

Received 5 September 2011.

Accepted 26 November 2011.

STANDARD OR DIALECT? A NEW ONLINE ELICITATION METHOD

Marjoleine SLOOS

University of Groningen

m.sloos@rug.nl

Abstract

In dialectology, it is often necessary to obtain a measure for the level of dialectal accent shown by individual speakers, especially if statistical analysis is needed. This also applies to studies on standard variants which are “coloured” by regiolects or dialects. In this paper I explore the feasibility of letting native speakers judge the degree of accentedness in Low-Alemannic German. Specifically, I investigate whether listeners who speak a similar dialect as the speakers who are evaluated assign different judgements than listeners who do not. A novel methodology is applied, which involves an on-line elicitation task using audio files. This experiment shows that listeners who speak different varieties of German form a homogeneous group, with respect to rating the level of accent.

Key words

Alemannic, speech perception, standard language, online survey, sentence shadowing

STANDARD ODER DIALEKT? EINE NEUE ONLINE ELIZITATIONS-TECHNIK

Zusammenfassung

In der Dialektologie wird häufig ein Maß für den Grad des dialektalen Akzents einzelner Sprecher benötigt, insbesondere, wenn eine statistische Analyse benötigt wird. Das gilt auch für Studien über Standardvarietäten, die durch Regiolekte oder Dialekte "gefärbt" sind. In diesem Artikel prüfe ich, inwieweit es möglich ist, Muttersprachler den dialektalen Akzent in nieder-alemannischem Deutsch einschätzen zu lassen. Insbesondere untersuche ich, ob Hörer, die einen ähnlichen Dialekt wie die bewerteten Sprecher sprechen, zu anderen Urteilen kommen als Hörer, die keinen ähnlichen Dialekt sprechen. Eine neuartige Methode mit einer

Online-Elizitations-Aufgabe anhand von Audio-Dateien kommt zum Einsatz. Das Experiment zeigt, dass Hörer, die verschiedene Varietäten des Deutschen sprechen, den erhobenen Bewertungen nach eine homogene Gruppe bilden.

Schlüsselwörter

Alemannisch, Sprachperzeption, Standarddeutsch, Online Elizitations-Aufgabe, shadowing

1. Introduction

The Low-Alemannic dialect is spoken in southwest Germany, where we find shows a continuum between standard German and the dialect. This means that speakers, who usually speak both the standard language as well as the Alemannic dialect, vary in the degree of standard and dialect (or dialectal accent) according to the situation (see also Auer 2005: 3). In more formal occasions they use more standard speech, whereas in casual situations the dialect is used. However, no diglossia or code-switching is used: there are no discrete differences between the different stylistic registers. In the present study, we investigate the influence of the dialect on the Alemannic variety of standard German. Furthermore, we will propose a way of quantifying the place on the dialect-standard language continuum, in other words, the degree of accentedness (DA).

In order to carry out dialectological and sociolinguistic analyses, it is often useful to quantify the degree of dialectal accent as a value on a scale between ‘perfect standard’ and ‘perfect dialect’. The DA is a useful variable in variationist studies of standard languages or regiolects, because different dialects may have a considerable influence on the data. At present, there is no reliable linguistic method to estimate the level of accentedness. First of all, there is no straightforward definition of what is standard and what is dialect. Opinions on the criteria for what is standard differ widely. Usually, the standard is assumed to be a variety that is largely accepted as the most prestigious one. This definition relies on sociological criteria rather than on linguistic ones (Milroy & Milroy 1993). Especially in case of a standard-dialect continuum, it is nearly impossible to set up proper criteria for standard and dialects: which features have to be taken into account to make the standard somewhat less standard? And if there are

no discrete levels, then it is inherently impossible to establish the exact degree of standardness.

Of course, there have been attempts to define the degree of standardness or accentedness. Dialectometry, originally used to express distances between dialects, could also be used to define the distance between the standard and dialects, as shown by Heeringa (2004) and Wieling (2012). Recently, Grieve et al. (2011) developed a method to compute dialect boundaries statistically by a combination of factor analysis, cluster analysis and spatial correlation. Whereas dialectometry aggregates different variables, the latter method resembles the traditional dialectological approach, which relies on determining dialect boundaries on the basis of the variation of particular individual variables. Although this method could also be applied in studies that investigate the standard-dialect continua, it involves a practical problem because fairly complex and time consuming procedures are needed. In addition, both methods require a detailed phonetic analysis of a sufficient amount of data.

A relatively efficient way to gather data on the DA is to ask native speakers to assess the relative level of “standardness”. This can be done on the basis of a small fragment of spoken language. For instance, Smakman & Van Bezooijen (2002) reported that Dutch listeners showed a high degree of consensus in the ratings for judgements about the degree of standardness used in recordings that covered the past few decades.

Intuitively, one would expect that listeners who are dialect speakers themselves would tend to categorize a slight accent in their variety as standard, whereas standard-language speakers would show less “tolerance” regarding the standard. However, Smakman & Van Bezooijen (2002) found that listeners with different demographic and sociolinguistic backgrounds (except for gender, see below) rated speech fragments similarly for the level of standardness. Smakman & Van Bezooijen investigated the rating of standardness from a diachronic point of view: speech fragments from different periods in the standard language are rated for standardness. These authors made no predictions about any correlation between listeners’ age or geographical background and the DA. However, this might be different in synchronic studies, in which different speakers vary in their level of dialectal accent.¹ For instance, we could hypothesize that

¹ Grondelaers et al. (2010) found that speakers with different accents of Dutch were also rated similarly by listeners of different demographic backgrounds. However, this study concentrates on language attitude, rather than on the level of accentedness.

listeners who speak the same variety as the speakers are more likely to accept dialectal features in the standard pronunciation. The present study investigates whether listeners of the same language community as the speakers, viz. the Alemannic variety of standard German, tolerate more dialectal features in what they regard as the standard than listeners from other language communities. This is the main question of the experiment reported on in this paper.

A remarkable finding by Smakman & Van Bezooijen (2002) was that female listeners displayed more tolerance regarding standardness than male speakers, i.e. particular speech fragments were judged as standard by females, but less standard by males. The hypothesis was that females would be more critical than males in judging standardness, and that this would be reflected in lower scores for standardness. The present study will investigate this hypothesis for German as well.

To investigate the DA in the Alemannic variety of Standard German, a new, online elicitation method was created. This proved to be a convenient and reliable method to obtain data in a relatively short period of time. The methodology is described in section 2. On the basis of these results, in section 3 it will be shown that age, as well as gender and the educational level of the speaker are significant predictors of the degree of accent in the standard language. However, it will turn out that there are some listener biases too, which will be discussed in more detail in section 4. Finally, section 5 contains a short conclusion.

2. Methodology

The stimuli were prepared in three stages, which are described in more detail in this section. First, recordings were made of a German female speaker in a neutral variety and this material was used for a shadowing task. Secondly, participants performed a shadowing task and their speech was recorded. Finally, speech samples were made of each participant in the shadowing task, which served as stimuli in the online elicitation.

2.1 Original recordings in Standard German

The original recordings, which served as stimuli in the shadowing task, were obtained from a female speaker speaking a northwestern variant of Standard German, which is usually perceived as the most “neutral” standard pronunciation. These recordings were made in an isolated booth in the New Media Center of Freiburg University in Germany. In sum, 108 sentences were recorded; 72 were collected from the regional newspaper “Badische Zeitung” and 36 were collected for the purpose of another study (Sloos, submitted a). In addition, six sentences were recorded which served for familiarization.

2.2 Shadowed speech recordings

Thirty native speakers of German, who were born and raised in the Alemannic area, and who varied in the degree and use of the dialect, participated in the shadowing task. Speech shadowing, which involves the rapid repetition of auditory stimuli, was first used to show that speakers correct mispronunciations and errors they hear when they have to repeat speech immediately (Marslen-Wilson 1975). Since the subjects are under time pressure while shadowing, their speech is near-spontaneous and they do not correct mispronunciations. By using stimuli in the standard language, the subjects are also forced to use standard language, but because of the time pressure, they will do so in a natural way, i.e. with their own accent.

A well-known fact in sociolinguistics is that older speakers tend towards dialectal varieties more than younger speakers (see e.g. Chambers & Trudgill 1998). To be able to test whether age also plays a role in the DA in the standard, subjects from an age range of 20-77 were selected. Nineteen females and eleven males participated in the recording sessions. The shadowing sessions were conducted at the subjects’ home or office, with E-prime standard 2.0 software (Schneider et al. 2002). The sound files were presented and recorded over Sennheiser PC-151 headphones and microphone with a Marantz PRC 620 recorder. Nine subjects had an intermediate level of education (mid professional education), 21 subjects were highly educated (higher professional or university education). None of the subjects reported hearing or speaking problems. All subjects were born and lived in the Low-Alemannic area, in the southwest of Germany.

Except for three of the youngest subjects (20-30 years), all subjects, regardless of age, gender and education, actively used dialect in their daily lives, according to self-assessment. All subjects stated they had at least a passive command of the dialect.

2.3 Material for the on-line survey

The recordings of the shadowed speech served as stimuli in an online elicitation to investigate the DA. This elicitation was placed on internet and participants were recruited via the personal network of the author and a local association for dialect speakers. One audio sample of four sentences with a duration of approximately 20 seconds was compiled of the recordings of each subject who participated in the shadowing task. To restrict the duration of the survey to maximally 15 minutes, the audio files were divided over two different versions of the online elicitation. In the online survey, each sentence occurred only once.

Additionally, a sample of the speaker who provided the stimulus material was included. This standard-language speaker's sample was identical for both versions. This was done for three reasons: first, this was intended to check the reliability of the test, since a reliable outcome should assign the lowest accent scores to this sample in both versions. Secondly, it was expected that different listeners would use different criteria for assigning DA scores (see below). Finally, listeners may vary in their selection strategies, viz. the range of judgement values could vary per speaker, for instance, some respondents might use the whole range of possible judgements, others might only use three possibilities in the middle of the scale). Since both versions of the survey were completed by different respondents, with different selection strategies, the identical sample of the original recordings served as a reference point.

2.4 Structure of the online survey

Both versions of the survey had exactly the same structure. Some background information, instructions and contact information were provided in written standard German. The elicitation task consisted of three parts: two pages with eight audio samples each and questions about the dialect level and estimated age of the speaker.

Regarding the main task, the judgement of the dialect level, the respondents were asked to select one value out of seven possibilities.

- Perfect standard
- Near standard
- More standard than dialectal
- In between standard and dialect
- More dialectal than standard
- Near dialectal
- Perfect dialectal

Respondents could listen to a sample as many times as they wished before entering their rating. The volume of each audio sample could be adjusted according to the respondents' wishes.

At the end of both pages, the respondents were asked to comment on their criteria (optional). The final part of the survey consisted of a short questionnaire about the respondents' age, gender, and the place where they and their parents had been born and raised. Furthermore the respondents were asked to estimate the relative frequency with which they used dialect on a daily basis (with 1 = only dialect, 5 = only standard) and their competence in both the standard language and dialect (with 1 corresponding to perfect/native command of the standard resp. the dialect, and 5 corresponding to no command over the standard resp. the dialect).

3. Results

In sum, 44 completed submissions were received (see also Table 1). Version A was returned by 23 listeners, 11 of whom were Alemannic (3 females and 8 males) and 12 were other speakers of German (6 females and 6 males). Version B was returned by 21 listeners, 12 Alemannic (6 females and 6 males) and 9 other speakers (6 females and 3 males). The age of the respondents ranged between 18 and 71.

Version	Variety	male	female	sum
A	Alemannic	8	3	11
A	Other	6	6	12
B	Alemannic	6	6	12
B	Other	3	6	9
	sum	23	21	44

Table 1. Respondents of the online survey

A Cronbach's alpha test for agreement (Cronbach 1951) was performed, which scored 0.669 for all respondents to version A, and a remarkably high agreement in version B, namely 0.971, which suggests that the results are reliable. The criteria for the accent rating that were mentioned by the respondents included intonation, deletions, and deviant pronunciation (e.g. diphthongization) of vowels.

In both versions, the speaker of the stimulus material was assigned the lowest score, i.e. the highest level of standard pronunciation, which also suggests that the respondents were reliable on estimating the DA. However, individual differences in selection strategies, in combination with a relatively large scale, led to a slightly different rating for the speaker of the stimulus material: 1.6 in version A and 1.2 in version B. Moreover, the range of values being assigned by listeners in version A and version B of the survey differed. The mean ratings per speaker in version A ranged between 1.6 and 3.7, whereas in version B, the values ranged from 1.2 to 5.3. To compare the two surveys, this has to be modified such that the value for the speaker of the stimulus material as well as the range is identical. Therefore, linear transformation was applied to scale A such that the range of A became identical to the range of version B. The extrapolation formula is provided in (1).

$$(1) \quad y = B_0 + ((x_A - A_0) * (B_n - B_0) / (A_n - A_0))$$

$$y = 1.2 + ((x_A - 1.6) * (5.3 - 1.2) / (3.7 - 1.6))$$

where

y = the degree of accentedness (DA)

x_A = the value in A to be transformed

A₀ = the lowest value in version A

A_n = the highest value in version A

B₀ = the lowest value in version B

B_n = the highest value in version B

This y value, which indicates the DA, was computed for each rating in version A.

Subsequently, linear modelling was applied to the data with the speaker bounded and listener bounded variables, which are explained in Table 2.

	<i>variable type</i>	<i>variable</i>	<i>description</i>
1	speaker bounded	gender	
2	speaker bounded	education	high or mid
3	speaker bounded	real age	
4	speaker bounded	estimated age	age of the speaker as estimated by the listener
5	listener bounded	age	
6	listener bounded	gender	
7	listener bounded	level of dialect	self-estimated level of dialect
8	listener bounded	level of standard	self-estimated level of standard German
9	listener bounded	dialect use	frequency of use on a daily base
10	listener bounded	command over the standard language	self-estimated command over standard German
11	listener bounded	command over the native variety	self-estimated command over the dialect

Table 2. The variables investigated in the online survey

For the speakers, there are three significant predictors: gender, education and age. Males are rated as having a more dialectal accent than females $t = 4.23$, $p < 0.001$ and speakers with an intermediate level of education also scored higher in accent level than speakers with higher education $t = 4.51$, $p < 0.001$. Initially, no effect was found for real age; however, the estimated age of the speakers by the respondents showed a strong positive correlation with the accent level $t = 10.935$, $p < 0.001$ (but see below).

The high degree of homogeneity between the respondents' ratings based on their variety is confirmed by the linear model: no effect can be found for variety. The self-estimated respondents' level of standardness also positively correlated with the rating, $t = 5.01$, $p < 0.001$. Since the highest level of standardness (native variety) is rated as 1, this positive correlation has to be interpreted such that less standard speaking

respondents tend to rate accents as more dialectal than more standard-like speakers. On the other hand, respondents who report using the standard language more often than dialect rate accents higher for the level of accent (i.e. more standard-like) $t = 3.77$, $p = 0.001$. Finally, gender of the respondents also influences the ratings, males assigning higher dialectal accent scores more frequently than females ($t = -3.48$, $p = 0.001$).

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.205	0.548	-4.021	< 0.001***
Speakers' Estimated Age	0.079	0.007	10.935	< 0.001***
Speakers' Education Level <i>mid</i>	0.959	0.213	4.511	< 0.001***
Speakers' gender <i>male</i>	0.865	0.205	4.225	< 0.001***
Respondents' dialect use	0.298	0.079	3.771	< 0.001***
Respondents' gender <i>male</i>	-0.589	0.169	-3.483	< 0.001***
Respondents' Standard Level	0.530	0.105820	5.012	< 0.001***

Table 3. Linear model of the level of dialect in the Alemannic variety of Standard German.

As mentioned above, older people tend to use more dialectal features in their speech than younger people. It is likely that this tendency is reflected in their pronunciation of the standard language. However, we see that the *estimated* age of the speakers by the respondents is a much stronger predictor than their real age (Sloos submitted b). This correlation shows that laypeople are also aware of the fact that older speakers tend to use more dialect, although they project the level of dialect onto the estimation of age. This shows a correlation between the accent level and the estimated age, but of course the estimated age is not a predictor of the accent level. Therefore the analysis was run once more, leaving estimated age out of the model. The results show that in this case real age of the speaker functions as a predictor of the accent level (Table 4).

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.153	0.577	-0.265	0.791
Speakers' Age	0.025	0.005	4.702	< 0.001***
Speakers' Education Level <i>mid</i>	1.578	0.229	6.881	< 0.001***
Speakers' Gender <i>male</i>	1.493	0.214	6.963	< 0.001 ***
Respondents' Dialect Use	0.229	0.086	2.671	0.008**
Respondents' Gender <i>male</i>	-0.921	0.177	-5.206	<0.001***
Respondents' Standard Level	0.587	0.116	5.086	<0.001***

Table 4. Linear model of the level of dialect in the Alemannic variety of Standard German (real age instead of estimated age).

Table 4 shows that the ratings for the degree of accent correlate with the age, education level and gender of the speakers as well as the dialect use, gender and degree of standardness of the listeners.

4. Listeners' biases

The speakers' age, gender and level of education are strong predictors for the rating of their accent level. Older speakers, males and lower educated speakers tend to a higher DA. This is in line with many sociolinguistic studies that showed that so-called NORMs (Non-Mobile Older Rural Males) tend towards the most dialectal speech. More surprisingly, as respondents, male respondents tend to assign lower accent levels than female respondents. This means that they are more "tolerant" towards the standard variety, which is sharply in contrast with the results of Smakman & Van Bezooijen (2002), who found that the females in their survey showed more "tolerance" towards the standard. This is probably correlated with the respondents' dialect use and their standard level. Dialect use refers to proportion of dialect and standard that people use on a daily basis. Males tend towards more frequent dialect use and one could argue that the level of standard language use by men is probably lower on average than that by women. If it would be the case that males rate others and themselves alike, they are therefore expected to rate lower for accent level than females. In order to test whether the

respondents' dialect levels correlate with the other respondents' bound variables, a post hoc linear model on the self-estimated respondents' dialect level was performed. Whereas the respondents' gender and respondents' dialect use are both significant predictors of the dialect level (gender $t = -5.32$, $p < 0.001$ and dialect use $t = 11.7$, $p < 0.001$), this is not true for respondents' age and respondents' standard level, as shown by Table 5.

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.637	0.359	0.775	0.0764
Respondents' Gender <i>male</i>	-0.537	0.101	-5.322	< 0.001***
Respondents' Age	-0.003	0.003	-0.899	0.369
Respondents' Dialect Use	0.597	0.051	11.672	< 0.001***
Respondents' Standard Level	0.030	0.066	0.458	0.647

Table 5. Linear model of the respondents' dialect level

As Table 5 shows, the respondents' dialect level corresponds with their dialect use, and also with their gender. It shows that respondents assign *lower* ratings to their own dialect level, meaning they consider themselves as speaking *more* dialectally than females.

5. Conclusion

In this paper, the ability of listeners to judge the accent of the Alemannic variety of standard German reliably was investigated. This investigation made use of an online experiment, which could be performed in listeners' homes and provides reliable results in a relatively short amount of time. It turned out that regardless of the variety the respondents themselves speak, the homogeneity of their judgements is relatively high. This suggests that this procedure is an adequate way of estimating the level of accent. This is confirmed by the fact that the speaker of the most neutral variety was consistently rated as the most standard speaker. Moreover, we did not find evidence that the variety that the listeners speak, and whether this variety is the same as the variety

under investigation, has an effect on their ratings. As predicted, the ratings showed that age, gender and educational level of the speakers contribute to the degree of accentedness. But there are also some listener biases: gender, dialect use and self-estimated command of the standard are also predictors for the ratings. This suggests that, ideally, in order to obtain the most homogeneous ratings, listeners should be selected more carefully and gender, dialect use and command of the standard should be controlled for.

Acknowledgements

This paper benefitted much of valuable comments of Charlotte Gooskens and Jeroen van de Weijer. Furthermore, I am very much indebted to the Müttersproochsgesellschaft ‘Native Speakers Association’ (of Alemannic) in Freiburg, for contacting the subjects for the recordings and respondents for the survey.

References

- AUER, Peter (2005) “Europe’s sociolinguistic unity, or: A typology of European dialect/standard constellations”, in N. DELBEQUE, J. van der AUWERA & D. GEERAERTS (eds.), *Perspectives on Variation*, Berlin/New York: De Gruyter Mouton, 65,7-42.
- CHAMBERS, JACK. K. & Peter TRUDGILL (1998) *Dialectology*, Cambridge: Cambridge University Press.
- CRONBACH, Lee J. (1951) “Coefficient alpha and the internal structure of tests”, *Psychometrika*, 16, 297-334.
- GRIEVE, Jack, Dirk SPEELMAN & Dirk GEERAERTS (2011) “A statistical method for the identification and aggregation of regional linguistic variation”, *Language Variation and Change*, 23, 1-29.
- GRONDELAERS, Stefan, Roeland VAN HOUT & MIEKE STEEGS (2010) “Evaluating Regional Accent Variation in Standard Dutch”, *Journal of Language and Social Psychology*, 29, 1, 101-116.
- HEERINGA, Wilbert (2004) *Measuring dialect pronunciation differences using Levenshtein distances*, PhD Dissertation, Groningen University.

- MARSLEN-WILSON, W. D. (1975) "Sentence perception as an interactive parallel process", *Science* 189: 4198, 226-228.
- MILROY, James & Lesley MILROY (1992) "Social network and social class: Toward an integrated sociolinguistic model", *Language and Society*, Cambridge: Cambridge University Press, 1-26.
- SCHNEIDER, Walter, Amy ESHMAN & Anthony ZUCCOLOTTO (2002) "E-Prime: A User's Guide", Psychology Software Tools, Pittsburgh: PA.
- SLOOS, Marjoleine (submitted a) "Interaction between frequency and morphological category in the Alemannic variety of Standard German".
- SLOOS, Marjoleine (submitted b) "Coders versus acoustics: a linguistic halo effect", *Labphon*.
- SMAKMAN, Dick & Renée Van BEZOOIJEN (2002) "Pronunciation in standard Dutch", in J. BERNIS & J. van MARLE (eds.), *Present-day dialectology*, Berlin: Mouton de Gruyter, 333-352.
- WIELING, Martijn (2012) *A Quantitative Approach to Social and Geographical Dialect Variation*, PhD Dissertation, Groningen University.