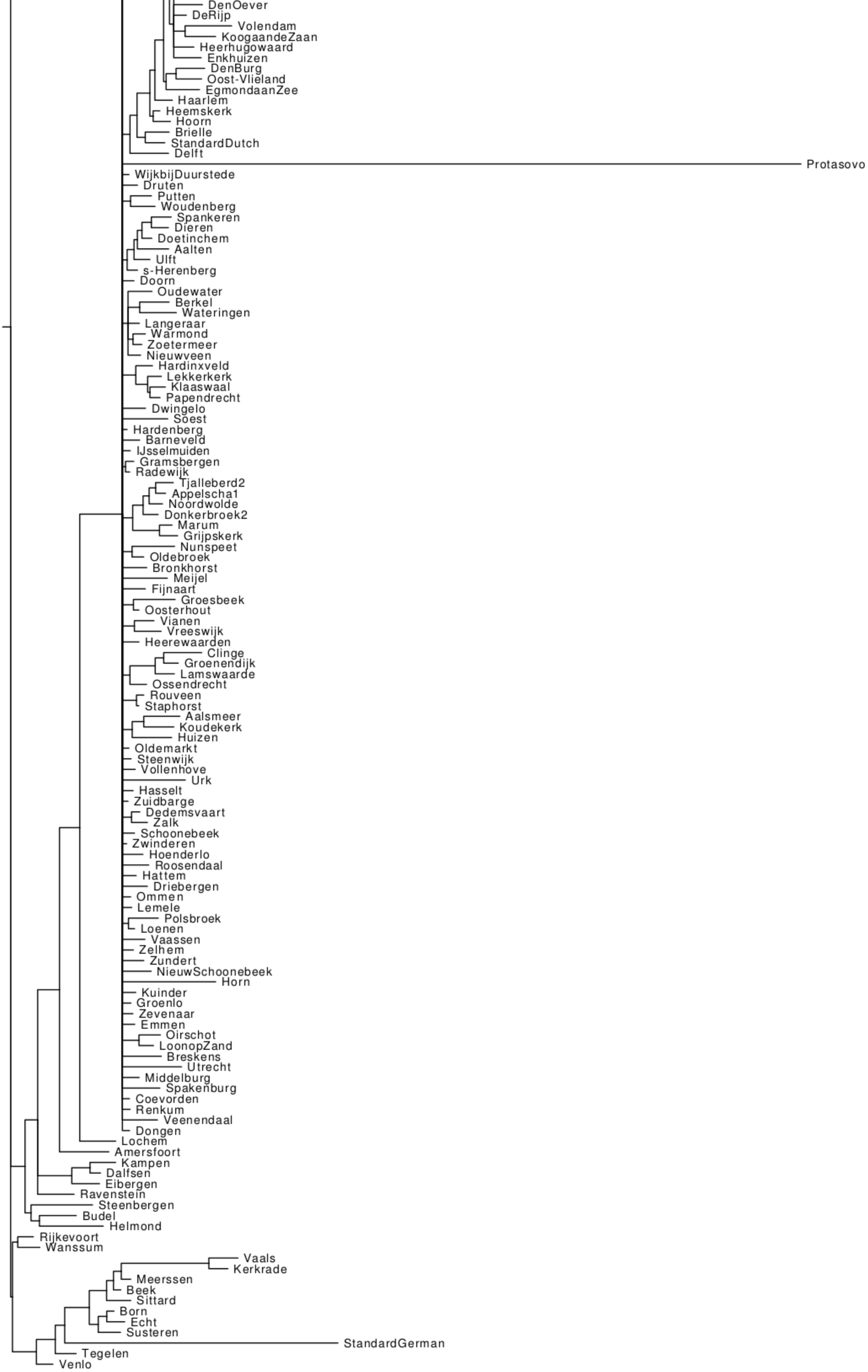

Venray
Rijkevoort
Bergentheim
Ommen
Lemele
Warffum
Eenrum
Middelstum
Roodeschool
Zuidhorn
Adorp
Hoogezand
Stadskanaal
Veendam
Zuidlaren
Zijldijk
Stedum
Peize
Eelde
Tjamsweer
Oldehove
Niekerk
Bierum
Schildwolde
Onstwedde
Wessingtange
Finsterwolde
Scheemda
Bellingwolde
Winschoten
Wagenborgen
Groningen
Garmerwolde
Zoutkamp
Ruinen
Hollandscheveld
Almkerk
Drongelen
Dussen
Zevenbergen
Wijhe
Hattem
Gemert
Bakel
Zeeland
Geldrop
DenDungen
SintOedenrode
Riethoven
Druten
Heerewaarden
Meijel
Horn
Odoorn
Lage
Uelsen
Itterbeck
Hoogstede
Nordhorn
Neuenhaus
Tilligte
Vasse
Lattrop
Usselo
Langeveen
Wilsum
Ferwerd
Anjum
Westergeest
Tietjerk
Rottevalle
Veenwouden
Spannum
Bergum
Oudega
Surhuisterveen
Bakkeveen
Holwerd
Grouw
Stiens
Hallum
Sexbierum
Makkum
IJlst
Beets
Donkerbroek1
Jubbega
Ureterp
Appelscha2
Tjalleberd1
Workum
Langweer
Oudeschoot
Koudum
Lemmer
West-Terschelling
Oosterend
Hindeloopen
Schiermonnikoog
Dokkum
Franeker
Heerenveen
Staveren
Leeuwarden
Sneek
Kollum
Hollum
Nes
SintAnnaparochie
Bolsward
Harlingen
Midsland
Urk
Oudenbosch
Fijnaart
Middelharnis
Zierikzee
Renesse
Kapelle
Westkapelle
Goes
Middelburg
Breskens
Schagen
DenBurg
Oost-Vlieland
Opperdoes
DenOever
DeRijp
Monnickenwerf
Enkhuizen
Heerhugowaard
KoogaandeZaan
EgmondaanZee
Protasovo
WijkbijDuurstede
Assen
Tjalleberd2


Enkhuizen
Heerhugowaard
KoogaandeZaan
EgmondaanZee
Protasovo
WijkbijDuurstede
Assen
Tjalleberd2
Appelscha1
Noordwolde
Donkerbroek2
Marum
Grijpskerk
Almelo
Wierden
Tubbergen
Ootmarsum
Oldenzaal
Rijssen
Haaksbergen
Hengelo
Putten
Hoenderlo
Woudenberg
Barneveld
Spakenburg
Delft
Spankeren
Dieren
Ulft
s-Herenberg
Doetinchem
Zevenaar
Doorn
Veenendaal
Renkum
Oudewater
Soest
Driebergen
Polsbroek
Loenen
Berkel
Wateringen
Nieuwveen
Langeraar
Warmond
Zoetermeer
Deil
Hardinxveld
Lekkerkerk
Klaaswaal
Papendrecht
Dwingelo
Hardenberg
Gramsbergen
Radewijk
Borger
IJsselmuiden
Vollenhove
Hasselt
Nunspeet
Oldebroek
Oldemarkt
Kuinder
Steenwijk
Coevorden
Eext
Bronkhorst
Volendam
Emlichheim
Kampen
Dalfsen
Eibergen
Groesbeek
Oosterhout
Vianen
Vreeswijk
Clinge
Groenendijk
Lamswaarde
Zundert
Roosendaal
Ossendrecht
Rouveen
Staphorst
Heemskerk
Hoorn
Brielle
Aalsmeer
KatwijkaanZee
Koudekerk
Huizen
Zuidbarge
Dedemsvaart
Zalk
Schoonebeek
Wilp
Bathmen
Zwinderen
Haarlem
StandardDutch
Roswinkel
Vriezenveen
NieuwSchoonebeek
Norg
Orvelte
Vaassen
Zelhem
Groenlo
Delden
Emmen
Oirschot
LoonopZand
Goirle
Koekange
Beilen
Utrecht
Laren
Ravenstein
Lochem
Dongen
Aalten
Grolloo
Amersfoort
Steenbergen
Helmond
Budel
Vaals
Kerkrade
Meerssen
Beek
Sittard
Born
Echt
Susteren
StandardGerman
Tegelen
Venlo
Wanssum

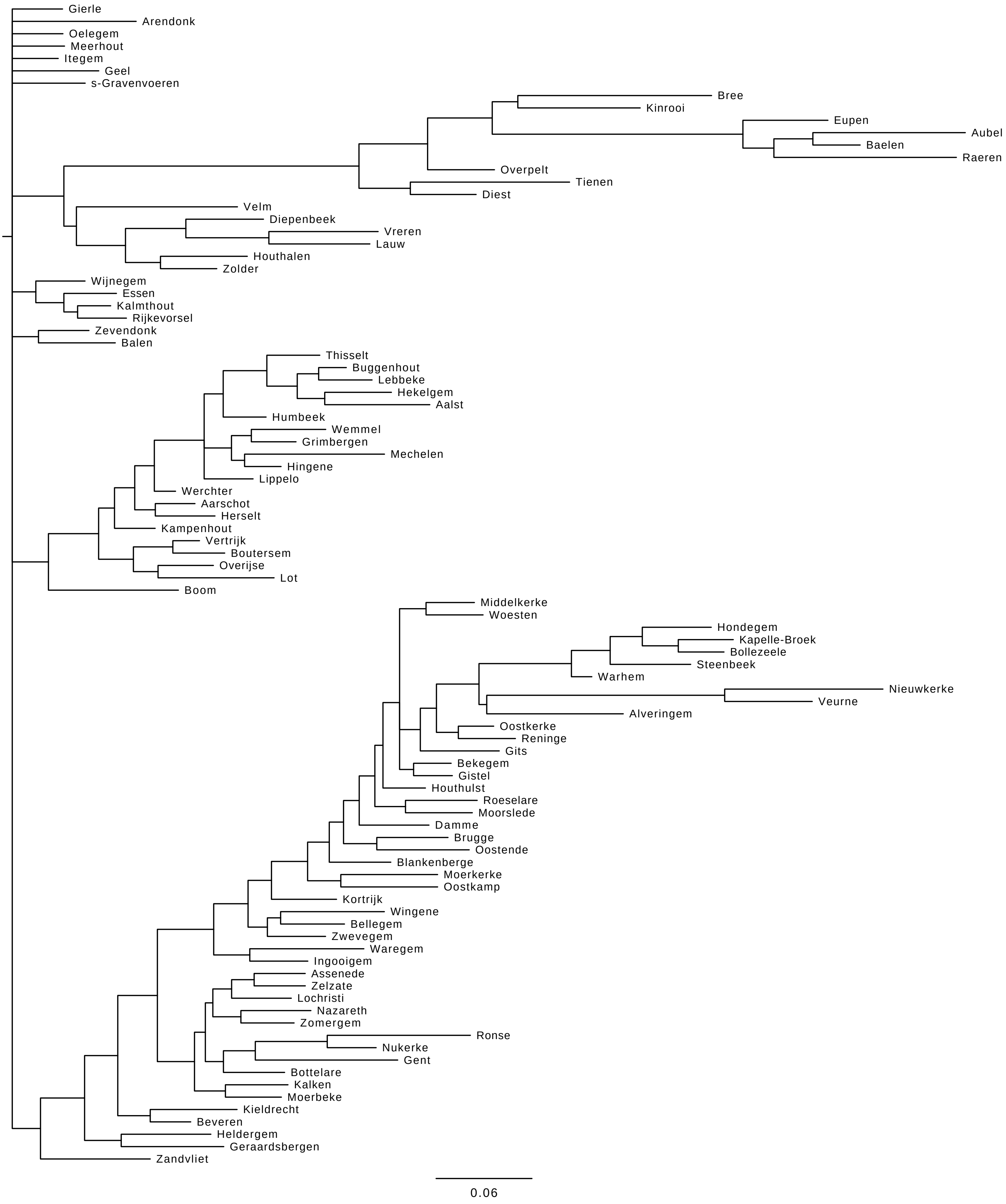
7.2 Tree of Netherlandic dialects – gamma rate variation

The tree spans two pages.

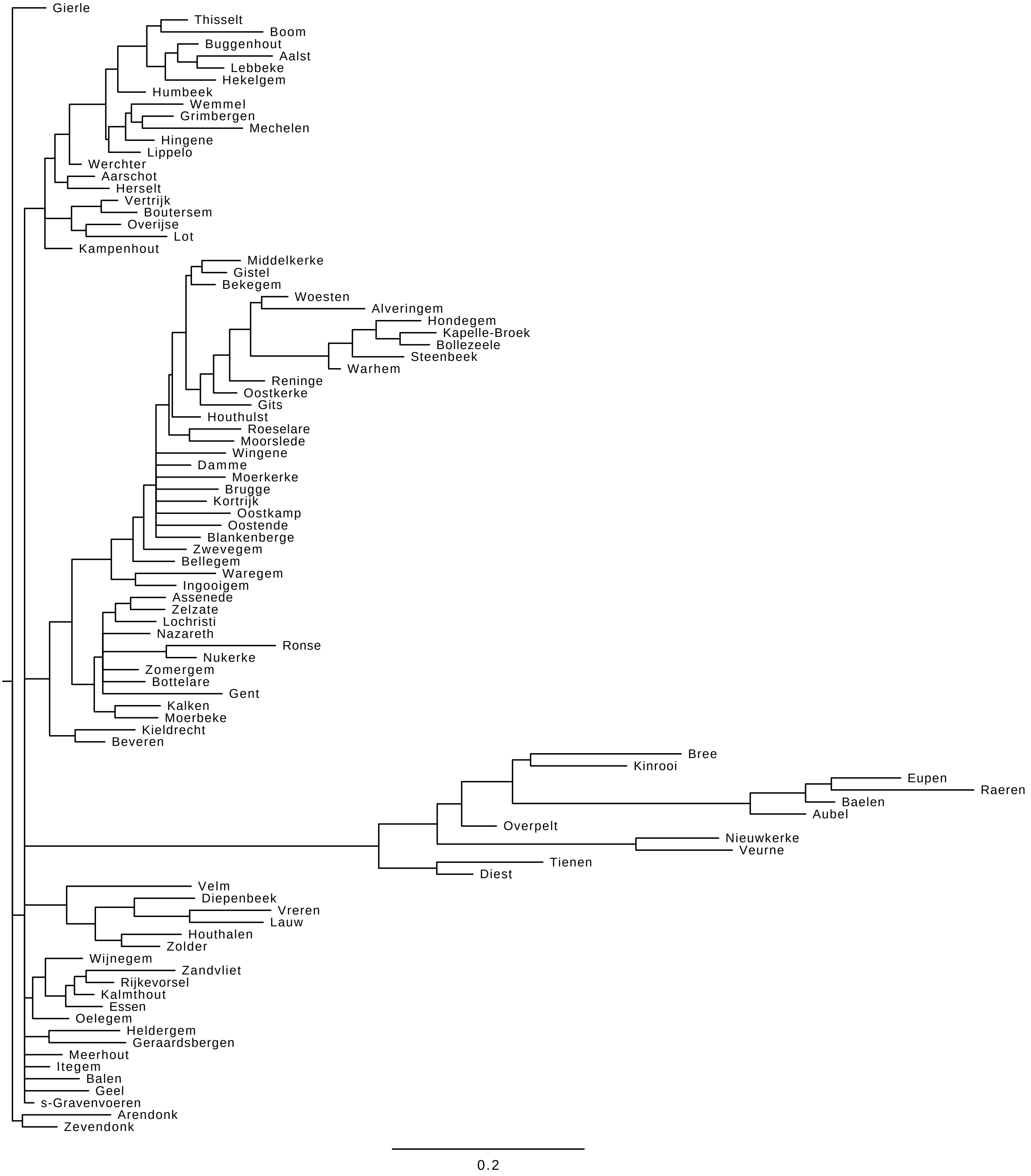

Venray
Bergentheim
Warffum
Eenrum
Zuidhorn
Middelstum
Adorp
Roodeschool
Zijldijk
Bierum
Stedum
Garmerwolde
Hoogezand
Stadskanaal
Zuidlaren
Zoutkamp
Tjamsweer
Oldehove
Onstwedde
Wessingtange
Finsterwolde
Scheemda
Veendam
Groningen
Bellingwolde
Winschoten
Wagenborgen
Schildwolde
Niekerk
Odoorn
Assen
Peize
Eelde
Borger
Eext
Roswinkel
Norg
Orvelte
Beilen
Grolloo
Ruinen
Hollandscheveld
Koekange
Almkerk
Deil
Drongelen
Zevenbergen
Goirle
Dussen
Wijhe
Gemert
Zeeland
Bakel
Geldrop
DenDungen
SintOedenrode
Riethoven
Lage
Uelsen
Wilsum
Itterbeck
Emlichheim
Hoogstede
Nordhorn
Neuenhaus
Almelo
Wierden
Tubbergen
Ootmarsum
Oldenzaal
Tilligte
Vasse
Lattrop
Usselo
Haaksbergen
Hengelo
Rijssen
Vriezenveen
Langeveen
Delden
Laren
Wilp
Bathmen
Ferwerd
Anjum
Westergeest
IJlst
Beets
Tietjerk
Surhuisterveen
Rottevalle
Veenwouden
Bakkeveen
Donkerbroek1
Jubbega
Ureterp
Appelscha2
Tjalleberd1
Workum
Holwerd
Grouw
Langweer
Stiens
Spannum
Oudeschoot
Koudum
Lemmer
Makkum
Bergum
Oudega
Sexbierum
Hallum
Schiermonnikoog
West-Terschelling
Oosterend
Hindeloopen
Dokkum
Franeker
Leeuwarden
Kollum
Hollum
Nes
SintAnnaparochie
Bolsward
Harlingen
Midsland
Sneek
Staveren
Heerenveen
Oudenbosch
Middelharnis
Zierikzee
Renesse
Kapelle
Westkapelle
KatwijkaanZee
Goes
Schagen
Opperdoes
Monnickenwerf
DenOever
DeRijp
Volendam
KoogaandeZaan
Heerhugowaard
Enkhuizen
DenBurg
Oost-Vlieland
EgmondaanZee
Haarlem
Heemskerk
Hoorn


DenOever
DeRijp
Volendam
KoogaandeZaan
Heerhugowaard
Enkhuizen
DenBurg
Oost-Vlieland
EgmondaanZee
Haarlem
Heemskerk
Hoorn
Brielle
StandardDutch
Delft
Protasovo
WijkbijDuurstede
Druten
Putten
Woudenberg
Spankeren
Dieren
Doetinchem
Aalten
Ulft
s-Herenberg
Doorn
Oudewater
Berkel
Wateringen
Langeraar
Warmond
Zoetermeer
Nieuwveen
Hardinxveld
Lekkerkerk
Klaaswaal
Papendrecht
Dwingelo
Soest
Hardenberg
Barneveld
IJsselmuiden
Gramsbergen
Radewijk
Tjalleberd2
Appelscha1
Noordwolde
Donkerbroek2
Marum
Grijpskerk
Nunspeet
Oldebroek
Bronkhorst
Meijel
Fijnaart
Groesbeek
Oosterhout
Vianen
Vreeswijk
Heerewaarden
Clinge
Groenendijk
Lamswaarde
Ossendrecht
Rouveen
Staphorst
Aalsmeer
Koudekerk
Huizen
Oldemarkt
Steenwijk
Vollenhove
Urk
Hasselt
Zuidbarge
Dedemsvaart
Zalk
Schoonebeek
Zwinderen
Hoenderlo
Roosendaal
Hattem
Driebergen
Ommen
Lemele
Polsbroek
Loenen
Vaassen
Zelhem
Zundert
NieuwSchoonebeek
Horn
Kuinder
Groenlo
Zevenaar
Emmen
Oirschot
LoonopZand
Breskens
Utrecht
Middelburg
Spakenburg
Coevorden
Renkum
Veenendaal
Dongen
Lochem
Amersfoort
Kampen
Dalfsen
Eibergen
Ravenstein
Steenbergen
Budel
Helmond
Rijkevoort
Wanssum
Vaals
Kerkrade
Meerssen
Beek
Sittard
Born
Echt
Susteren
StandardGerman
Tegelen
Venlo

7.3 Tree of Belgian dialects – equal rate variation

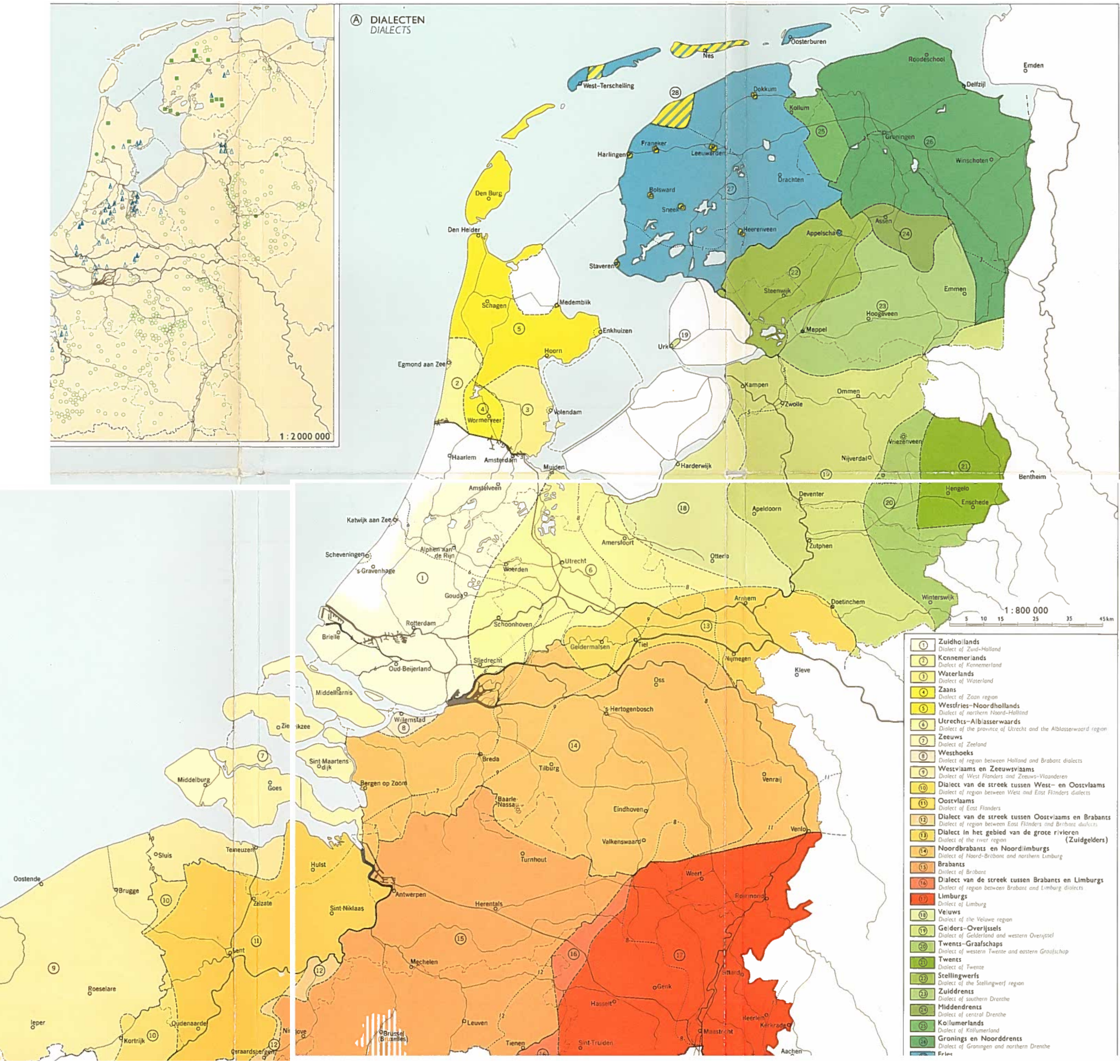

Gierle
Arendonk
Oelegem
Meerhout
Itegem
Geel
s-Gravenvoeren
Bree
Kinrooi
Eupen
Aubel
Baelen
Raeren
Overpelt
Tienen
Diest
Velm
Diepenbeek
Vreren
Lauw
Houthalen
Zolder
Wijnegem
Essen
Kalmthout
Rijkevorsel
Zevendonk
Balen
Thisselt
Buggenhout
Lebbeke
Hekelgem
Aalst
Humbeek
Wemmel
Grimbergen
Mechelen
Hingene
Lippelo
Werchter
Aarschot
Herselt
Kampenhout
Vertrijk
Boutersem
Overijse
Lot
Boom
Middelkerke
Woesten
Hondegem
Kapelle-Broek
Bollezeele
Steenbeek
Warhem
Nieuwkerke
Veurne
Alveringem
Oostkerke
Reninge
Gits
Bekegem
Gistel
Houthulst
Roeselare
Moorslede
Damme
Brugge
Oostende
Blankenberge
Moerkerke
Oostkamp
Kortrijk
Wingene
Bellegem
Zwevegem
Waregem
Ingooigem
Assenede
Zelzate
Lochristi
Nazareth
Zomergem
Ronse
Nukerke
Gent
Bottelare
Kalken
Moerbeke
Kieldrecht
Beveren
Heldergem
Geraardsbergen
Zandvliet
0.06

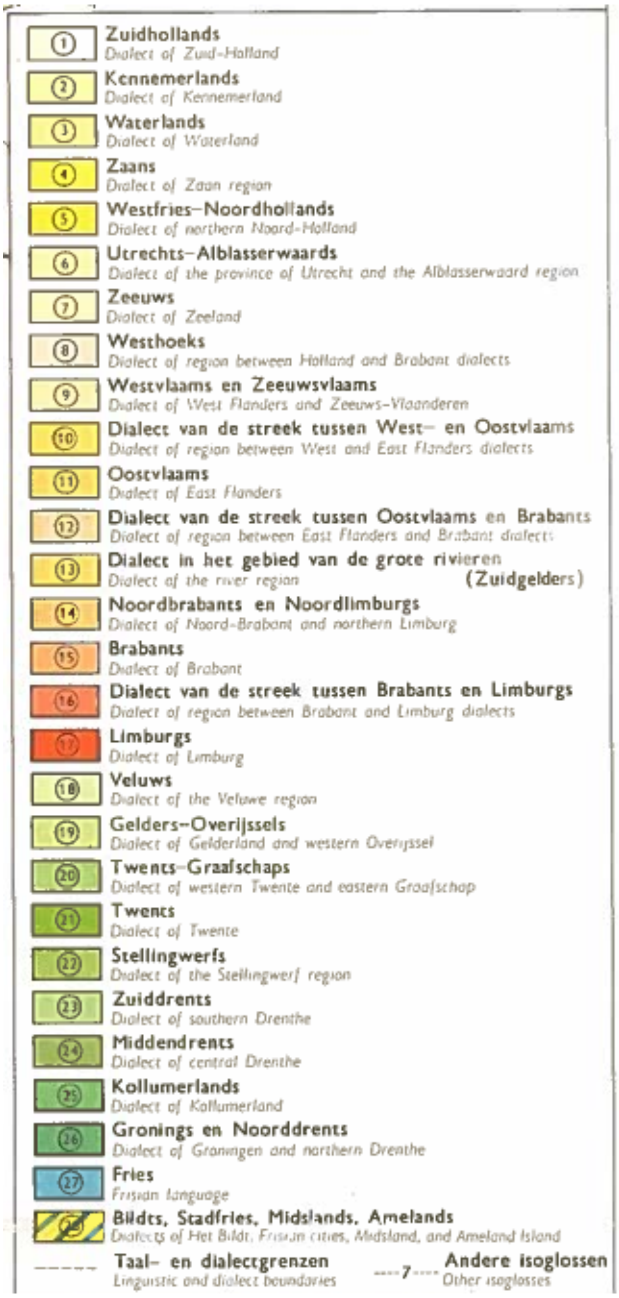
7.4 Tree of Belgian dialects – gamma rate variation


Gierle
Thisselt
Boom
Buggenhout
Aalst
Lebbeke
Hekelgem
Humbeek
Wemmel
Grimbergen
Mechelen
Hingene
Lippelo
Werchter
Aarschot
Herselt
Vertrijk
Boutersem
Overijse
Lot
Kampenhout
Middelkerke
Gistel
Bekegem
Woesten
Alveringem
Hondegem
Kapelle-Broek
Bollezeele
Steenbeek
Warhem
Reninge
Oostkerke
Gits
Houthulst
Roeselare
Moorslede
Wingene
Damme
Moerkerke
Brugge
Kortrijk
Oostkamp
Oostende
Blankenberge
Zwevegem
Bellegem
Waregem
Ingooigem
Assenede
Zelzate
Lochristi
Nazareth
Ronse
Nukerke
Zomergem
Bottelare
Gent
Kalken
Moerbeke
Kieldrecht
Beveren
Bree
Kinrooi
Eupen
Raeren
Baelen
Aubel
Overpelt
Nieuwkerke
Veurne
Tienen
Diest
Velm
Diepenbeek
Vreren
Lauw
Houthalen
Zolder
Wijnegem
Zandvliet
Rijkevorsel
Kalmthout
Essen
Oelegem
Heldergem
Geraardsbergen
Meerhout
Itegem
Balen
Geel
s-Gravenvoeren
Arendonk
Zevendonk
0.2

7.5 Dialect map by Daan and Blok (1969)

The first page shows the map, the second page shows the legend.


Ⓐ DIALECTEN
DIALECTS
1 : 2 000 000
1 : 800 000
0 5 10 15 25 35 45 km
① Zuidhollands
Dialect of Zuid-Holland
② Kennemerlands
Dialect of Kennemerland
③ Waterlands
Dialect of Waterland
④ Zaans
Dialect of Zaan region
⑤ Westfries-Noordhollands
Dialect of northern Noord-Holland
⑥ Utrechts-Alblasserwaards
Dialect of the province of Utrecht and the Alblasserwaard region
⑦ Zeeuws
Dialect of Zeeland
⑧ Westhoeks
Dialect of region between Holland and Brabant dialects
⑨ Westvlaams en Zeeuwsvlaams
Dialect of West Flanders and Zeeuws-Vlaanderen
⑩ Dialect van de streek tussen West- en Oostvlaams
Dialect of region between West and East Flanders dialects
⑪ Oostvlaams
Dialect of East Flanders
⑫ Dialect van de streek tussen Oostvlaams en Brabants
Dialect of region between East Flanders and Brabant dialects
⑬ Dialect in het gebied van de grote rivieren (Zuidgelders)
Dialect of the river region
⑭ Noordbrabants en Noordlimburgs
Dialect of Noord-Brabant and northern Limburg
⑮ Brabants
Dialect of Brabant
⑯ Dialect van de streek tussen Brabants en Limburgs
Dialect of region between Brabant and Limburg dialects
⑰ Limburgs
Dialect of Limburg
⑱ Veluws
Dialect of the Veluwe region
⑲ Gelders-Overijssels
Dialect of Gelderland and western Overijssel
⑳ Twents-Graafschaps
Dialect of western Twente and eastern Graafschap
㉑ Twents
Dialect of Twente
㉒ Stellingwerfs
Dialect of the Stellingwerf region
㉓ Zuiddrents
Dialect of southern Drenthe
㉔ Middendrents
Dialect of central Drenthe
㉕ Kollumerlands
Dialect of Kollumerland
㉖ Gronings en Noorddrents
Dialect of Groningen and northern Drenthe
Fries
West-Terschelling
Nes
Oosterburen
Roodeschool
Emden
Delfzijl
Dokkum
Kollum
Groningen
Winschoten
Franeker
Harlingen
Leeuwarden
Drachten
Bolsward
Sneek
Heerenveen
Appelscha
Assen
Staveren
Steenwijk
Emmen
Hoogeveen
Meppel
Den Burg
Den Helder
Medemblik
Schagen
Enkhuizen
Urk
Hoorn
Egmond aan Zee
Volendam
Wormerveer
Kampen
Ommen
Zwolle
Vriezenveen
Haarlem
Amsterdam
Muiden
Harderwijk
Nijverdal
Bentheim
Amstelveen
Deventer
Hengelo
Enschede
Katwijk aan Zee
Apeldoorn
Amersfoort
Alphen aan de Rijn
Scheveningen
's-Gravenhage
Utrecht
Otterlo
Zutphen
Woerden
Gouda
Arnhem
Doetinchem
Winterswijk
Schoonhoven
Rotterdam
Brielle
Geldermalsen
Tiel
Nijmegen
Kleve
Sliedrecht
Oud-Beijerland
Oss
Middelharnis
's-Hertogenbosch
Willemstad
Zierikzee
Sint Maartensdijk
Breda
Tilburg
Middelburg
Goes
Bergen op Zoom
Venraij
Baarle-Nassau
Eindhoven
Venlo
Valkenswaard
Sluis
Terneuzen
Hulst
Turnhout
Oostende
Brugge
Weert
Roermond
Zelzate
Antwerpen
Herentals
Sint-Niklaas
Gent
Mechelen
Sittard
Roeselare
Genk
Hasselt
Heerlen
Ieper
Oudenaarde
Leuven
Kerkrade
Kortrijk
Ninove
Brussel (Bruxelles)
Maastricht
Geraardsbergen
Tienen
Sint-Truiden
Aachen



1	**Zuidhollands** *Dialect of Zuid-Holland*
2	**Kennemerlands** *Dialect of Kennemerland*
3	**Waterlands** *Dialect of Waterland*
4	**Zaans** *Dialect of Zaan region*
5	**Westfries–Noordhollands** *Dialect of northern Noord-Holland*
6	**Utrechts–Alblasserwaards** *Dialect of the province of Utrecht and the Alblasserwaard region*
7	**Zeeuws** *Dialect of Zeeland*
8	**Westhoeks** *Dialect of region between Holland and Brabant dialects*
9	**Westvlaams en Zeeuwsvlaams** *Dialect of West Flanders and Zeeuws-Vlaanderen*
10	**Dialect van de streek tussen West– en Oostvlaams** *Dialect of region between West and East Flanders dialects*
11	**Oostvlaams** *Dialect of East Flanders*
12	**Dialect van de streek tussen Oostvlaams en Brabants** *Dialect of region between East Flanders and Brabant dialects*
13	**Dialect in het gebied van de grote rivieren (Zuidgelders)** *Dialect of the river region*
14	**Noordbrabants en Noordlimburgs** *Dialect of Noord-Brabant and northern Limburg*
15	**Brabants** *Dialect of Brabant*
16	**Dialect van de streek tussen Brabants en Limburgs** *Dialect of region between Brabant and Limburg dialects*
17	**Limburgs** *Dialect of Limburg*
18	**Veluws** *Dialect of the Veluwe region*
19	**Gelders–Overijssels** *Dialect of Gelderland and western Overijssel*
20	**Twents–Graafschaps** *Dialect of western Twente and eastern Graafschap*
21	**Twents** *Dialect of Twente*
22	**Stellingwerfs** *Dialect of the Stellingwerf region*
23	**Zuiddrents** *Dialect of southern Drenthe*
24	**Middendrents** *Dialect of central Drenthe*
25	**Kollumerlands** *Dialect of Kollumerland*
26	**Gronings en Noorddrents** *Dialect of Groningen and northern Drenthe*
27	**Fries** *Frisian language*
28	**Bildts, Stadfries, Midslands, Amelands** *Dialects of Het Bildt, Frisian cities, Midsland, and Ameland Island*
-----	**Taal– en dialectgrenzen** *Linguistic and dialect boundaries*
----7----	**Andere isoglossen** *Other isoglosses*