The Dutch-German dialect border: relating linguistic, geographic and perceptual distances

Folkert de Vriend (1), Charlotte Giesbers (2), Roeland van Hout (2) & Louis ten Bosch (1)

(1) Centre for Language and Speech Technology, Radboud University Nijmegen
(2) Centre for Language Studies, Radboud University Nijmegen

1 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 dialects spoken in the north of the state border area belong to the Kleverlandish dialect continuum, which extends from Duisburg in Germany to Nijmegen in the Netherlands as is shown in Fig. 1. The heart of the Kleverlandish dialect continuum is located in Germany, but parts of the Dutch provinces of Gelderland, Limburg and Brabant also belong to this area. The Kleverlandish dialect area is defined by the Uerdingen Line in the south, the diphthongization 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 Dutch-German national border was drawn in 1830 right through this dialect continuum. The geographic details of the Kleverlandish dialect area become clear from Fig. 1. The figure shows how the area crosses the Dutch-German state border (the state border is depicted with a dotted line).

Figure 1: Geographic details of the Kleverlandish dialect area

This paper focuses on data collected in a sub-area of the Kleverlandish dialect continuum by Giesbers (2008). The area lies on the Dutch-German state border between the Dutch towns of Nijmegen (in the north) and Venray (in the south) as is depicted in Fig. 2. The nearest larger location on the German side of the state border is the town of Kleve.

The dialect speakers in the research area can still use their dialect as a lingua franca for cross-border communication, but today the number of dialect speakers is rapidly decreasing (see Giesbers 2008).

Figure 2: Position of the research area, relative to the Netherlands

The Kleverlandish area, in its original form, is a prototypical example of a dialect continuum. There are no natural borders nor sharp dialect borders. When closely related language varieties in an area form a continuum, their distribution is marked by a direct, monotonous relationship between geographic and linguistic distance, as is formulated by Chambers and Trudgill (1998: 5) in the following way:

> "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 precise on the basis of geographic distance: the larger the geographic distance the larger the linguistic distance. Perhaps there is some variability over the area, but the default model for a perfect dialect continuum can be defined as follows:

dialect continuum model
linguistic distance = f(geographic distance)

In this model f is a monotonous function, and, in a particularly simple case, a linear function.

After the establishment of the state border in 1830 the Kleverlandish dialects in the two countries came under the heading and influence of the two respective standard languages, Dutch and German. In addition, political, administrative and cultural developments were different in the area that was divided now over two countries. What could have been the impact on the dialects? The central research hypothesis in Giesbers (2008) was that the German-Dutch state border has given rise to a linguistic gap in the Kleverlandish dialect continuum. If that applies the dialect continuum model could be expanded 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:

dialect gap model
linguistic distance = f(gap)

The gap can only have two values. It is zero when the two locations involved are not separated by the state border. It has a specific, fixed value when the state border is involved. Again, an error term could be added to account for smaller differences. The usefulness of the two models is supported by historical marriage data collected in the research area by Giesbers (2008). In the period 1850-1870 30% of the marriages were mixed, indicating a continuous socio-geographical network structure, with no real state border rupture. Nowadays the figures of mixed marriages have dropped to a percentage of less than 5%, pointing out that the state border nowadays constitutes a gap in marital exchange. If marriage data reflect the intensity of cross-border contact and contact exchange, the question arises which other properties correlate to the communicative or social contact structure of the area.

Intensive contacts between speakers are an essential condition for dialects to not diverge over time. No objective data, however, are available on the socio-geographic communicative or contact structure of the research area. The alternative is to ask people living in the research area about their perceptions of relevant contact phenomena. By collecting perceptual data of people living in the area, we can investigate if perceptual socio-geographic distances run parallel to or are more directly related to the linguistic distances found than to the actual geographic distances. Linguistic distances may better resemble social contact factors like the locations where one’s friends or relatives are living or the locations where one goes shopping. Perceptual data may attribute to our understanding of the new linguistic structures of the dialects in the Kleverlandish dialect area. Given that perspective, the linguistic distances as perceived in the area need to be included as well, as a possible link between the linguistic data and the perceptual socio-geographic data. This perspective can be summarized under an explanatory model based on perceptual distances:

perceptual distance model
linguistic distance = f(perceptual distance)

All three models use the concept of distance. Purely formally, a distance is a mathematical concept that attains to each pair of points (p1, p2) a real number D(p1, p2) such that the following three properties are being met:

- D is 0 or positive, D=0 only if p1=p2, else D>0
- D is symmetric
- D obeys the triangle inequality

Geographic measures “as the crow flies” meet these three properties and are therefore interpretable as genuine (in this case two-dimensional) distances. This is not necessarily true for the linguistic and perceptual distances that we discuss in section 3 of this paper. These "distances" are in fact dissimilarities and dissimilarities do not need to meet the triangle inequality which states that for any triangle, the length of a given side must be less than or equal to the sum of the other two sides but greater than or equal to the difference between the two sides.

A common method for analyzing dissimilarities is Multi-Dimensional Scaling (MDS). Point sets with a dissimilarity matrix that does not obey the triangle inequality can still be projected and visualized in a two-dimensional space (see for instance Nerbonne, Heeringa & Kleiweg 1999 or Spruit 2008), which is acceptable only when the reduction leads to minor violations of the two-dimensional distance model (see also De Vriend et al 2008 in which a procedure is described for dealing with this problem in three-dimensional space). Here we will use MDS to analyze the different dissimilarity measures as distances. By visually comparing the plots with each other we are able to test the successfulness of the models. In addition we will use (Pearson product-moment) correlations to compare the different distance topologies.

In section 2 we first describe the data collected by Giesbers (2008) in the Kleverlandish area. In section 3 we derive distances from this data and test and evaluate the three models. We discern (1) geographic distance, (2) linguistic distance, (3) perceptual distance. The perceptual distances consist of perceptual linguistic distances and perceptual socio-geographic distances. The perceptual socio-geographic distances are based on data for friends, relatives, and shopping behaviour.

2 Data collection

Locations

Ten locations in the northern Kleverlandish dialect area were selected, five on each side of the border, as is shown in Fig. 3. The area does not contain any natural borders and the ten locations lie in a connected area close to the state border. In the selection process, locations on the Dutch side of the border were paired with locations on the German side of the border, matching them on size, infrastructure and distance to the border. Five cross-border pairs of locations were selected, as indicated in Figure 3. Hülm (Germany) and Siebengewald (Netherlands) for instance lie both at the border (their centres have a distance of 3.7 kilometres). Two locations with a larger distance to the border are Gennep and Goch. They have a general distance over the road of 17 kilometres. The population size of the locations varies between 721 (Hülm) and 19,961 (Goch) in Germany, and 777 (Ven-Zelderheide) and 11,403 (Groesbeek) in the Netherlands.

Figure 3: The ten locations on both sides of the Dutch-German state border; cross-border paired locations are connected by lines

Linguistic data

The linguistic data was elicited with a dialect questionnaire. In each of the ten locations selected in the research area two dialect speakers were asked to carefully give the dialect words for a list of hundred nominal words related to everyday entities. Heeringa (2000, 2004) concludes that an arbitrary list of 100 words is sufficient to determine linguistic distances between dialect varieties. This list was also used by Heeringa, Nerbonne, Van Bezooijen and Spruit (2007). The list is meant for measuring lexical and phonetic-phonological variation between closely related (Germanic) language varieties. All answers by the respondents were recorded. Only respondents who indicated that they spoke their dialect on an everyday basis were interviewed. Of the two respondents per location one was a younger (30-40 yrs) and one an older (> 60 yrs) dialect speaker. Here we will use only the data from the older speakers since these data are less likely to have been influenced by the standard languages of the two areas (Dutch and German) and are therefore expected to be more close to typical dialect speech. The data from the younger dialect speakers however does not change in any substantial way the conclusions that can be drawn on the data from the older speakers (cf. Giesbers 2008).

The recordings were transcribed on a detailed phonetic level. Table 1 shows an example of the phonetic transcriptions made. It gives the pronunciation for the concept of "aardappel" (potato) as realized by the older respondent of the location Gennep. The transcription system used was a combination of German and Dutch SAMPA.

Location	**Concept**	**Phonetic transcription**
Gennep	Aardappel	ERdAp@l

Table 1: Example of the phonetic transcriptions used

Because of the narrow transcriptions many subtle differences were transcribed, like the voicing of fricatives, differences in place and manner in /r/ pronunciations and vowel height, rounding and length.

Perceptual data
All perceptual data was elicited with a written survey among 268 respondents from the selected ten locations, stratified for age (between 30 and 40, 60 and older) and gender. The informants were fairly evenly distributed over the 40 cells (the numbers varying between 6 and 9). They were recruited through a regional Dutch and German newspaper that both published a longer article about the research project in question and the questionnaire (this happened in 2003). The criteria that had to be met by the informants were specified. 340 informants returned the questionnaire, 268 met the selection criteria. All respondents grew up in the location where they were living today and they spoke the local dialect.

For the perceptual linguistic data the respondents were asked to name ten neighbouring locations with a dialect they considered most similar to their own local dialect (they had map of the whole region available in the newspaper). Subsequently, the respondents were asked to rank order (from 1 to 10) the ten locations they had chosen from most similar to less similar. The same procedure was applied for the socio-geographic data.

The socio-geographic data was divided into three types. All three types tell us more about the amount of social contact between the locations in the research area. Dialect speakers were asked in what locations they had the most friends, the most family and where they went shopping most often. In all three cases, they had to mention and rank order five locations.

3 Results

3.1 Geographic and linguistic distances

For obtaining geographic distances we used a Dutch route planner website by the Dutch automobile association ("ANWB"). The geographic distances were not the distances "as the crow flies" but we took the shortest travel distance when following the normal road infrastructure. Although travel distances are not equal to distances "as the crow flies", they are often quite comparable. The longest travel distance in our data set is the travel distance between Groesbeek and Hülm; 26,1 kilometres. The distance "as the crow flies" between Groesbeek and Hülm however is only 20,7 kilometres. The smallest distance in our data set is between Goch and Hülm; 3.4 kilometres.

To test our sample we looked at the Dutch, the German and the Dutch-German distances separately. In Table 2 the range and mean for the three types of geographic distances are given. No significant differences between the three groups of distances were found ($F(2,42) = 0,787$, $p = 0,462$).

	Dutch	**German**	**Dutch-German**
Range	4,90 – 22,00	3,70 - 26,10	3,40 - 20,00
Mean	12,14	12,98	10,25

Table 2: Geographic distances

It means that the geographic distances are not related to the two countries the locations belong to. The locations were selected in such a way that they were geographically balanced, but the border might have had some particular effect on the actual connectedness between locations. However, the distances are evenly distributed.

We performed an MDS analysis (Alscal) on our geographic distances. A two dimensional solution returned an excellent fit (Stress = 0,024, RSQ = 0,977) The result is plotted in Fig. 6. We mirrored the *x* axis to make it easier to visually compare the result of the MDS analysis to the topology of the distances “as the crow flies” depicted in Figure 3. The plot shows that the topology of the geographic distances we obtained from the route planner is very similar to the topology depicted in Figure 3. The order of the locations is the same and Groesbeek and Kranenburg are slightly separated from the rest.

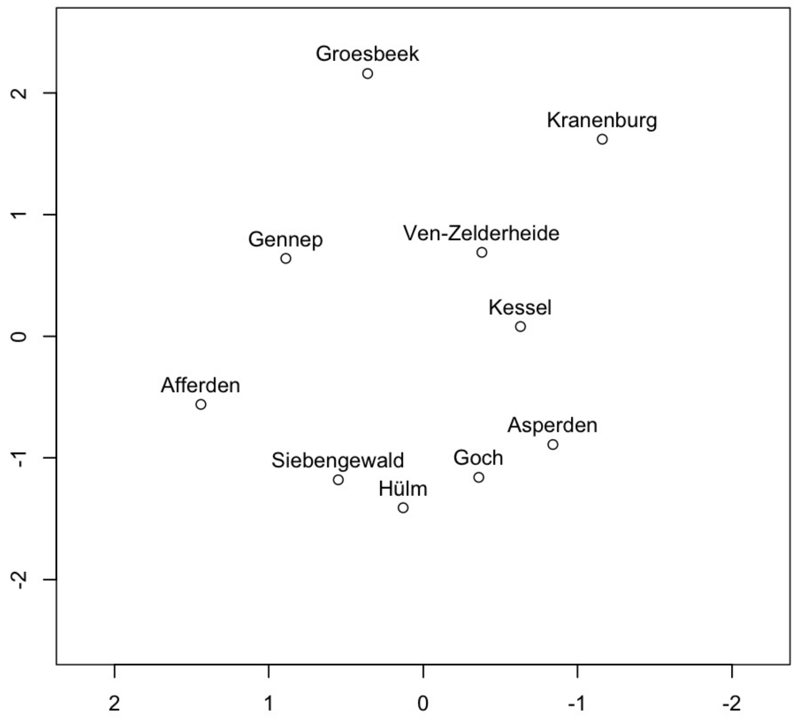

Figure 6: Two-dimensional MDS plot of the geographic distances

The linguistic dissimilarities between the ten locations were computed using the Levenshtein method in which all pairs of words (strings) were being compared. The distance between the two strings 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. (see also Heeringa 2004) We used the dialectometric software RuG/L04 (Kleiweg) for computing the distances. With the RuG/L04 software we obtained a 10 by 10 dissimilarity matrix for the locations. We did the same type of MDS analyses for the descriptive linguistic distances and got a nearly perfect two-dimensional solution (Stress = 0,050, RSQ = 0,990). For this plot, depicted in Fig. 7, we mirrored both the x-axis and the y-axis. The topology of the linguistic distances is different from that of the geographic distances plotted in Fig. 6.

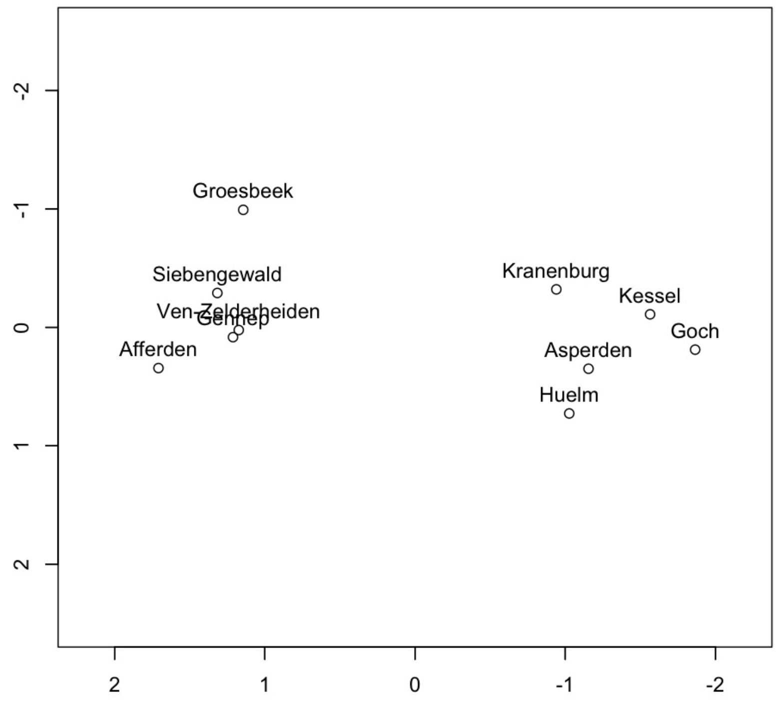

Figure 7: Two-dimensional MDS plot of the linguistic distances

The most remarkable outcome in Fig. 7 is that distances *within* the two countries are much smaller than the distances *between* the countries. The distances between the countries are always larger than whatever distance within a country. The continuum model no longer applies, and the gap model seems to give the required topology.

Can we investigate the topological structures and their relations in more detail? We calculated the distances between all 45 pairs of locations for both the geographic and linguistic distances. The continuum model predicts a high correlation, the gap model no correlation at all. The correlation is .256, with a one-tailed p value of .045. (We opted for a one-tailed test, for if a correlation exists its value must be positive.) The conclusion is that the correlation is significant but low. Geographical distances hardly play a role as an explanatory factor in explaining the linguistic distances. Do we need to reject now the continuum model completely?

To test the continuum model in more detail, we divided the location pairs into three groups: *Dutch* couples, *German* couples and *Dutch-German* couples. Given the gap model the linguistic distance within Dutch-German couples should be relatively large and constant. The distance within the Dutch couples and within the German couples 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 couples are given in Table 3.

	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)

Table 3: Correlation values for linguistic distances and the three types of location couples; one-tailed p values

The correlation for the Dutch-German location couples is not significant. The correlations for the Dutch and German location couples is clearly higher, although the correlation for the Dutch couples is just not significant. Given the low number of location couples, it is clear that the statistical test of the correlations does not have much power.

We visualized the relationship between the geographic and linguistic distances in a direct way by transforming the linguistic distances to a similar scale as the geographic distances (we used the same maximum). Scaling does not change the intrinsic structural characteristics of the distance matrices and is of benefit to the interpretation. On the diagonal line we now expect to find the couples adhering to the continuum model, in which the linguistic distance equals the geographic distance. The scatter plot is given in Fig. 9, where the three groups of couples are distinguished by different symbols. The triangles mark the cross-the-border couples.

The Dutch-German couples have a distinct location. The geographic distance varies between 3 and 26 kilometres, but their (scaled) linguistic distance ranges between 17 and 26 kilometres. There is no further explanation for the variation within the range found. The smallest distance of 17 has the same size as the largest distance found for Dutch-Dutch and German-German couples. The gap model obviously applies to the cross-the-border couples. The distances between the locations within each of the countries are smaller, but they show at the same time a diagonal pattern. Both the circles (Dutch-Dutch) and the plusses (German-German) globally give an increase for linguistic distance as the geographic distance gets larger. The strength of this relationship was indicated by the correlation in Table 3. That means that within the countries the continuum model applies, although in a moderate fashion. The restriction of the continuum model within the two countries may have to do with the differential impact of the standard languages involved on the local dialects.

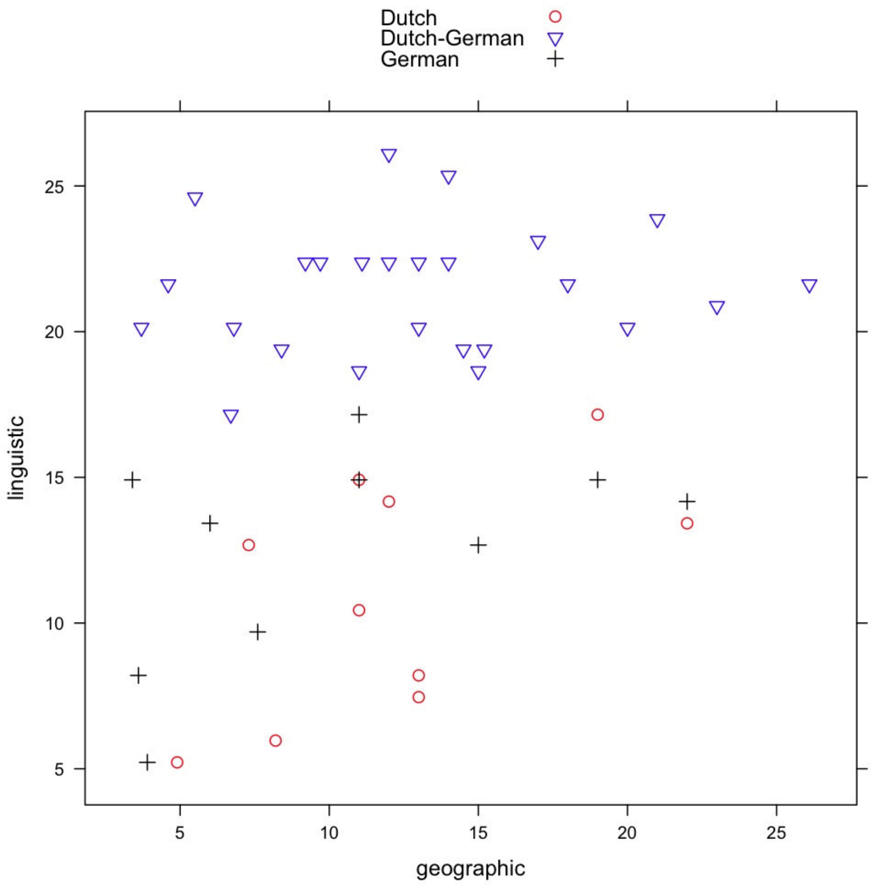

Figure 9: geographic distance by linguistic distance, on the same scale of magnitude; a distinction is made between Dutch-Dutch, German-German and Dutch-German couples

As stated earlier, intensive contact between speakers is an essential condition for dialects to not diverge over time. This is especially true when different standard languages have an increasing influence on the position and the structure of local dialects. The continuum model no longer applies for the whole research area, but only moderately for the within country linguistic differences. Cross-the-border dialects are nowadays separated by a plain gap.

3.2 Linguistic and perceptual distances

Can socio-geographical patterns perhaps explain the linguistic distances, or, to formulate it in a more moderate way, do they parallel linguistic distances? Respondents were asked about their own perceptions of contact phenomena. The phenomenon or concept affiliated most directly to linguistic distance is the way the linguistic distances are being perceived in the research area. Respondents listed and ranked the ten nearest dialects. The outcomes were analysed in detail for respondent effects of country, place of living, age, and gender (Giesbers 2008). The strongest outcome was that Dutch respondents choose Dutch dialects (77%) and that German respondents choose German dialects (78%). This country effect outweighed all other effects. There were no age and gender effects, and the differences between the locations where one lived were minor in comparison to the country effect. We have collapsed the perceptual linguistic distance data for each of the ten locations under study. Each location got its own vector by adding the weighed choices of its own respondents. The highest rank of a respondent got 10 points, the lowest rank 1 point. The resulting vector with rank values for a specific location can be compared with the vector of another location. When the two locations make the same choice the dissimilarity between the vectors will be low. When they choose differently, the dissimilarity will be high. The result is again a ten by ten dissimilarity matrix, like we had for the linguistic distances.

We performed an MDS analysis on the perceptual linguistic dissimilarities and again a two-dimensional structure turned out to be nearly perfect (Stress = 0,022, RSQ = 0,998). That means that the dissimilarities can be interpreted as two-dimensional distances. The resulting MDS plot is given in Fig. 10. We mirrored the x-axis for this plot.

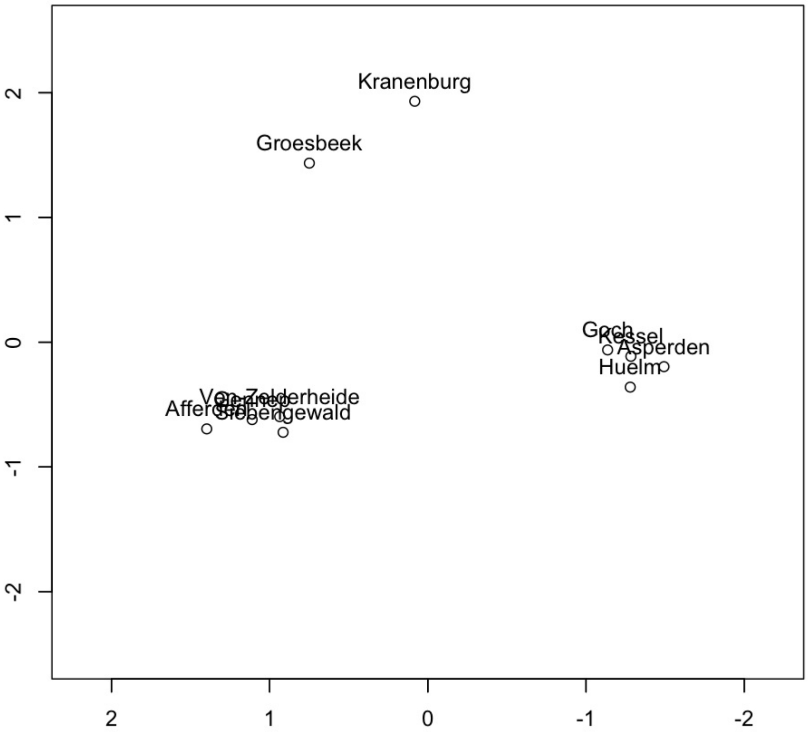

Figure 10: Two-dimensional MDS plot of the perceptual linguistic distances. The cluster on the left consists of Afferden, Ven-Zelderheide, Siebengewald and Gennep. The cluster on the right consists of Goch, Asperden, Kessel and Huelm

We see a topology in Fig 10 comparable to the one found for linguistic distance. There is a gap separating the two clusters with Dutch and German locations. Noteworthy are the towns of Groesbeek and Kranenburg that lie separated in the top of the plot (the 'north'), and, moreover, the gap between them is smaller than for the other locations. The perceptual distances for these two towns have maintained more of the geographic structure than the other locations. Do we see the similarity between the linguistic distance and the perceived linguistic distance back in their correlation? The relevant correlations are given in Table 4.

	Geographic distance (N=45)	**Perceptual Linguistic distance** (N=45)
Linguistic distance	0,256 (p = 0,045)	0,762 (p = 0,000)

Table 4: Correlations for linguistic distances and the geographic and perceptual linguistic distances; one-tailed p values

As can be seen in Table 4 the linguistic distances correlate much better with the perceptual distances than with the geographic distances. The correlation is fairly strong, strong enough to conclude that actual and perceptual linguistic distance have a plain overlap in their socio-geographical structure and that the perceptual distance model has explanatory value.

Can other data than language related data tell us more about the socio-geographical structure? The next type of perceptual data studied were the contact data Giesbers collected about friends, family, and shopping. The data were analysed in the same way as the perceptual distance data. Again, no effect was found for the respondent variables of gender and age. The weighted data were computed per location and 10 by 10 dissimilarity matrices were obtained for all three contact variables. MDS returned excellent results for a two-dimensional representation for all three contact variables. The two-dimensional MDS results for *friends* (Stress = 0,070, RSQ = 0,972) are given in Fig. 12. Again we mirrored the x-axis.

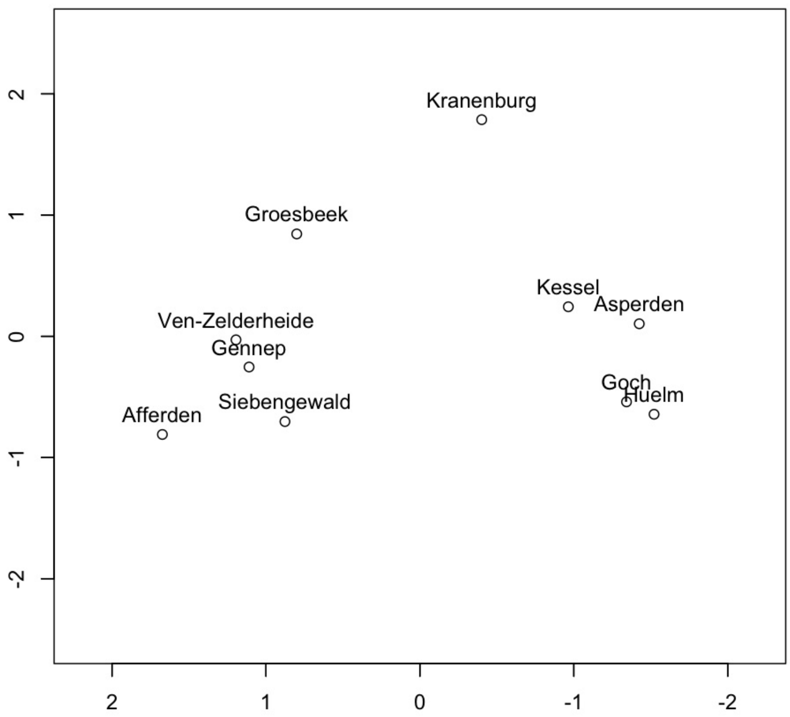

Figure 12: Two-dimensional MDS plot of the contact variable friends

The plot in Fig. 12 shows that the general topology for *friends* resembles the MDS plot for linguistic distances depicted in Fig. 7. Is the similarity higher than for the perceptual linguistic distance configurations and what is the correlation for the other two contact variables? The correlations can be found in Table 5.

	Shopping (N=45)	**Family** (N=45)	**Friends** (N=45)
Linguistic	0,623 (p = 0,000)	0,737 (p = 0,000)	0,818 (p = 0,000)

Table 5: Correlations for linguistic distance and shopping, family or friends

The contact variable friends shows the highest correlation, but the correlation for the other two contact variables are clearly present. The linguistic distance is not a property on its own, but is embedded in the way the socio-geographic structure is being perceived by the respondents.

To get a more detailed picture of the relationship between the variables of friends and linguistic distance, we used the same visualization method as applied in Fig. 9. 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 Fig. 14, for three groups of couples: Dutch-Dutch, German-German and Dutch-German.

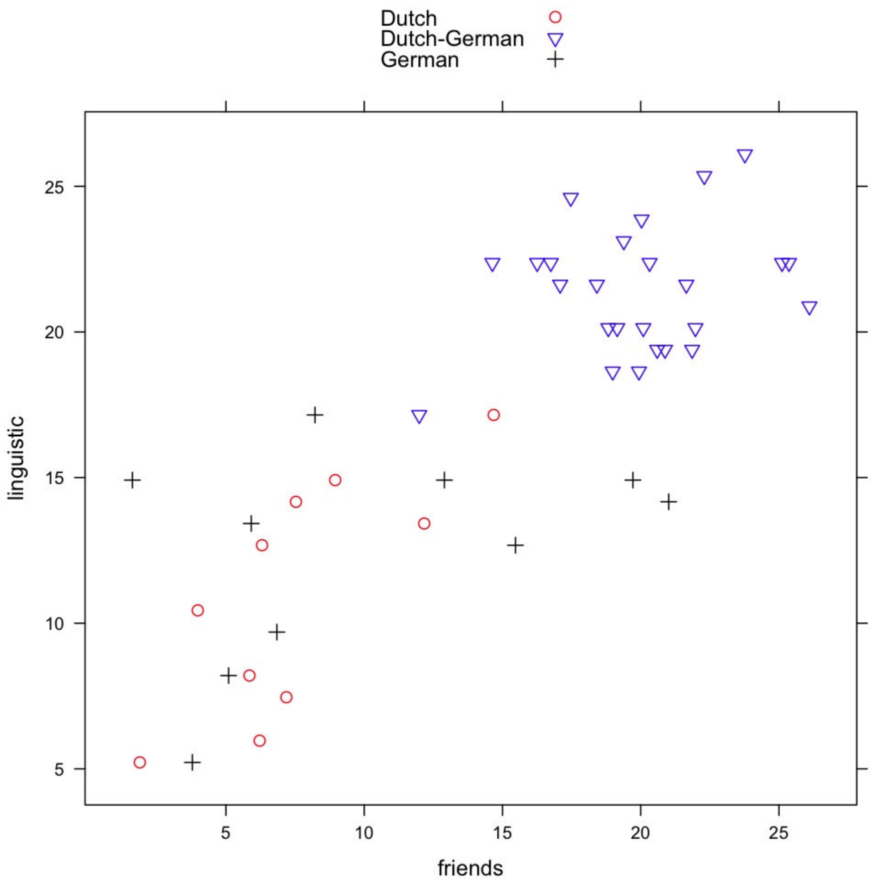

Figure 14: linguistic distance by friends distance, on the same scale of magnitude; a distinction is made between Dutch-Dutch, German-German and Dutch-German couples

Figure 14 shows a diagonal pattern for all three groups of couples together, although the relationship is not perfect, but scattered. The diagonal relationship applies to all three groups of couples, meaning that the two variables really share a similar configuration or topology. Fig. 14 strengthens our interpretation of the linguistic distances found as belonging to an overarching socio-geographic pattern that has developed for the research area over the last two centuries.

4 Conclusions

We proposed three different models for describing the relation between linguistic distances and geographic and perceptual distances and tested to what extent the models are corroborated by the different spatial data sets we collected and their spatial constellations. We started out with the continuum model and showed that nowadays it no longer applies to the Kleverlandish research area. Cross-the-border dialects today are separated by a plain gap. The gap model applies to the cross-the-border location couples very precisely. For the location couples within each of the countries the continuum model still seems to apply, although only in a moderate and imperfect fashion.

In the next step, we tried explaining the linguistic distances with other spatial configurations, and without geographic distances. We proposed a perceptual model to relate linguistic distance to perceptual distances, including both perceptual linguistic distances and perceptual socio-geographic contact distances (relatives, friends, shopping). The perceptual model fitted the linguistic distances much better than the continuum or gap models which are based on geographic distances. Especially the spatial configuration for friends data very much resembled the linguistic data.

The dialect variation in our research turns out to be more closely related to socio-geographic structures than to the geographic spatial configuration. This can be seen as the consequence of dialect variation paralleling or reflecting contact data, dialect variation being the product of social entities (people, groups of people) interacting with each other. When there is no or scarce social interaction and no cohesive social system, like between locations across the state border in the Kleverlandish area, dialect variation maximizes, certainly because the dialects in question function under the influence of two separate standard language varieties, Dutch and German.

In conclusion, in determining linguistic distance the geographic distances between locations are of lesser importance than the frequency of contact and interaction. This is especially interesting in the light of today's improved mobility and telecommunication possibilities. Both of these aspects contribute to the declining influence of one's geographic location on the possibilities for social interaction. For the study of language variation between closely related languages (neighbouring dialects), our conclusions can be the starting point for a deeper study of socio-geographic network structures of dialect areas. Which socio-geographic variables are really pivotal in understanding and explaining spatial structures in linguistic variation? We have shown that it is possible to include perceptual contact variables in explaining spatial linguistic configurations.

5 References

- Chambers, J.K. & Trudgill, P. (1998). Dialectology. Cambridge: Cambridge University Press.
- De Vriend, F., Kunst, J.P., Ten Bosch, L., Giesbers, C. & Van Hout, R. (2008). Evaluating the Relationship between Linguistic and Geographic Distances using a 3D Visualization. Proceedings of The sixth international conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco.
- Giesbers, C.H.E., Van Hout, R. & Van Bezooijen, R. (2006). De geografische en perceptieve afstand tussen dialecten: de rol van de staatsgrens in het Kleverlandse dialectcontinuüm. In Koole, T., Nortier, J. & Tahitu, B. (Eds.), Artikelen van de vijfde sociolinguïstische conferentie (pp. 175-186). Delft: Eburon.

- Giesbers, C. (2008). Dialecten op de grens van twee talen. Een dialectologisch en sociolinguïstisch onderzoek in het Kleverlands dialectcontinuüm. PhD thesis, Radboud University, Nijmegen.
- Heeringa, W., Nerbonne, J., Niebaum, H., Nieuweboer, R. & Kleiweg, P. (2000). Dutch-German contact in and around Bentheim. Languages in Contact. In: Studies in Slavic and General Linguistics 28. Gilbers, D.G., Nerbonne, J. & Schaeken, J. (eds.) Amsterdam-Atlanta: Rodopi.
- Heeringa, W. (2004). Measuring Dialect Pronunciation Differences using Levenshtein Distance, PhD thesis Rijksuniversiteit Groningen, Groningen.
- Heeringa, W., Nerbonne, J., Van Bezooijen, R. & Spruit, M. (2007). Geografie en inwoneraantallen als verklarende factoren voor variatie in het Nederlandse dialectgebied. In: Nederlandse Taal- en Letterenkunde, Volume 123(1), Uitgeverij Verloren, Hilversum, 70–82.
- Kleiweg, P. RuG/L04, Software for dialectometrics and cartography. http://www.let.rug.nl/~kleiweg/L04/.
- Nerbonne, J., Heeringa, W. & Kleiweg, P. (1999). Edit Distance and Dialect Proximity. In: Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison. Sankoff, D. & Kruskal, J.B. (eds.). Stanford: CSLI.
- Spruit, M. (2008). Quantitative perspectives on syntactic variation in Dutch dialects. LOT Dissertation Series 174.