THE DUTCH–GERMAN BORDER: RELATING LINGUISTIC, GEOGRAPHIC AND SOCIAL DISTANCES

FOLKERT DE VRIEND, CHARLOTTE GIESBERS,
ROELAND VAN HOUT AND LOUIS TEN BOSCH

Abstract *In this paper we relate linguistic, geographic and social distances to each other in order to get a better understanding of the impact the Dutch-German state border has had on the linguistic characteristics of a sub-area of the Kleverlandish dialect area. This area used to be a perfect dialect continuum. We test three models for explaining today's pattern of linguistic variation in the area. In each model another variable is used as the determinant of linguistic variation: geographic distance (continuum model), the state border (gap model) and social distance (social model). For the social model we use perceptual data for friends, relatives and shopping locations. Testing the three models reveals that nowadays the dialect variation in the research area is closely related to the existence of the state border and to the social structure of the area. The geographic spatial configuration hardly plays a role anymore.*

I. INTRODUCTION

The Dutch-German state border south of the river Rhine was established in 1830. Before that time, the administrative borders in this region frequently changed. The Kleverlandish dialect area, which extends from Duisburg in Germany to Nijmegen in The Netherlands, crosses the state border south of the Rhine. The area is demarcated by the Uerdingen line in the south, the diphthongisation line of the West Germanic 'i' in the West, and the border with the Low Saxon dialects of the *Achterhoek* area in the North-East. The geographic details of the area can be found in Figure 1 (the state border is depicted with a dashed-dotted line).

Figure 1. Geographic details of the Kleverlandish dialect area (Cornelissen 2003).

Hinskens, Kallen and Taeldeman (2000) pointed out that European state borders that cut across old dialect continua had a strong impact on dialect change. Boberg (2000) investigated the dialects on both sides of the border between Canada and the US and criticised Trudgill's gravity model (Trudgill, 1974). The gravity model says that language varieties may be subject to a 'gravity-like law'. In the model, population size plays the role of mass, so that settlements with large populations are particularly likely to adopt each other's changes. However, the effects of state borders are not taken into account in the model.

The Kleverlandish area, in its original form, is a prototypical example of a dialect continuum. There were no natural borders or sharp dialect borders. Kremer (1984; 1990) and Niebaum (1990) discussed the increased significance of the Dutch-German state border as a dialect border. Later, both Heeringa, Nerbonne, Niebaum, Nieuweboer and Kleiweg (2000) and Giesbers (2008) quantatively examined the effect of the Dutch-German state border. The area Heeringa et al. (2000) were interested in is situated north of the Rhine, around

Figure 2. Position of the research area of Giesbers (2008), relative to The Netherlands. The map also includes provincial borders.

the German town of Bentheim. Giesbers (2008) investigated a sub-area of the Kleverlandish dialect area that is situated south of the Rhine between the Dutch towns of Nijmegen (in the north) and Venray (in the south), as depicted in Figure 2. Both studies showed that the political border had a significant impact on the dialect continuum and separated the Dutch from the German dialects.

In this paper we test three models to explain the linguistic characteristics of the area depicted in Figure 2. The three models are the continuum model, the gap model and the social model.

When closely related language varieties in an area form a continuum, their distribution is marked by a direct, monotonic relationship between geographic and linguistic distance. Chambers and Trudgill (1985) formulate this as:

'If we travel from village to village, in a particular direction, we notice linguistic differences which distinguish one village from another. Sometimes the differences will be larger, and sometimes smaller, but they

will be cumulative. The further we get from our starting point, the larger the differences will become.'

A cumulative model implies that the linguistic distance can be estimated fairly precisely on the basis of geographic distance: the larger the geographic distance, the larger the linguistic distance. The default model for a perfect dialect continuum can be defined as follows:

continuum model
linguistic distance = f (geographic distance)

f is a monotonic increasing function in this model, and, in a particularly simple case, a linear function. An error term could be added to the model since there may be some variability across the area. Hard (1972) simulated a continuum model for the Rhenish Fan (cf. Bloomfield, 1933), a famous example of a cumulative dialect continuum consisting of a stepwise isoglossic structure, with a random variability component.

The Kleverlandish dialects in The Netherlands and Germany came under the hegemony and influence of the two respective standard languages after the establishment of the state border in 1830. In addition, political, administrative and cultural developments became different in the area that was divided then over two countries. What was the impact of the state border on the dialects? The central research hypothesis in Giesbers (2008) is that the Dutch-German state border has given rise to a linguistic gap in the Kleverlandish dialect continuum. If this is true, the continuum model could be extended by adding a constant value to *f* representing the state border gap. In its most outspoken form we may assume that the gap became the main determinant of the linguistic distance, overshadowing remaining differences and patterns of dialect variation. Such an outspoken model can be defined as:

gap model
linguistic distance = f (gap)

The gap can only have two values in this model. It is zero when two locations are not separated by the state border. It has a specific, fixed value when two locations are separated by the state border. Again, an error term could be added to account for variability across the area.

The usefulness of the continuum model and the gap model is supported by historical marriage data collected in the research area by Giesbers (2008). In the period 1850–1870 30 per cent of the marriages were mixed, indicating a continuous socio-geographic network structure, with no notable effect of the state border. Nowadays the number of mixed marriages has dropped to less than 5 per cent, indicating that the state border has formed a gap in marital exchange. Intensive contact between speakers is an essential condition for dialects to

122

continue to cohere. Could the linguistic distances in our research area be more directly related to social distances than to geographic distances? Assuming that marriage data reflect the intensity of cross-border contact, the question arises which other properties correlate with the social contact structure of the area. These properties could contribute to our understanding of the new linguistic structure in the Kleverlandish area. No objective data, however, are available on the social structure of the research area. Alternatively, one can ask people living in the area about how they perceive the geographic distribution of relevant social phenomena. We will look at perceptual social contact data for friends, relatives and shopping places. An explanatory model based on these social data can be defined as follows:

social model
linguistic distance = f(social distance)

The social model is similar to the continuum model but takes social distance as the determinant of linguistic distance instead of geographic distance. *f* is again a monotonic increasing function in this model and an error term could be added here too.

We want to test the explanatory power of the three models. In section 2 we first describe the data that will be used to test the models. These data were collected along the state border in the Kleverlandish area. In section 3 we derive distances from the data and test the three models. We first compare the continuum and gap models and then check to see if the social model has more explanatory power. In section 4 we discuss the results followed by a conclusion.

2. DATA COLLECTION

We collected data for 10 locations in the northern part of the Kleverlandish dialect area. Five locations on each side of the border were selected, as is shown in Figure 3. The area does not contain any natural borders and the 10 locations lie close to the state border. In the selection process, locations on the Dutch side of the border were paired with similar locations on the German side of the border based on information about population size, infrastructure and distance to the border. This resulted in five cross-border pairs of locations. The town centres of both Huelm (Germany) and Siebengewald (The Netherlands) for instance are 3.7 kilometres from the border. Two locations with a larger distance to the border are Gennep and Goch: 17 kilometres. The population size of the locations varies between 721 (Huelm) and 19,961 (Goch) in Germany, and 777 (Ven-Zelderheide) and 11,403 (Groesbeek) in The Netherlands. For these 10 locations we collected linguistic and social data.

Figure 3. The 10 locations on both sides of the Dutch-German state border; cross-border paired locations are connected by lines. The rightmost line is the state border.

2.1 Linguistic data

The linguistic data were collected by recording 20 respondents in the research area. In each of the 10 locations two dialect speakers (one younger and one older speaker) that spoke their dialect on a daily basis were asked to give their dialect words (nouns) for a list of a hundred concepts related to everyday life. We only used the data from the 10 older speakers (> 60 years) since these data are expected to be closer to typical dialect speech and less likely to have been influenced by the standard languages of the two areas (Dutch and German) (cf. Giesbers, 2008). The list of concepts we used was developed by Van Bezooijen to measure lexical and phonetic-phonological variation between closely related (Germanic) language varieties (cf. Heeringa, Nerbonne, Van Bezooijen and Spruit, 2007).

The recordings were transcribed on a detailed phonetic level. Table 1 shows an example of the phonetic transcriptions. It shows the pronunciation for the concept *aardappel* ('potato') as realised by the older respondent of the location

124

Table 1. Example of the phonetic transcriptions used.

Location	Concept	Phonetic transcription
Gennep (> 60 yrs)	Aardappel	ERdAp@l

Gennep. The transcription system used was a combination of German and Dutch X-SAMPA.

Many subtle differences were transcribed, such as the voicing of fricatives, differences in place and manner in [r] pronunciations and the height, rounding and length of vowels. The frequency of each phoneme and phoneme cluster in the dataset was representative for the range of sounds in the dialect speech (cf. Giesbers, 2008).

2.2 Social data

The social data were collected using a questionnaire that was filled in by 268 respondents from the 10 locations in the research area. The respondents belonged to two age groups (i.e., 30–40 years, 60 years and over) and they were balanced for age and gender. The respondents were recruited through a regional Dutch and a regional German newspaper. Both newspapers published an article about the research project together with the questionnaire. Of the 340 informants that returned the questionnaire, 268 also met the selection criteria: they grew up in the location where they were living today and they spoke the local dialect.

Respondents were asked to indicate and rank in which locations they had friends and family and where they went shopping. These three types of data give information on the degree of social contact between the locations in the area. Respondents were asked to name and rank order 10 locations for friends and family, and five for shopping locations.

3. METHODOLOGY AND RESULTS

We first obtained geographic distances and distances for the linguistic and social data to test the three models. Next, we performed Multi-Dimensional Scaling on the distances and plotted the two-dimensional result. We also plotted the distances after transforming them to a similar scale. By visually comparing the plots to each other we can interpret the explanatory power of the three models. To compare the different distance topologies (Pearson product-moment) correlations were also used. Finally, the three models were tested statistically using regression analysis.

3.1 The continuum model versus the gap model

We used a Dutch route planner website by the *Dutch* Automobile Association ('ANWB') to obtain geographic distances. The geographic distances obtained

Table 2. Range and mean for the geographic distances.

	Dutch	German	Dutch–German
Range	4.90–22.00	3.70–26.10	3.40–20.00
Mean	12.14	12.98	10.25

were not distances 'as the crow flies' but the shortest travel distance when following the normal road infrastructure. Travel distances are not equal to distances 'as the crow flies' and Gooskens (2005) for instance notes that in Norway they can be quite different because of the mountainous characteristics of the country. In our research area, however, there are no natural borders and the travel distances we obtained are quite comparable to distances 'as the crow flies'. The longest geographic distance in our data set is the distance between Groesbeek and Huelm; 26.1 kilometres. The smallest geographic distance is between Goch and Huelm; 3.4 kilometres.

The locations were selected in such a way that they were geographically balanced, but the border may have had an effect on the actual connectedness between locations. To validate our sample we looked at the Dutch, the German and the Dutch-German distances separately. Table 2 shows the range and mean for the three types of geographic distances. The distances are evenly distributed over the three groups of locations and no significant differences were found ($F_{(2.42)} = 0.787$, $p = 0.462$).

We used the Levenshtein method (Kruskal, 1983) to obtain linguistic distances. The Levenshtein method computes linguistic distances between locations based on all pairs of phonetic transcriptions (strings). The distance between the two phonetic transcriptions involved is calculated on the basis of the minimum number of operations needed for string A to be transformed into string B. The three types of operations permitted are insertion, deletion or substitution of characters.

We calculated Cronbach's Alpha (Cronbach, 1951) with the dialectometric software RuG/L04 to make sure that the number of words used is a sufficient basis for the Levenshtein analysis. With Cronbach's Alpha we can measure the minimum reliability of our Levenshtein distance measurements when applied to our dataset. In the social sciences values higher than 0.70 are considered sufficient (Nunnally, 1978). Cronbach's Alpha was 0.93 for our dataset of a hundred concepts.

Next, we used RuG/L04 to compute the linguistic distances with the Levenshtein method. We used the simplest version of the method in which phonetic overlap is binary: non-identical phones contribute to phonetic distance, identical ones do not. No feature-based segment distance table was used and no weights were assigned to the different types of operations. The RuG/L04
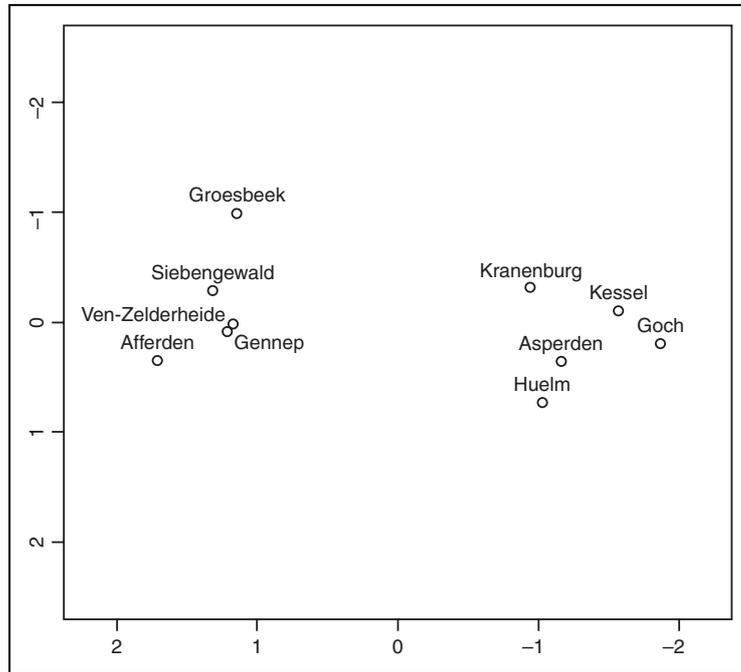
Figure 4. Two-dimensional MDS plot of the linguistic distances.

software was used to obtain a 10 by 10 dissimilarity matrix for the locations. The diagonal of this matrix is always zero and only half of the matrix is used since the lower half is the mirror image of the upper half.

To test the continuum model we first performed Multi-Dimensional Scaling (MDS) on the distances. Point sets are projected and visualised in two-dimensional space with MDS (see for instance Nerbonne, Heeringa and Kleiweg, 1999; Spruit, 2008). The MDS analysis on the linguistic distances (Alscal) gave a nearly perfect two-dimensional solution (Stress = 0.050, RSQ = 0.990). Both the x-axis and the y-axis are mirrored in this plot, depicted in Figure 4, to make it easier to visually compare the result of the analysis to the topology of the distances 'as the crow flies' depicted in Figure 3. The topology of the linguistic distances is clearly different from the geographic distances 'as the crow flies'.

The most remarkable outcome is that distances between locations within the *same* country are always (much) smaller than distances between locations that are in *different* countries. The continuum model does not apply to the linguistic distances. It is the gap model that seems to give the required topology.

127

Table 3. Correlation values for linguistic distance and the three types of location pairs; one-tailed p values.

	Geographic distance Dutch (N = 10)	Geographic distance German (N = 10)	Geographic distance Dutch-German (N = 25)
Linguistic distance	0.495 (p = 0.073)	0.577 (p = 0.041)	0.098 (p = 0.321)

To investigate the topological structures and their relations in more detail we also calculated the correlations between the geographic and the linguistic distances for all 45 pairs of locations. The continuum model predicts a high correlation, the gap model no correlation at all. The correlation is 0.291, with a one-tailed p value of 0.038. We opted for a one-tailed test, since the values must be positive if there is a correlation. Since the distances are not independent measurements, the probability values of the correlation coefficients were calculated using the Mantel test (Mantel, 1967). The conclusion is that the correlation is significant but low. Geographic distances hardly play a role in explaining the linguistic distances. Does this mean we have to reject the continuum model completely?

We divided the 45 location pairs into three groups to test the continuum model in more detail: 10 Dutch pairs, 10 German pairs and 25 Dutch-German pairs. The gap model predicts that the linguistic distances within Dutch-German pairs should be relatively large and constant. The linguistic distances in the Dutch pairs and in the German pairs may be arbitrary, but under the assumption of the remains of a continuum model, a relationship between geography and linguistic distance may still hold. The correlations for the three groups of pairs are given in Table 3. The correlation coefficients were calculated using a classical statistical test instead of the Mantel test because of the asymmetric nature of the three distance matrices.

The correlation for the Dutch-German location pairs is not significant. The correlations for the Dutch and German location pairs are clearly higher, although the correlation for the Dutch pairs fails to reach significance. Given the low number of location pairs, the statistical test of the correlations does not have much power to detect lower correlations.

Next, we visualised the relationship between the geographic and linguistic distances after having transformed the linguistic distances to a similar scale as the geographic distances (we used the same maximum). Standardising the scaling helps in interpreting the relationship and does not change the intrinsic structural characteristics of the distances. We now expect to find the location pairs adhering to the continuum model, in which the linguistic distance equals the geographic distance, on the diagonal. The scatter plot is given in Figure 5. The three groups of location pairs were given different symbols.
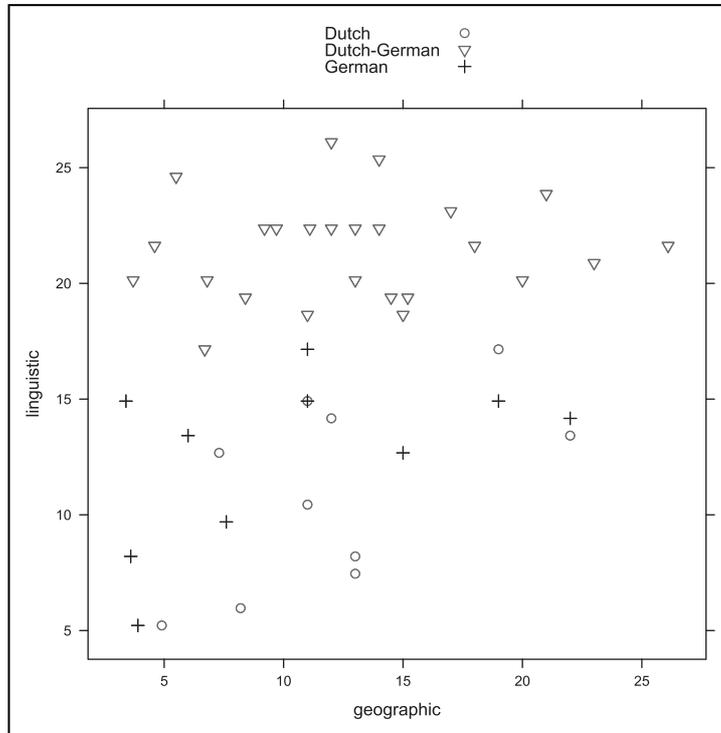
128

Figure 5. Linguistic distance by geographic distance for the three groups of location pairs on the same scale of magnitude.

The Dutch-German pairs (the triangles) show a distinct pattern. Their geographic distance varies between 3 and 26 kilometres, but their (scaled) linguistic distance ranges from 17 to 26 kilometres. There is no further explanation for the variation within the range found. The smallest distance in the Dutch-German pairs (17) is equal to the largest distance found for the Dutch and German pairs. The gap model clearly applies to the Dutch-German pairs. The distances for the Dutch and German pairs are smaller, and they are roughly on the diagonal. For both the Dutch pairs and the German pairs the linguistic distance on average increases as the geographic distance increases. This relationship was also reflected in the higher correlation values in Table 3. The continuum model no longer applies to the whole research area, but only (moderately) to the within-country linguistic distances. Location pairs across the Dutch-German border are nowadays separated by a clear linguistic gap.

Our interpretation of the plot in Figure 5 is further corroborated by the high correlation between linguistic distance and the gap as a nominal variable
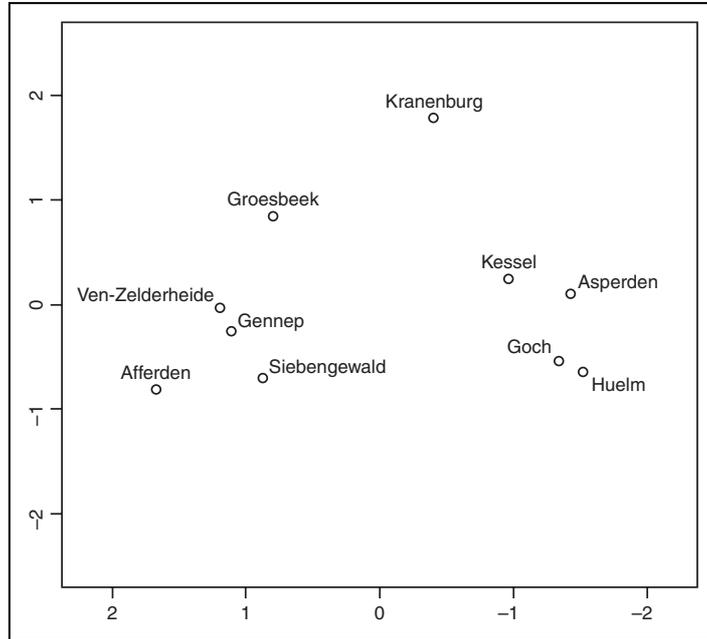
129

Figure 6. Two-dimensional MDS plot of the social variable friends.

(two locations are separated by the border, yes or no): 0.850 (N = 45, p = 0.000). Looking at the strong division between the Dutch-German location pairs and the Dutch or German location pairs in Figure 5, this high correlation was to be expected. When entering both the gap and geographic distance in a (linear) regression analysis, the effect of geographic distance turns out to be not significant, even when we use a one-tailed test. Our conclusion is that we need to reject the continuum model and accept the gap model.

3.2 The gap model versus the social model

Can other types of data tell us more about the linguistic structure in the Kleverlandish area? In this section we test the social model to see if it has more explanatory power than the gap model. We used the social data Giesbers (2008) collected about friends, family, and shopping for this. In these data no effect was found for the respondent variables *gender* and *age*. The weighted data were computed per location and three separate 10 by 10 dissimilarity matrices were obtained for each of the social variables. MDS returned excellent results for a two-dimensional representation for all three variables. The two-dimensional

130

Table 4. Correlation values for linguistic distance and the distance for shopping, family or friends; one-tailed p values.

	Shopping distance (N = 45)	Family distance (N = 45)	Friends distance (N = 45)
Linguistic distance	0.623 (p = 0.001)	0.737 (p = 0.001)	0.818 (p = 0.000)

MDS results for *friends* (Stress = 0.070, RSQ = 0.972) are given in Figure 6. Again we mirrored the x-axis.

The plot in Figure 6 shows that the general topology for *friends* resembles the MDS plot for the linguistic distances depicted in Figure 4. Next, we checked the correlations for linguistic distance and each of the three social variables. The correlations can be found in Table 4. Again we used the Mantel test to determine their statistical significance.

The social variable *friends* shows the strongest correlation, but the correlation for the other two variables is also clearly present. Linguistic distance is not a property on its own, but is embedded in the social structure of the research area.

To get a more detailed picture of the relationship between the variable *friends* and linguistic distance, we used the same visualisation method as applied in Figure 5. This time, we scaled up both the *friends* distances and the linguistic distances to the same maximum value of the geographic distances. Next, we plotted the relation between the *friends* distances and the linguistic distances in Figure 7, for the three groups of pairs: Dutch, German and Dutch-German.

Figure 7 shows that all location pairs are on the diagonal this time, although the relationship is not perfect, but scattered. The diagonal pattern applies to all three groups of pairs. This means that the two variables really share a similar configuration or topology. Figure 7 supports our finding that the linguistic distances appear to belong to an overarching socio-geographic pattern that has developed in the research area over the last two centuries.

A regression analysis was also performed to see if a combination of social variables gives an even better result. However, no better model could be found. The simple model in which *friends* is the only explanatory variable still gave the best result. We also entered both *friends* and geographic distances in a (linear) regression analysis. The effect of geographic distance here turned out not to be significant.

The explanatory power of the gap model (explained variance: 0.723) is higher than that of the social model (explained variance: 0.668). However, the difference between the two models is not big enough to make a final choice for one of them. Combining the two variables in a regression analysis returned a better model with a higher amount of explained variance (0.781). Both variables are significant, although the strongest predictor is the gap variable. This is
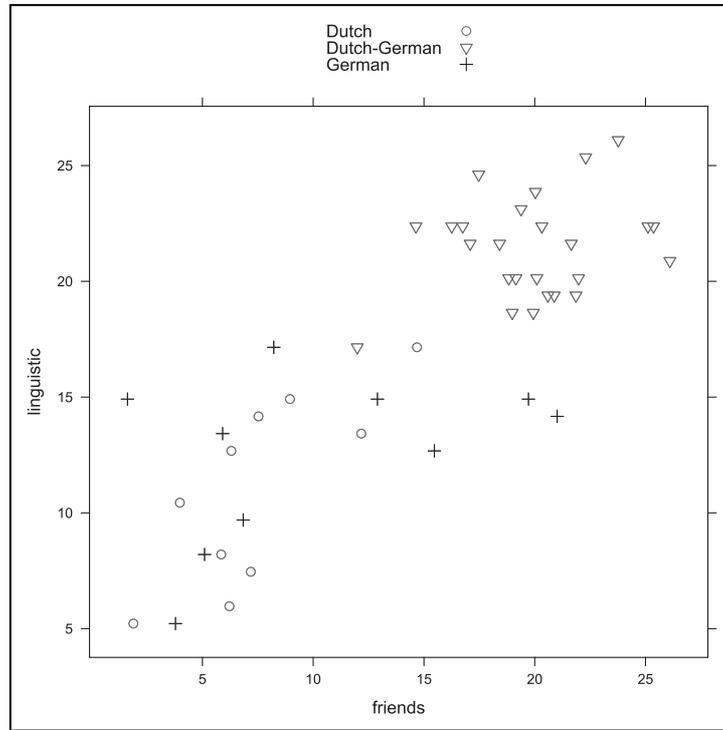
Figure 7. Linguistic distance by friends distance for the three groups of location pairs on the same scale of magnitude.

not surprising since the correlation value for the gap was higher. Explaining linguistic distance in the Kleverlandish area clearly requires combining the assumption of a gap and data about the social structure of the area.

4. DISCUSSION AND CONCLUSIONS

In this paper we tested three models to explain the pattern of linguistic variation in a sub-area of today's Kleverlandish dialect area: the continuum model, the gap model and the social model. We first compared the continuum model to the gap model and concluded that the continuum model nowadays no longer applies to the whole research area. It applies only moderately to the within-country locations. The gap model explains the linguistic distances between the cross-border pairs of locations very precisely. There, the state border has become the main determinant of the linguistic distances, overshadowing remaining differences and patterns of dialect variation. Next, we checked to see if the

social model had more explanatory power. We therefore looked at contact data about friends, relatives and shopping places and found that linguistic distance nowadays is embedded in the socio-geographic structure of the research area. Especially the friends data very much resembled the linguistic data. Combining the gap model and the social model turned out to be the most successful way to explain the linguistic distances. Clearly, the dialect variation in our research area is closely related to the existence of the state border and to the social structure of the area. The geographic spatial configuration hardly plays a role anymore.

We have shown that political and social contact variables nowadays are more important than geography in explaining the topological structure for linguistic variation in our research area. The pattern of linguistic variation that existed about 200 years ago in the Kleverlandish area was the result of human contact. Without natural or political borders or dominant population centres, the Kleverlandish dialect area developed into a dialect continuum in which linguistic distance was closely related to geography. The cohesive social system between locations across the state border diverged after the establishment of the state border in the area in 1830, and as a result, so did the dialect variation. Others have also pointed at the importance of such political and social variables (cf. Bailey, Wikle, Tillery and Sand, 1993; Boberg, 2000; Horvath and Horvath, 2001). What other variables might help to come to a better understanding of the topological structure of linguistic variation? Giesbers (2008) shows that the divergence of dialect variation in the Kleverlandish area was further stimulated by the influence the two standard language varieties, Dutch and German, had on the dialects. Gooskens (2005) points at the influence mass media and demographic factors like migration and immigration might have on the spreading of linguistic variables. Also, social contact data about mobility and telecommunication possibilities could be helpful.

REFERENCES

G. Bailey, T. Wikle, J. Tillery and L. Sand (1993), 'Some patterns of linguistic diffusion', *Language Variation and Change*, 3(3), 241–264.

L. Bloomfield (1933), *Language* (New York).

C. Boberg (2000), 'Geolinguistic diffusion and the U.S.-Canada border', *Language Variation and Change*, 12(1), 1–24.

J. K. Chambers and P. Trudgill (1998), *Dialectology* (Cambridge).

G. Cornelissen (2003), *Kleine Niederrheinische Sprachgeschichte (1300–1900): eine regionale Sprachgeschichte für das deutsch-niederländische Grenzgebiet zwischen Arnheim und Krefeld* (Geldern – Venray).

L. J. Cronbach (1951), 'Coefficient alpha and the internal structure of tests', *Psychometrika*, 16, 297–334.

F. De Vriend, J. P. Kunst, L. Ten Bosch, C. Giesbers and R. Van Hout (2008), 'Evaluating the relationship between linguistic and geographic distances using a 3D visualization', *Proceedings of the sixth international conference on language resources and evaluation* (Marrakech), 2212–2215.

C. Giesbers, R. Van Hout and R. Van Bezooijen (2006), 'De geografische en perceptieve afstand tussen dialecten: de rol van de staatsgrens in het Kleverlandse dialectcontinuüm', *Artikelen van de vijfde sociolinguïstische conferentie* (Delft), 175–186.

C. Giesbers (2008), 'Dialecten op de grens van twee talen. Een dialectologisch en sociolinguïstisch onderzoek in het Kleverlands dialectcontinuüm' (Ph.D. thesis, Radboud University Nijmegen).

C. Gooskens, Charlotte (2005). 'Traveling time as a predictor of linguistic distance', *Dialectologia et Geolinguistica*, 13, 38–62.

G. Hard (1972), 'Ein geographisches Simulationsmodell für die rheinische Sprachgeschichte', in E. Essen and G. Wiegelmann, eds, *Festschrift Matthias Zender, Studien zur Volkskultur, Sprache und Landesgeschichte* (Bonn), 5–29.

W. Heeringa, J. Nerbonne, H. Niebaum, R. Nieuweboer and P. Kleiweg (2000), 'Dutch-German contact in and around Bentheim. Languages in Contact', *Studies in Slavic and General Linguistics*, 28, 145–156.

W. Heeringa (2004), 'Measuring dialect pronunciation differences using Levenshtein distance' (Ph.D. thesis, Rijksuniversiteit Groningen).

W. Heeringa, J. Nerbonne, R. Van Bezooijen and M. Spruit (2007), 'Geografie en inwoneraantallen als verklarende factoren voor variatie in het Nederlandse dialectgebied', *Nederlandse Taal- en Letterenkunde*, 123(1), 70–82.

F. Hinskens, J. L. Kallen and J. Taeldeman (2000), 'Dialect convergence and divergence across European borders', *International Journal of the Sociology of Language*, 145.

B. M. Horvath and R. J. Horvath (2001), 'A multilocality study of a sound change in progress: The case of /l/ vocalization in New Zealand and Australian English', *Language Variation and Change*, 13(1), 37–57.

L. Kremer (1984), 'Die niederländisch-deutsche Staatsgrenze als subjektive Dialektgrenze', *Driemaandelijkse Bladen*, 36, 76–83.

L. Kremer (1990), 'Kontinuum oder Bruchstelle? Zur Entwicklung der Grenzdialekte zwischen Niederrhein und Vechtegebiet', in L. Kremer and H. Niebaum, eds, *Grenzdialekte. Studien zur Entwicklung kontinentalwestgermanischer Dialektkontinua* (Hildesheim, Zürich and New York), 85–123.

J. B. Kruskal (1983), 'An overview of sequence comparison: time warps, string edits, and macromolecules', *SIAM Rev.*, 25, 201–237.

N. Mantel (1967), 'The detection of disease clustering and a generalized regression approach', *Cancer Research*, 27, 209–220.

J. Nerbonne, W. Heeringa and P. Kleiweg (1999), 'Edit Distance and Dialect Proximity', in D. Sankoff and J. B. Kruskal, eds, *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison* (Stanford), 5–15.

H. Niebaum (1990), 'Staatsgrenze als Bruchstelle? Die Grenzdialekte zwischen Dollart und Vechtegebiet', in L. Kremer and H. Niebaum, eds, *Grenzdialekte. Studien zur Entwicklung kontinentalwestgermanischer Dialektkontinua* (Hildesheim, Zürich and New York), 49–83.

J. Nunnally (1978), *Psychometric theory* (New York).

RuG/L04, Software for dialectometrics and cartography. http://www.let.rug.nl/ kleiweg/L04/.

M. Spruit (2008), 'Quantitative perspectives on syntactic variation in Dutch dialects' (Ph.D. thesis, University of Amsterdam).

P. Trudgill (1974), 'Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography', *Language in Society*, 2, 215–246.