

German Today: An Extensive Speech Data Collection in the German Speaking Area of Europe

Ralf Knöbl,¹ Stefan Kleiner,¹
Caren Brinckmann¹ and Nina Berend¹

Abstract

The research project “German Today” aims to determine the amount of regional variation in (near-) standard German spoken by young and older educated adults, and to identify and locate the regional features. To this end, an extensive corpus of read and spontaneous speech is currently being compiled.

German is a so-called pluricentric language. With our corpus we aim to determine whether national or regional standards really exist. Furthermore, the linguistic variation due to different contextual styles (read vs. spontaneous speech) shall be analysed. Finally, the corpus will enable us to investigate whether linguistic change has occurred in the domain of the German standard language. The main focus of all research questions is on phonetic variation (lexical variation is only of minor interest).

Read and spontaneous speech of four secondary school students (aged seventeen to twenty) and two fifty- to sixty-year-olds is recorded in 160 cities throughout the German-speaking area of Europe. All participants read a number of short texts and word lists, name pictures, translate from English, and take part in a sociobiographic interview and a map task experiment. The resulting corpus will comprise over 1000 hours of orthographically and (in part) phonetically transcribed speech.

1 Introduction

Due to historical reasons, spoken German does not have one single standard form but rather a multitude of national or regional standards which are influenced by the respective dialects.

In many dialectological studies and atlases, the phonetic domain has been thoroughly analysed and mapped.² However, research concerning the opposite side of the linguistic continuum, namely the German standard language, has been remarkably sparse. It is a commonly known fact that almost everywhere in the German-speaking area, even in most formal speaking styles, regionalisms of some kind can be detected. It is also generally known that these regionalisms are less prominent in most northern regions of Germany where the traditional dialects have ceased to be spoken among the younger generation.

In all regions where the traditional dialects continue to be everyday vernaculars they exert a strong phonetic influence on the locally spoken Standard German. But due to a high mobility of the working population and the influence of the mass

¹ Institut für Deutsche Sprache, Mannheim, Germany

e-mail: knoeb1@ids-mannheim.de, kleiner@ids-mannheim.de, brinckmann@ids-mannheim.de, berend@ids-mannheim.de

² see *Marburger Sprachatlanten*, <http://web.uni-marburg.de/dsa/atlantent.html> (accessed: June 24, 2007).

media, there has been an ongoing trend for linguistic assimilation of the local dialects to the standard language along with a sharp overall decline in dialect speakers. This recent development has been especially strong in the urbanised regions of central and southern Germany.



Figure 1: The traditional dialect areas of German (from Dingeldein, 1992: 30)

Nevertheless there are only few studies aiming at a comprehensive description of the features of regionalised Standard German (see 1.2). To fill this scientific gap, our project aims to describe the phonetic features and their areal range in the whole area where German is used as an official language (i.e. Germany, Austria, Switzerland, Luxemburg, South Tyrol, Eastern Belgium and Liechtenstein).

1.1 Research questions

The following research questions shall be answered in the course of our project:

- How much regional variation in near-standard speech can be found in younger vs. older educated people?
- Which regional features are still in use and where?
- Have new isoglosses emerged at (relatively) new political borders, i.e. do national standards (Germany vs. Austria vs. Switzerland) or regional standards (e.g. Bavarian Standard German) exist?
- How do empirically collected pronunciation data differ from the forms codified in pronunciation dictionaries?

- What kind of linguistic forms do we find in different contextual styles (read vs. spontaneous speech)?

1.2 Previous studies

The only fully comparable previous study is the *Atlas zur Aussprache des Schriftdeutschen in der Bundesrepublik Deutschland* (König, 1989). The speech data that was used for its compilation was collected in 1976-77. The speech of forty-four university students from forty-four different places spread homogeneously over the whole area of then West Germany was recorded. The students read a wordlist, a text passage and answered questions in a short sociobiographic interview. Only the data from the wordlists has been transcribed in narrow phonetic transcript and analysed. Comparable empirical studies exist for Switzerland (Hove, 2002) and Austria (Bürkle, 1995). Unfortunately, there is no comparable study for the area of the former German Democratic Republic.

2 Concept and design

2.1 Stimuli

Several different stimuli are used for speech elicitation:

- two short texts (*Northwind and Sun* + 500-word popular scientific text)
- 800-word text/sentences (specifically compiled for this corpus)
- seventy-five pictures (picture naming)
- twenty-five English words and ten English sentences (for translation into German)
- word list with approx. 1000 words (including minimal pairs)
- sociobiographic interview (approx. thirty minutes)
- map task experiment (Anderson et al., 1991; approx. fifteen minutes).

The different stimuli are mainly used to elicit different levels of formality in speech, especially the read vs. spontaneous contrast is an important issue. In the spontaneous speech domain, the sociobiographic interview is carried out between a researcher and a local participant, whereas in the map task experiment two participants from the same city are interacting. Thus, the interview may yield radically different linguistic forms in comparison to the more informal map task situation. This is primarily the case in areas where dialect is the everyday vernacular among the participants (esp. in Switzerland). Picture naming and translation from English are used to check if certain words are pronounced differently when they are written down and read out loud or elicited without providing the written form.

2.2 Cities selected for the recordings

The cities where the participants were recorded were selected according to different criteria. Firstly, the forty-four cities in former West Germany analysed in König (1989) had to be included in our survey. In that way, real-time language change in the

past thirty years at these places may be detected. All in all, we plan to include 160 places in our recordings. Thus, the whole German-speaking area can be covered with a grid of recordings that leaves out none of the traditional dialect areas of the German language so that we shall be able to detect even minor deviations in the regional standards. Recordings take place at population centres (e.g. Berlin, Hamburg, Cologne, Munich, Vienna, Zurich) as well as small towns in sparsely populated areas.



Figure 2: Cities selected for the recordings

2.3 Recordings

2.3.1 Participants

For our main corpus, four speakers (two female, two male) aged seventeen to twenty are recruited at one secondary school in each of the 160 cities. For our secondary corpus, two speakers (one female, one male) aged fifty to sixty are recruited at adult education centres in eighty cities (out of the 160). Both groups of participants have secondary school education. A further requirement is that the participant has to be born at or near the place of recording and at least one of her/his parents has to be from the region as well.

2.3.2 Recruitment of the participants

For the acquisition of participants for the main corpus, a letter with a request for cooperation and a short description of the project is sent to the headmaster of a secondary school in the selected cities. The spontaneous positive responses to this letter come close to 40 percent. By following e-mail requests and in some cases a second letter, the positive responses could be raised to about 60 percent. For the remaining 40 percent of the cities where either no response at all or a negative response was given, another secondary school is selected. In case there should be only one secondary school in town, we have to switch to another small town in the vicinity.

The problem of negative responses to our request for cooperation is especially prominent with our recordings at the adult education centres because even in big cities there is only one such institution. On top of that, the positive responses from the adult education centres are only at about 30 percent. We have yet to find a different strategy for the recruitment of the older participants in our study.

2.3.3 Field recordings and equipment

For the field recordings several microphone and recorder types were tested. Finally, a combination of Sennheiser HSP4 headset cardioid microphones and Marantz PMD671 solid state recorders was chosen for the recordings of the main corpus. Initial problems with hum on some field recordings were countered by using rechargeable batteries instead of AC mains power. (For most of the recordings at the adult education centres a Sennheiser MKE2 omni-directional clip-on microphone and a Mayah FlashMan solid state recorder are used).

The recording quality is 16 Bit, 44,1 kHz. For the individual tasks (reading, picture naming, translation) mono recordings are made, in the interviews and map tasks both participants use separate microphones and their speech is recorded on separate audio tracks.

The recordings are carried out by one researcher. Due to the fact that the recordings take place in schools, there is sometimes significant ambient noise. Although the schools are usually able to provide a room in a quiet part of the school building, it can become very noisy on the floors, especially during recesses. Sometimes sound reflection from the walls is a problem, too, when the recordings take place in large rooms with blank walls (classrooms). However, both kinds of problems are alleviated by the headset cardioid microphones that are very effective at suppressing all ambient noise.

2.4 Documentation

All recordings have to be documented in a standardised fashion in order to be usable for statistical analyses. For example, for comparing younger and older speakers we have to document their age at the time of the recordings. In order to unify the documentation for all existing and future in-house corpora, several projects collaborated with the in-house archive and developed a set of standardised XML metadata schemas. These XML schemas build on existing internal and external documentation schemas such as IMDI (2003, 2004) and OLAC (2006) and take into account the workflow of speech corpus production. Another major objective was to minimise re-

dundancy which was achieved by designing separate schemas for documenting speakers, recording sessions, and entire corpora.

The resulting schemas are currently being tested within our project and are regularly revised following consultation with the archive and other projects. The XML schemas are accompanied by a manual containing element definitions, guidelines, and examples.

2.5 Annotation

Annotating over 1000 hours of read and spontaneous speech is a rather daunting task. As a first step, the speech data is transcribed orthographically. Similar to our efforts regarding the standardisation of documentation schemas, the conventions for orthographic transcriptions of spontaneous speech are discussed and agreed upon between several in-house projects, building on existing transcription conventions (Goedertier and Goddijn, 2000; Kohler et al., 1994). For example, punctuation is not marked and lexical capitalisation is applied.

The orthographic transcription of read speech is carried out semi-automatically using a Praat script (Boersma and Weenink, 2007) that fills the intervals between manually set word boundaries with the respective orthographic material. Spontaneous speech is transcribed completely manually in 2-3 second inter-pause stretches using Praat. Especially the map task data can be rather dialectal, making it hard to understand for transcribers not familiar with the respective dialect. All in all, read speech can be transcribed much faster and more reliably than spontaneous speech.

Since our resources are somewhat limited, up to now primarily read speech has been transcribed and aligned. By focussing on read speech instead of transcribing the whole corpus in a non-selective way, it will be possible to carry out all analyses pertaining to regional variation without having to wait several years only for the orthographic annotation to be complete. Nevertheless, the interviews are also being transcribed, at least in part. In addition, two avenues to overcome the transcription bottleneck are currently being explored:

1. evaluation of available automatic speech recognition software (for spontaneous speech) and alignment tools (for read speech) in terms of robustness and usability
2. crowdsourcing the orthographic transcription task (e.g. similar to Distributed Proofreaders¹).

In the future, broad phonetic transcriptions aligned on the phone level and several variationalist annotations will be added to the orthographically transcribed speech. Speech data, transcriptions and annotations, as well as documentation data are stored in a database.

3 Concluding remarks: current state of the project

Currently, the project has reached the final stage of the recording phase. Up to this date (30 June 2007) we have made recordings at 110 secondary schools and 45 adult

¹ <http://www.pgdp.net/c/> (accessed: June 24, 2007).

education centres (525 participants). Orthographic annotations include the read speech of 250 and the interviews of 50 participants. We expect to finalise our recordings by the end of 2007. The orthographic transcription of the read data will be completed by summer 2008.

References

- Anderson, A., M. Bader, E. Bard, E. Boyle, G. M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson and R. Weinert (1991) 'The HCRC Map Task Corpus'. *Language and Speech* 34, 351–6.
- Boersma, P. and D. Weenink (2007) Praat: doing phonetics by computer (Version 4.6.09) [Computer program]. Available on-line from <http://www.praat.org/> (accessed: 24 June 2007).
- Bürkle, M. (1995) Zur Aussprache der unbetonten Silben im österreichischen Standarddeutsch. Die unbetonten Silben. Frankfurt am Main: Lang.
- Dingeldein, H. (1992) 'Deutsch in Deutschland'. *Konturen. Magazin für Sprache, Literatur und Landschaft* 1/1992, 29–3.
- Goedertier, W. and W. Goddijn (2000) Protocol voor Orthografische Transcriptie. Available on-line from http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/annot/orthography/ort_prot.pdf (accessed: 24 June 2007).
- Hove, I. (2002) Die Aussprache der Standardsprache in der deutschen Schweiz. Tübingen: Niemeyer.
- IMDI Team (2003) IMDI Metadata Elements for Session Descriptions, Version 3.0.4. Nijmegen: MPI. http://www.mpi.nl/IMDI/documents/Proposals/IMDI_MetaData_3.0.4.pdf (accessed: 24 June 2007).
- IMDI Team (2004) IMDI Metadata Elements for Catalogue Descriptions, Version 3.0.0. Nijmegen: MPI. http://www.mpi.nl/IMDI/documents/Proposals/IMDI_Catalogue_3.0.0.pdf (accessed: 24 June 2007).
- Kohler, K., G. Lex, M. Pätzold, M. Scheffers, A. Simpson, W. Thon (1994) Handbuch zur Datenaufnahme und Transliteration in TP14 von VERBMOBIL – .0. Kiel: IPDS.
- König, W. (1989) Atlas zur Aussprache des Schriftdeutschen in der Bundesrepublik Deutschland. Ismaning: Hueber.
- OLAC Metadata (2006) <http://www.language-archives.org/OLAC/metadata-20060405.html> (accessed: 24 June 2007).