On the asymmetry of wh-doubling in varieties of German and Dutch

Barbiers, S; Delbar, N.; Fanselow, G.; Nauta, S.; Rothert, J.

Appendices of: On the asymmetry of wh-doubling in varieties of German and Dutch

Sjef Barbiers, Natasja Delbar, Gisbert Fanselow, Sophia Nauta, Johannes Rothert

Appendix 1 Subexperiments

Appendix 1.1 Dutch

As the existence of two dialects in both Dutch and German in the domain of the location of complexity has been made plausible by the above data analysis, a natural question to ask is if these dialects have some regional basis. To this end, we investigated both the Standard language and dialects of Dutch. The data of the pertaining two experiments constitute the pooled data set we have analyzed above.

The Dutch language area includes The Netherlands and the Dutch speaking part of Belgium. We also included Frisian, which has been considered to be a language distinct from Dutch and its dialects. The survey consisted of two rounds. In the first round we wanted to establish whether the various dialect regions in the Dutch language area are distinct with respect to the judgements on the doubling construction and the location of complexity. Since the number of participants per dialect region in the first round was too low for statistical analysis , we decided to have a second round with speakers of standard Dutch, both in the Dutch and the Belgian part of the Dutch language area.

Appendix 1.1.1 Dutch dialects

The dialect groups investigated and the number of subjects per dialect group are given in (25).

(25) Dialect groups and number of participants (n=71) in survey round 1, Dutch language area

West-Flemish	18	East-Flemish	6	Belgian Brabantish	8
Belgian Limburgish	5	whil Limburgish	8	Dutch Brabantish	6
Zeeuws (Zeeland)	2	Hollandic	3	Gronings	11
Frisian	4				

The results of the first round are summarized in Figure 10, with average scores on conditions a (simple), b (right complexity) and c (left complexity) (1 = very unnatural; 7 = completely natural).

Figure 1

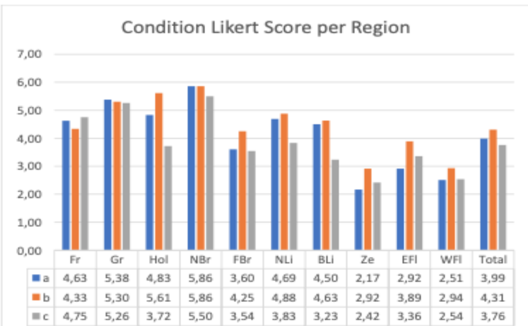

Figure 10 Condition a: simple wh doubling - Condition b: right complexity - Condition c: left complexity

Although it is not possible to establish whether the differences depicted in Figure 10 are statistically significant, due to the low number of respondents, there are several tendencies that are worth mentioning.

First, we see that simple wh doubling (condition a) is judged to be quite natural in most dialect groups, except in Zeeuws, West- and East-Flemish. For West-Flemish, this replicates the results reported in SAND Volume I, map 91b, but not for the other two dialect groups, which have many dialects on map 91b that allow simple wh doubling. In Zeeuws and West-Flemish, and to a lesser extent in East-Flemish as well, the low acceptance rate of simple doubling seems to correlate with low grades on condition b and c as well.

A second finding is that all dialect groups have a preference for right complexity, except for Frisian where we find the reverse. In most cases however, these are only slight differences. It seems fair to conclude that in Frisian, Gronings and Northern Brabantish all three conditions are judged to be natural, while in Hollandic and Belgian Limburgian right complexity is clearly prefered over left-complexity.

If the respondents of the first round are analyzed as one undifferentiated group, then statistical analysis is possible. Table 3 summarizes the means for the three experimental conditions both for the complete and the reduced data set. Figure 11 gives the plots of the difference between the mean judgments of right and left complexity for each of the subjects, and the histograms in Figure 12 show the frequency distribution of the speakers relative to the different naturalness that they attribute to conditions b and c. The left hand histogram takes into account all 71 dialect speakers, the right hand histogram shows the results if the 27 respondents are excluded who rated condition a (simple doubling) lower than 3

Table 1: Mean ratings; Left: all participants (n = 71), right: after exclusion (n = 45)

left	3.76	left	3.79
right	4.31	right	4.34
simple	3.99	simple	4.03

Table 3

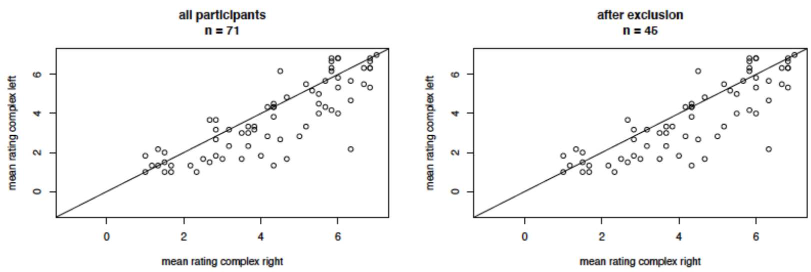

Figure 11

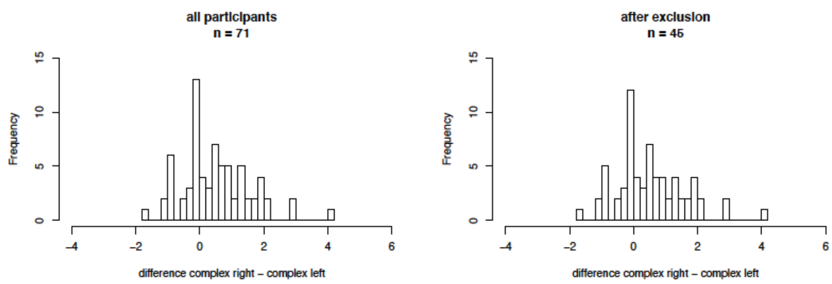

Figure 12

As before, we fitted 3 Gaussian finite mixture models to the data that differed in the number of underlying normal components (1, 2, or 3, Models M1, M2 and M3). Using the likelihood ratio test procedure, we compared the log likelihoods of these models to find the model that best fits the data. The details of the statistical analysis for the pooled Dutch dialect data are given in Appendix 1.1.4.

As illustrated by Figure 13, the statistical analysis for the complete Dutch dialect set revealed that the data are best analysed as being characterized by three normal distributions.

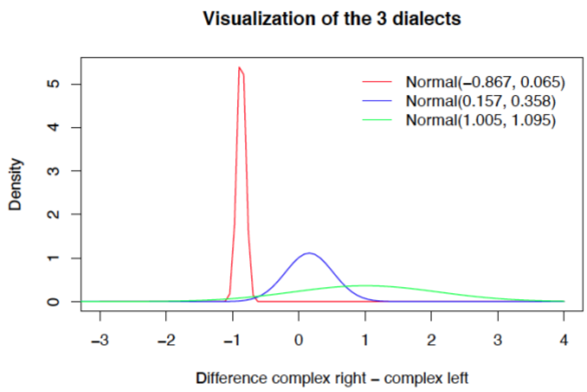

Figure 13

Numerically, there is one normal distribution(dialect) that has its peak at a (around 1) preference for left complexity. The two other distributions prefer right complexity in different shapes.

When it comes to the reduced data set containing only judgments form participants actually accepting the construction, the statistical analysis yields no clear result. There are two models that cannot be distinguished statistically, one depicted in Figure 14 with two dialects that looks much like the constellation we identified for the pooled Dutch data, and one depicted in Figure 15 with three dialects that looks like what we have found for the complete dialect set.

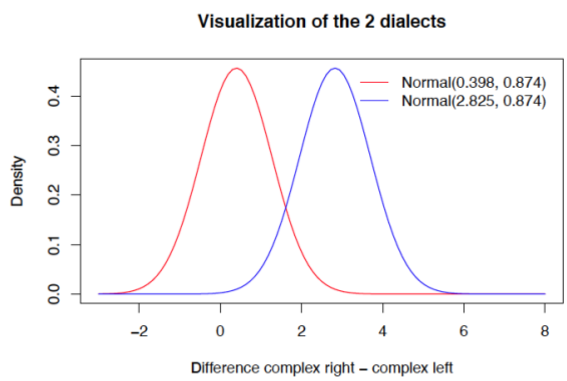

Figure 14

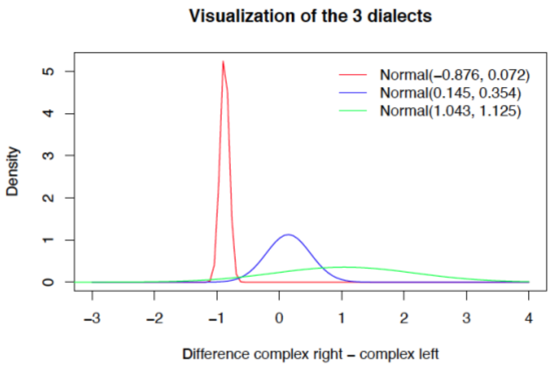

Figure 15

The two dialect model is qualitatively the same as the model identified for the complete data set. In other words, on the basis of the available evidence, we cannot decide if the judgments of the Dutch dialect speakers show a qualitative difference to the complete data set or not.

Details of the statistical analysis

COMPLETE DATA SET
Log likelihoods of the three models:

M1 - model with 1 normal component: -103.8478 (df = 2))
M2 - model with 2 normal components: -101.1544 (df = 4)
M3 - model with 3 normal components: -96.04396 (df = 6)

Likelihood ratio test:
Comparison of M1 and M2:
Evidence that M1 and M2 do not have the same log likelihoods (χ^2 = 5.387, df2)

- Given that M2 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M1 and M3:
Evidence that the two models also do not have the same log likelihoods (χ^2 = 15.61, df = 4).

- Given that M3 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M2 and M3:
Evidence that the two models also do not have the same log likelihoods (χ^2 = 10.22
, df = 2).

- Given that M3 has a higher log likelihood than M2, it is a better fit to the data.

In summary, a model assuming 3 underlying normal componants fits the complete data best.

REDUCED DATA SET
Log likelihoods of the three models:

M1 - model with 1 normal component: - 97.60681 (df = 2))
M2 - model with 2 normal components: -95.00222 (df = 4)
M3 - model with 3 normal components: - -90.47565 (df = 8)

Likelihood ratio test:
Comparison of M1 and M2:
Evidence that M1 and M2 do not have the same log likelihoods ($\chi^2 =$ 5.209, df2)

- Given that M2 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M1 and M3:
Evidence that the two models also do not have the same log likelihoods ($\chi^2 =$ 14.26, df = 4).

- Given that M3 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M2 and M3:
According to a likelihood ratio test where we compared M2 and M3, there is no evidence that the two models do not have the same log likelihoods (χ^2 = 1.66, df = 2).

- For reasons of parsimony, M2 is a better fit to the data.

In summary, a model assuming 2 underlying normal componants fits the complete data best.

Appendix 1.1.2 Standard Dutch in Belgium and The Netherlands

In this round there were 90 respondents, all native speakers of Dutch. 50 of them were from the Dutch speaking area of Belgium, 40 from The Netherlands. Material and methods were as described above. The comparison of the responses from Belgium and the Netherlands can be found in Figure 16. We see that the three types of sentences are slightly more natural for speakers of Dutch in The Netherlands than in Belgium. The average total scores in the two rounds are very similar. There is a slight preference for right complexity over left complexity in the entire language area.

Condition a: simple wh doubling - Condition b: right complexity - Condition c: left complexity

Figure 2

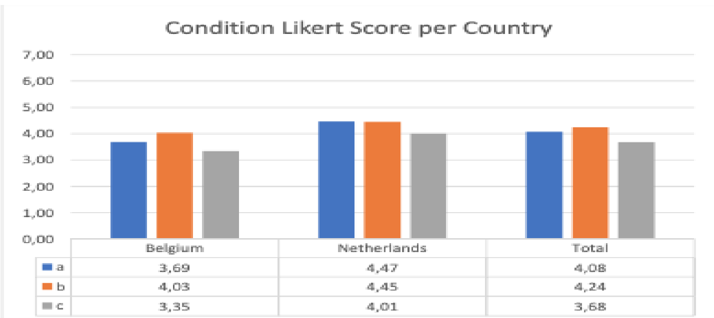

Figure 16

The statistical analysis of the judgments in this round does not differ in any interesting way from the results found for the pooled data set. The means for the simple, right and left complex construction both for the complete and the reduced data set are represented in Table 4. There is a small (.53 on the 7 point Likert scale) advantage of the right complex construction over the left complex one for those speakers who accept the construction at all.

Table 1: Mean ratings; Left: all participants (n = 90), right: after exclusion (n = 67)

left	3.64	left	4.07
right	4.22	right	4.60
simple	4.04	simple	4.59

Figure 17 gives the plots of the difference between the mean judgments of right and left complexity for each of the subjects, and the histograms in Figure 18 show the frequency distribution of the speakers relative to the different naturalness that they attribute to conditions b and c. The left hand histogram takes into account all 90 standard speakers, the right hand histogram shows the results if the 33 respondents are excluded who rated condition a (simple doubling) lower than 3

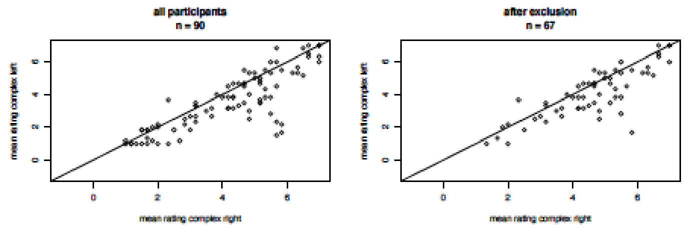

Figure 17

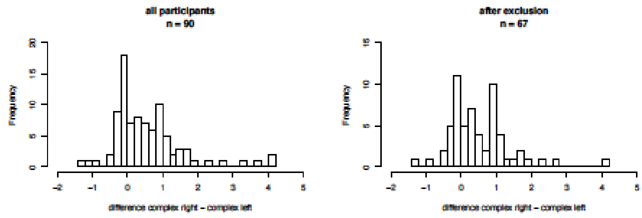

Figure 18

Just as for the pooled data set, the likelihood ratio test identifies models that work with 2 normal distributions as the best fit of the data, both when all responses are taked into account and when the analysis is confined to the responses of the standard speakers accepting the doubling construction at all. The normal distributions are depicted in Figure 19 for the complete data set, and in Figure 20 for the reduced one. The details of the statistical analysis are given below.

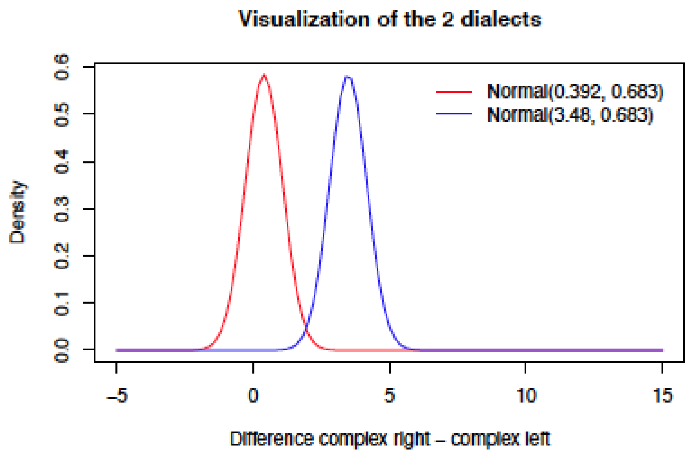

Figure 19

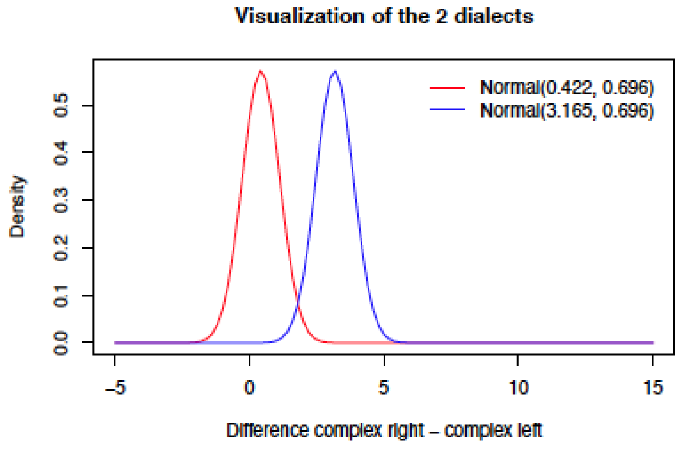

Figure 20

Details of the statistical analysis

COMPLETE DATA SET
Log likelihoods of the three models:

M1 - model with 1 normal component: -128.0528 (df = 2))
M2 - model with 2 normal components: -112.3936 (df = 4)
M3 - model with 3 normal components: - -112.394 (df = 6)

Likelihood ratio test:
Comparison of M1 and M2:
Evidence that M1 and M2 do not have the same log likelihoods (χ^2 = 31.32, df2)

- Given that M2 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M1 and M3:
Evidence that the two models also do not have the same log likelihoods (χ^2 = 31.32, df = 4).

- Given that M3 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M2 and M3:
When comparing M2 and M3, the result of the likelihood ratio test can not be interpreted, as the χ^2 value is negative χ^2 -0.0007821, df = 2). However, the two likelihoods are nearly identical and we therefore do not have reason to assume that there is any difference between both models in terms of how well they fit the data.

- For reasons of parsimony, M2 is a better fit to the data.

In summary, a model assuming 2 underlying normal componants fits the complete data best.

REDUCED DATA SET
Log likelihoods of the three models:

M1 - model with 1 normal component: -77.67644 (df = 2))
M2 - model with 2 normal components: -72.0106 (df = 4)
M3 - model with 3 normal components: - -72.01122 (df = 6)

Likelihood ratio test:
Comparison of M1 and M2:
Evidence that M1 and M2 do not have the same log likelihoods (χ^2 = 11.33, df2)

- Given that M2 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M1 and M3:
Evidence that the two models also do not have the same log likelihoods (χ^2 = 11.33, df = 4).

- Given that M3 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M2 and M3:
When comparing M2 and M3, the result of the likelihood ratio test can not be interpreted, as the χ^2 value is negative χ^2 = - -0.00123, df = 2). However, the two likelihoods are nearly identical and we therefore do not have reason to assume that there is any difference between both models in terms of how well they fit the data.

- For reasons of parsimony, M2 is a better fit to the data.

In summary, a model assuming 2 underlying normal componants fits the complete data best.

Appendix 1.1.3 Summary: Dutch surveys

Both the pooled data set and the data for the speakers of the Standard language have revealed a stable but small acceptability advantage for right complex over left comple doubling constructions. For both cases, one can establish the existence of two normal distrubutions that might be interpreted as dialects. For speakers of one dialect, there is no real difference in acceptability between left and right complexity, for speakers of the other, the right complex construction is judged better than the left complex construction, but it is only very few speakers who share such intuitions.

As a whole, the dialect data share the small acceptability advantage of the Standard language data set, but it could not be determined if the dialect data of speakers accepting the construction are indicative of the existence of two dialects (comparable then to the Standard language data) or of three dialects (that can be identified the complete dialect data set). The low number of participants in certain regions made it impossible to break down the overall result to particular regions in a statistically sound way.

Appendix 1.1.4 Pooled Dutch data: details statistical analysis

COMPLETE DATA SET
Log likelihoods of the three models:

M1 - model with 1 normal component: -231.9798 (df = 2))
M2 - model with 2 normal components: -216.0339 (df = 4)
M3 - model with 3 normal components: -216.0346 (df = 6)

Likelihood ratio test:
Comparison of M1 and M2:
Evidence that M1 and M2 do not have the same log likelihoods (χ^2 = 31.89, df2)

- Given that M2 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M1 and M3:

Evidence that the two models also do not have the same log likelihoods (χ^2 = 31.89, df = 4).

- Given that M3 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M2 and M3:
When comparing M2 and M3, the result of the likelihood ratio test can not be interpreted, as the χ^2 value is negative χ^2 = -0.001389, df = 2). However, the two likelihoods are nearly identical and we therefore do not have reason to assume that there is any difference between both models in terms of how well they fit the data.

- For reasons of parsimony, M2 is a better fit to the data.

In summary, a model assuming 2 underlying normal componants fits the complete data best.

REDUCED DATA SET
Log likelihoods of the three models:

M1 - model with 1 normal component: -153.9273 (df = 2))
M2 - model with 2 normal components: -144.106 (df = 4)
M3 - model with 3 normal components: - -144.1062 (df = 6)

Likelihood ratio test:
Comparison of M1 and M2:
Evidence that M1 and M2 do not have the same log likelihoods (χ^2 = 19.64, df2)

- Given that M2 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M1 and M3:
Evidence that the two models also do not have the same log likelihoods (χ^2 = 19.64, df = 4).

- Given that M3 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M2 and M3:
When comparing M2 and M3, the result of the likelihood ratio test can not be interpreted, as the χ^2 value is negative χ^2 = - -0.0003207, df = 2). However, the two likelihoods are nearly identical and we therefore do not have reason to assume that there is any difference between both models in terms of how well they fit the data.

- For reasons of parsimony, M2 is a better fit to the data.

In summary, a model assuming 2 underlying normal componants fits the complete data best.

Appendix 1.2 German
We intended to carry out the experiment in the three dialect areas for which we expected particularly high numbers or speakers accepting wh-doubling, viz. Berlin/Brandenburg, the Ruhr area, and Bavaria. Based on the description of Anyadi & Tamrazian (1993), we hoped for a large number of participants on the Ruhr-area using the free relative-clause dialect. Participants for Brandenburg and the Ruhr-area were recruited with the help of the universities of Potsdam (SONA) and Bochum (Agata Renans), respectively. The participants at these two places were students, and the audio material was recorded such that it would represent the regional variants of the Standard language rather than a true dialect. For Bavaria, we recruited participants with the help of the lab of Mechthild Habermann at the University of Erlangen/Nürnberg. Participants were born and raised in the region where the local variety of German is spoken.

For Bavaria, we tried to recruit participants only that actively used the local dialect (Oberostfränkisch), and they were not students but dialect speakers previously recruited for a different study. Although recruitment went relatively well. There were not enough participants who actually completed the whole experiment. Given the sociolinguistic differences, we decided to not analyse the few complete Bavarian responses any further.

Appendix 1.2.1 Ruhr Area/North Rhine Westfalia

76 participants participated in the experiment, of which 62 gave a mean rating for the simple doubling construction above 3, which we use as a criterion for identifying the participants for which wh-doubling is acceptable. Table 6 gives the mean ratings for simple, left and right complex doubling constructions, both for all participants and those for which the construction is acceptable. As we can see, the difference between left and right complexity is, again, numerically quite small.

Table 1: Mean ratings; Left: all participants (n = 76), right: after exclusion (n = 68)

left	4.16	left	4.40
right	4.50	right	4.76
simple	4.66	simple	4.97

Table 6

The by now familar plots and histograms for the differences in the judgment between left and right complexity contain can be found in Figures 21 and 22

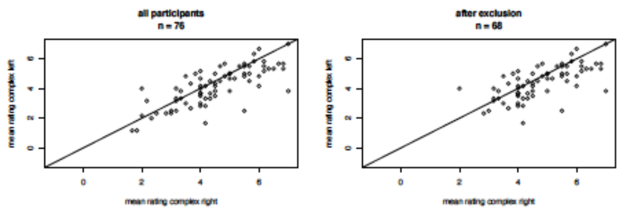

Figure 21

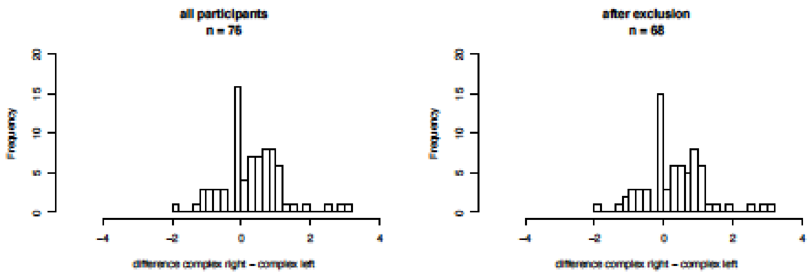

Figure 22

The likelihood ratio test yielded different results for the complete data set and data of the speakers who actually accept the construction. The complete set of data collected in Bochum comes with the three normal distributions shown in Figure 23:

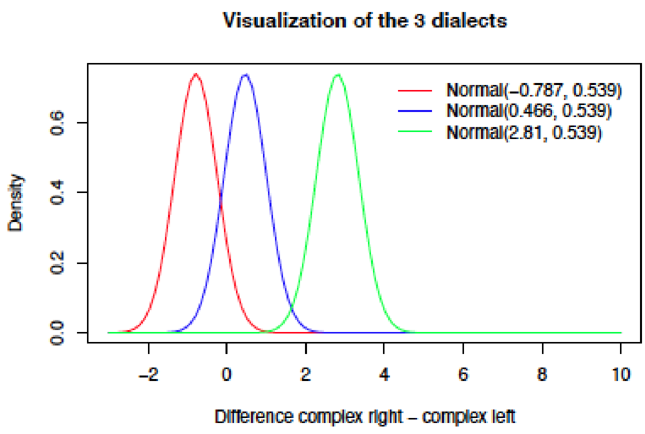

Figure 23

In contrast, there is no evidence whatsoever for multimodality in the reduced set, which is characterized by the normal distribution shown in Figure 24. See Appendix 2.5 for the details of the analysis.

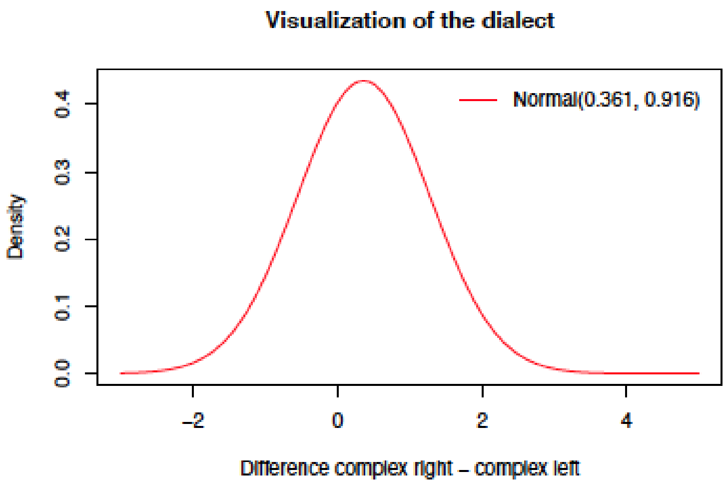

Figure 24

If we confine our attention to those speakers who find doubling acceptable, we observe that the data from Bochum are the most uniform we have seen so far, since it seems to be characterized by a single dialect only. It is tempting to understand this in terms of the fact that, unlike what holds for the data reported so far, the Bochum data are not composites of data collected in different regions, but come from a single regional variety. If this is the decisive factor, the Bochum study would make the idea plausible that the two complexity dialects we have identified in the composite data set might have a regional basis after all.

Details of the statistical analysis

COMPLETE DATA SET

Log likelihoods of the three models:

M1 - model with 1 normal component: -98.53606 (df = 2))
M2 - model with 2 normal components: -95.23294 (df = 4)
M3 - model with 3 normal components: - -92.80487 (df = 6)

Likelihood ratio test:
Comparison of M1 and M2:
Evidence that M1 and M2 do not have the same log likelihoods (χ^2 = 6.606, df3)

- Given that M2 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M1 and M3:
Evidence that the two models also do not have the same log likelihoods (χ^2 = 11.46, df = 4).

- Given that M3 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M2 and M3:
Evidence that the two models also do not have the same log likelihoods (χ^2 = 4.856, df = 1).

- Given that M3 has a higher log likelihood than M2, it is a better fit to the data.

In summary, a model assuming 3 underlying normal componants fits the complete data best.

REDUCED DATA SET
Log likelihoods of the three models:

M1 - model with 1 normal component: - -90.51226 (df = 2))
M2 - model with 2 normal components: -87.83919 (df = 5)
M3 - model with 3 normal components: - -86.06428 (df = 6)

Likelihood ratio test:
Comparison of M1 and M2:
According to a likelihood ratio test where we compared M1 and M2, there is no evidence that the twomodels do not have the same log likelihoods (χ^2 = 5.346, df = 3).

- For reasons of parsimony, M1 is a better fit to the data.

Comparison of M1 and M3:
According to a likelihood ratio test where we compared M1 and M3, there is no evidence that the twomodels do not have the same log likelihoods (χ^2 = 8.896, df = 4).

- For reasons of parsimony, M1 is a better fit to the data.

Comparison of M2 and M3
According to a likelihood ratio test where we compared M1 and M3, there is no evidence that the two models do not have the same log likelihoods (χ^2 = 3.55, df = 1).

- For reasons of parsimony, M2 is a better fit to the data.

In summary, a model assuming 1 underlying normal component fits the complete data best.

Appendix 1.2.2. Berlin/Brandenburg
49 participants took part in our experiment in the Berlin-Brandenburg area, of which 39 gave mean ratings above 3 for the simple doubling construction. Table 7 gives the mean ratings for simple. left and right complex doubling constructions, both for all participants and those for which the construction is acceptable. As we can see, the difference between left and right complexity is again numerically rather small.

Table 1: Mean ratings; Left: all participants (n = 49), right: after exclusion (n = 39)

left	3.95	left	4.43
right	4.62	right	4.99
simple	4.84	simple	5.21

Table 7

Numerically, acceptability was higher than in the pilot study mentioned above, by 1.17 for the right complex condition and by 1.12 for the left complex condition in the reduced data set. The plots in Fig 25 again show the mean rating for condition b (complex right) on the x-axis and the mean rating for condition c (complex left) on the y-axis for each participant. The histograms in Fig. 26 show he data distribution when considering the difference between complex right and complex left.

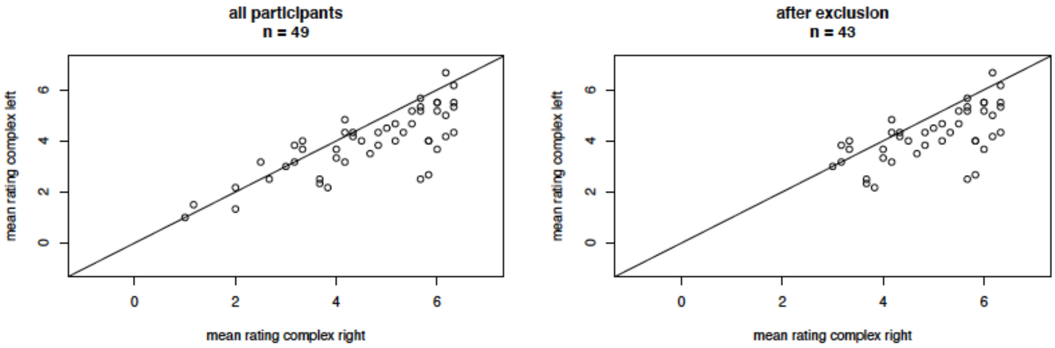

Figure 25

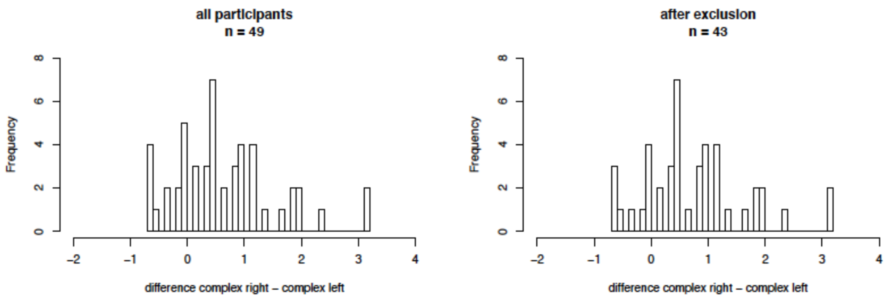

Figure 26

Both for the complete and the reduced data sets, the likelihood ratio test indicates that a model assuming three underlying normal components fits the data best. The results are depicted in Figure 27 (complete data set) and Figure 28 (responses of the participants who actually accept the construction). Note that one of the distributions cannot be plotted in both cases because it has too a small standard deviation ((Normal(μ = -0.665, σ = 0.005) and (Normal(μ = -0.663, σ = 0.004)). for the complete and reduced data sets, respectively).

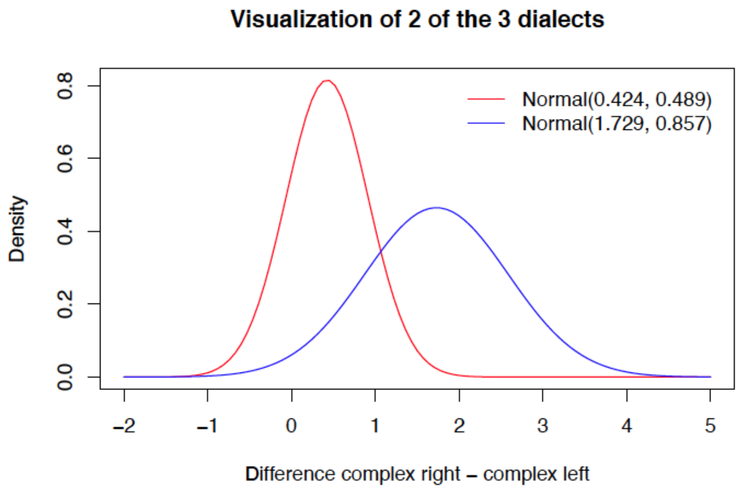

Figure 27

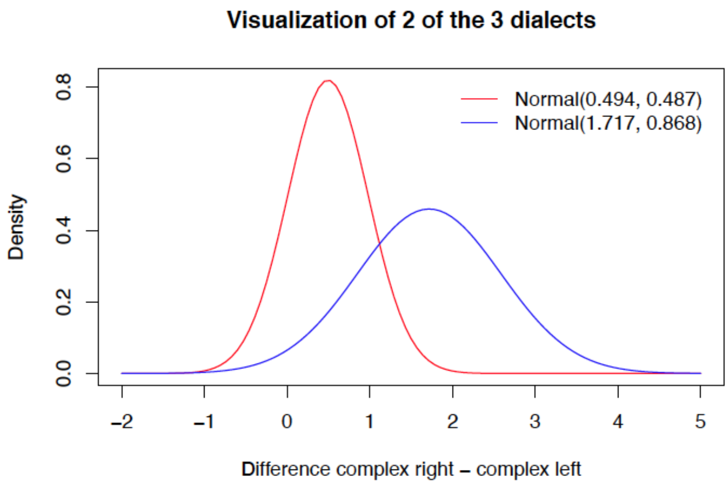

Figure 28

The Berlin-Brandenburg data thus turn out to be less uniform dialectally than those from the Bochum / Ruhr area study. One could speculate that the dialectal situation of the capital area has become more diversified in the last three decades because of the large influx of new inhabitants (Potsdam grew by 50,000 in the last 20 years, now having some 180,000 inhabitants, Berlin grew by 280,000, now having some 3,364,000 inhabitants, while Bochum lost 30,000, having gone down to 365,000 people). However, the overall picture is not much different from the other results but the one from the Ruhrgebiet: the "dialect" with a clear right advantage is constituted by very few participants only.

Details of the statistical analysis

COMPLETE DATA SET
Log likelihoods of the three models:

M1 - model with 1 normal component: -64.46088 (df = 2))
M2 - model with 2 normal components: -61.48317 (df = 4)
M3 - model with 3 normal components: -49.99041 (df = 8)

Likelihood ratio test:
Comparison of M1 and M2:
According to a likelihood ratio test where we compared M1 and M2, there is no evidence that the two models do not have the same log likelihoods ((χ^2 = 5.955, df2)

- For reasons of parsimony, M1 is a better fit to the data.

Comparison of M1 and M3:
Evidence that the two models also do not have the same log likelihoods (χ^2 = 28.94, df = 6).

- Given that M3 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M2 and M3:
Evidence that the two models also do not have the same log likelihoods (χ^2 = 22.99, df = 4).

- Given that M3 has a higher log likelihood than M2, it is a better fit to the data.

In summary, a model assuming 3 underlying normal componants fits the complete data best.

REDUCED DATA SET
Log likelihoods of the three models:

M1 - model with 1 normal component: - -56.73501 (df = 2))
M2 - model with 2 normal components: -54.49895 (df = 4)
M3 - model with 3 normal components: -45.46978 (df = 8)

Likelihood ratio test:
Comparison of M1 and M2:
According to a likelihood ratio test where we compared M1 and M2, there is no evidence that the two models do not have the same log likelihoods (χ^2 = 4.472, df = 3).

- For reasons of parsimony, M1 is a better fit to the data.

Comparison of M1 and M3:
Evidence that the two models do not have the same log likelihoods (χ^2 = 22.53, df = 6).

- Given that M3 has a higher log likelihood than M1, it is a better fit to the data

Comparison of M2 and M3
Evidence that the two models also do not have the same log likelihoods (χ^2 = 18.06, df = 4).

- Given that M3 has a higher log likelihood than M2, it is a better fit to the data.

In summary, a model assuming 3 underlying normal component fits the complete data best.

Appendix 1.2.3 Pooled German data: Details statistical analysis
COMPLETE DATA SET
Log likelihoods of the three models:

M1 - model with 1 normal component: -164.8961 (df = 2))
M2 - model with 2 normal components: -158.8801 (df = 4)
M3 - model with 3 normal components: -158.102 (df = 6)

Likelihood ratio test:
Comparison of M1 and M2:
Evidence that M1 and M2 do not have the same log likelihoods (χ^2 = 12.03, df2)

- Given that M2 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M1 and M3:

Evidence that the two models also do not have the same log likelihoods ($\chi^2 = 13.59$, df = 4).

- Given that M3 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M2 and M3:
According to a likelihood ratio test where we compared M2 and M3, there is no evidence that the twomodels do not have the same log likelihoods ($\chi^2 = 1.556$, df = 2).

- For reasons of parsimony, M2 is a better fit to the data.

In summary, a model assuming 2 underlying normal componants fits the complete data best.

REDUCED DATA SET
<u>Log likelihoods of the three mo</u>dels:

M1 - model with 1 normal component: -149.810 (df = 2))
M2 - model with 2 normal components: -145.245 (df = 4)
M3 - model with 3 normal components: - 144.4152 (df = 6)

<u>Likelihood ratio test</u>:
Comparison of M1 and M2:
Evidence that M1 and M2 do not have the same log likelihoods ($\chi^2 = 9.131$, df2)

- Given that M2 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M1 and M3:
Evidence that the two models also do not have the same log likelihoods ($\chi^2 = 10.79$, df = 4).

- Given that M3 has a higher log likelihood than M1, it is a better fit to the data.

Comparison of M2 and M3:
According to a likelihood ratio test where we compared M2 and M3, there is no evidence that the twomodels do not have the same log likelihoods ($\chi^2 = 1.66$, df = 2).

- For reasons of parsimony, M2 is a better fit to the data.

In summary, a model assuming 2 underlying normal componants fits the complete data best.

Appendix 2: List of Target Sentence Pairs with their contexts in the Dutch surveys
Condition a = wh-doubling, Condition b = right complexity, Condition c = left complexity

T1context Karel vertelt mij dat de kleuterjuf alleen de ouders van Christof zal uitnodigen.

1. T1a Wie denk jij wie zij uitnodigt?
2. T1b Wie denk jij wie van de ouders zij uitnodigt?
3. T1c Wie van de ouders denk jij wie zij uitnodigt?

T2context Katja meent dat de wiskundeleraar duidelijk de leerling uit de rijke buurt voortrekt.

1. T2a Wie denk jij wie hij voortrekt?
2. T2b Wie denk jij wie van de leerlingen hij voortrekt?
3. T2c Wie van de leerlingen denk jij wie hij voortrekt?

T3context

Andreas beweert dat de professor alleen de student van natuurkunde voor de onderscheiding voordraagt.

1. T3a Wie denk jij wie zij voordraagt?
2. T3b Wie denk jij wie van de studenten zij voordraagt?
3. T3c Wie van de studenten denk jij wie zij voordraagt?

T4context

Anna zei dat de baas alleen de medewerker van de PR-afdeling meeneemt naar de jaarbeurs.

1. T4a Wie denk jij wie hij meeneemt?
2. T4b Wie denk jij wie van de medewerkers hij meeneemt?
3. T4c Wie van de medewerkers denk jij wie hij meeneemt?

T5context

Bert meent dat de rechercheur alleen de oudste van de keukenbedienden verdenkt.

1. T5a Wie denk jij wie hij verdenkt?
2. T5b Wie denk jij wie van de keukenbedienden hij verdenkt?
3. T5c Wie van de keukenbedienden denk jij wie hij verdenkt?

T6context Britt zegt dat de vader de brutale jongen uit de buurt beschuldigt.

1. T6a Wie denk jij wie hij beschuldigt?
2. T6b Wie denk jij wie van de buurjongens hij beschuldigt?
3. T6c Wie van de buurjongens denk jij wie hij beschuldigt?

T7context Kristiaan gelooft dat de journaliste van alle politici alleen Rutte bewondert.

1. T7a Wie denk jij wie zij bewondert?
2. T7b Wie denk jij wie van de politici zij bewondert?
3. T7c Wie van de politici denk jij wie zij bewondert?

T8context

1. Carla beweert dat haar zoon van de Nederlandse zangeressen vooral Anouk aanbidt.
2. T8a Wie denk jij wie hij aanbidt?
3. T8b Wie denk jij wie van de Nederlandse zangeressen hij aanbidt?
4. T8c Wie van de Nederlandse zangeressen denk jij wie hij aanbidt?

T9context De trainer gelooft dat het meisje de jongste speler uit het team leuk vindt.

1. T9a Wie denk jij wie zij leuk vindt?
2. T9b Wie denk jij wie van de spelers zij leuk vindt?

3. T9c Wie van de spelers denk jij wie zij leuk vindt?

T10context Erika gelooft dat de medewerkster alleen de trouwste klant opbelt.

1. T10a Wie denk jij wie zij opbelt?
2. T10b Wie denk jij wie van de klanten zij opbelt?
3. T10c Wie van de klanten denk jij wie zij opbelt?

T11context De advocaat vertelt dat de getuige de leider van de bankovervallers wil belasten.

1. T11a Wie denk jij wie zij wil belasten?
2. T11b Wie denk jij wie van de bankovervallers zij wil belasten?
3. T11c Wie van de bankovervallers denk jij wie zij wil belasten?

T12context Greet meent dat de chauffeur de collega van de kwaliteitscontrole van diefstal beschuldigt.

1. T12a Wie denk jij wie hij beschuldigt?
2. T12b Wie denk jij wie van de collega's hij beschuldigt?
3. T12c Wie van de collega's denk jij wie hij beschuldigt?

T13context Hans is er vast van overtuigd dat de jonge activiste de CDA-politicus wil aanspreken.

1. T13a Wie denk jij wie zij wil aanspreken?
2. T13b Wie denk jij wie van de politici zij wil aanspreken?
3. T13c Wie van de politici denk jij wie zij wil aanspreken?

T14context Gianna vertelt dat de collectant alleen de rijkste inwoner om geld vraagt.

1. T14a Wie denk jij wie hij vraagt?
2. T14b Wie denk jij wie van de bewoners hij vraagt?
3. T14c Wie van de bewoners denk jij wie hij vraagt?

T15context Hannes zegt dat de begunstiger alleen de meest getalenteerde schilder ondersteunt.

1. T15a Wie denk jij wie hij ondersteunt?
2. T15b Wie denk jij wie van de schilders hij ondersteunt?
3. T15c Wie van de schilders denk jij wie hij ondersteunt?

T16context Irina weet zeker dat de conciërge de huurder van de eerste verdieping gaat aangeven.

1. T16a Wie denk jij wie hij gaat aangeven?
2. T16b Wie denk jij wie van de huurders hij gaat aangeven?

3. T16c Wie van de huurders denk jij wie hij gaat aangeven?

T17context Karel vertelde dat de gravin de kandidaat van de PVV steunt.

1. T17a Wie denk jij wie zij steunt?
2. T17b Wie denk jij wie van de kandidaten zij steunt?
3. T17c Wie van de kandidaten denk jij wie zij steunt?

T18context Annette meent dat de professor de kritische onderwijsassistente vermijdt.

1. T18a Wie denk jij wie hij vermijdt?
2. T18b Wie denk jij wie van de onderwijsassistenten hij vermijdt?
3. T18c Wie van de onderwijsassistenten denk jij wie hij vermijdt?