

Digitising a corpus of Austrian dialect recordings from the 20th century

Christian Huber

Phonogrammarchiv
Austrian Academy of Sciences
christian.huber@oeaw.ac.at

Benjamin Fischer

Phonogrammarchiv & ACDH-CH
Austrian Academy of Sciences
benjamin.fischer@oeaw.ac.at

Abstract

The *Corpus of Austrian Dialect Recordings from the 20th Century* comprises 2442 dialect recordings from the *Phonogrammarchiv*'s holdings on magnetic tape from fieldwork conducted in the years 1951 to 1995 by German philologists Eberhard Kranzmayer, Maria Hornung, Werner Bauer, Herbert Tatzreiter and others. They cover all provinces of Austria as well as linguistic varieties of German spoken in Northern Italy, Hungary and former Yugoslavia and Czechoslovakia. In a project cooperation of the *Phonogrammarchiv* with the Research Department "Variation and Change of German in Austria" (both Austrian Academy of Sciences) and the Austrian Science Fund Special Research Programme "German in Austria" (F60), these recordings are now digitised, annotated and analysed, and will be made searchable in a database. In this article we introduce and discuss the corpus and address various digitisation and metadata-related issues with special attention to real-world questions and problems encountered.

Keywords: Dialect recordings, dialect corpora, digitisation, metadata, geodata

1 Introduction

Languages and their dialects are subject to constant change. Older varieties are gradually passing out of usage and linguistic phenomena connected to them have often vanished from the language of everyday life. They can be witnessed authentically only in contemporaneous audio recordings, which are available to varying degrees from the late 19th century onwards. However, the use of historically grown collections of language recordings creates challenges that do not arise in modern corpora, which are generated within a specific research context and infrastructure. In historic corpora, recordings were often made not only at different times but also with different objectives, adhering to different methods, using different recording technologies and employing different documentation practices. That is, before such corpora can be subject to further exploitation in linguistic or other research, one has to deal with questions of data organisation as well as the preservation of their sonic content.

The accessibility and sound quality of historic recordings depends on the state of preservation of their carrier media. Traditional analogue sound carriers such as wax cylinders, gramophone discs or magnetic tape are subject to natural decay: the more critical the condition of the carrier, the worse the signal quality. Once the carrier can no longer be played, the recordings on it are lost forever. Therefore, it is necessary to digitise perishable sound documents as long as the carriers can still be properly played in order to preserve the recorded contents in the long term and make them available for future generations. Digital audio data are no longer bound to an individual data carrier but can be losslessly copied as often as desired. In this way they can be electronically preserved for a virtually indefinite period of time.

In the same vein, all existing written documentation on paper (likewise perishable) needs to be digitised and made searchable in order to preserve as well as to make accessible all information that goes beyond what can be heard on the recordings. Historic metadata may be organised according to

different principles and may be available in varying detail. To ensure a high level of searchability, they must be structured in a uniform way and, where possible, be enriched and emended.

While details may vary to some degree, depending on the respective corpus or collection, these issues are vital in making a corpus of historic linguistic recordings usable for further research procedures. However, the question of how to prepare such a corpus for further use, generally, does not seem to have received much attention in the literature. Therefore, in this article, we will pay due attention to these questions from the perspective of the corpus at hand.¹

1.1 The cooperation

The present project is a cooperation of the *Phonogrammarchiv* with the Research Department “Variation and Change of German in Austria” (both institutions of the Austrian Academy of Sciences) and the Austrian Science Fund Special Research Programme “German in Austria” (F60). Its aim is to digitise, annotate and analyse Austrian dialect recordings on magnetic tape from the *Phonogrammarchiv*’s holdings covering five decades (1950s–1990s) and provide a structured and searchable description in order to ensure their lasting preservation and usability. Despite the significant scientific interest, the systematic evaluation of these recordings as well as their general accessibility have hitherto been hampered by the fact that they existed only in an analogue format. This requires special storage conditions and players, which are nowadays available only in specialised institutions. Transferring them to an easily accessible and easy-to-handle digital format and enhancing them with an adequate, carefully designed electronic metadata description and database ensures the usability and searchability of the corpus not only for linguistic purposes but makes it possible to exploit these recordings in wider scientific as well as general context. The data will be used in research by the Research Department “Variation and Change of German in Austria” and in various subprojects of the Special Research Programme “German in Austria”.

However, the recordings are not only of purely linguistic interest but also offer material relevant to disciplines such as folklore and folklife studies or research on oral history and oral traditions. Therefore, the searchable corpus will be further developed into an online platform and this way become accessible to other interested researchers and also to the general public.²

The project started in January 2019 and, after completing the digitisation of the analogue audio tapes, it is presently concerned primarily with systematising and enriching the descriptive metadata based on the available archival materials. A first stage (2019) was dedicated to digitising and archiving a subcorpus of 1768 dialect recordings from 1951 to 1983, based on a set of recordings included in UNESCO’s national register “Memory of Austria” in 2018. A further 674 recordings are currently (2020/2021) being dealt with in a second stage.

1.2 Organisation of the article

In section 2 we introduce the corpus. We briefly outline the history of sound recordings of Austrian varieties of German in the *Phonogrammarchiv* and then proceed with addressing the phases of fieldwork that yielded the recordings in the present corpus. Additionally, we discuss the methods employed in the field by various surveyors as compared to previous approaches and approaches taken by other dialect survey projects. We also compare the methods used by various researchers included in our corpus, taking into account how they analyse their audio recordings and incorporate them into

¹ The present article updates Huber et al. (2019), which reflects the initial stage of the project.

² For more detail on the cooperation of the *Phonogrammarchiv* with the Research Department “Variation and Change of German in Austria” and the Special Research Programme “German in Austria” see also Lenz et al. (2020).

their research. Since the present corpus represents only a part of the *Phonogrammarchiv*'s holdings of dialect recordings, we explicate the factors that played a role in assembling and partitioning it in its present shape. We explain what makes this corpus exceptional and sets it apart from traditional field studies in German dialectology at that time and point out some of the possibilities it offers for sociolinguistic investigations and research into historical dialect data.

Section 3 is concerned with the procedures undertaken in the *Phonogrammarchiv* in order to prepare the corpus for further use. We address various digitisation and metadata-related issues from an archival perspective and pay special attention to real-world questions and problems as encountered in the case at hand and discuss the approaches we chose to overcome them.

In section 4 we present some statistical details as they emerge from the presently available data and discuss our approach to dealing with problems in representing geographical metadata in various places in the electronic documentation. We also touch upon the ethical and legal questions the project will have to cope with in the dissemination of the linguistic and audio data.

Section 5 presents a brief outlook.

2 A corpus of dialect recordings on magnetic tape

Employing sound recordings for documenting linguistic varieties of German played a prominent role at the *Phonogrammarchiv* (founded in 1899) from the outset. The *Phonogrammarchiv*'s first German dialect recordings were made in 1901,³ and in 1903 the documentation of Austrian dialects was defined as one of its major tasks (Hajek 1928a: 11, Schabus 1999: 26). Consequently, in the following years a great number of samples of varieties of German spoken in the Austro-Hungarian empire were recorded by researchers such as Josef Schatz (Tyrol and Vorarlberg, 1903–1906), Primus Lessiak (Carinthia, 1903–1904), Franz Scheirl (Salzburg, 1905), Joseph Seemüller (various areas in the Austro-Hungarian empire and Alsace, 1906–1911), Jakob Reimer (Lower Austria, 1909), or Walter Steinhauser (1911–1913) and Anton Pfalz (various German-speaking areas).⁴ These recordings include elicited questionnaire data (primarily the 40 *Wenker sentences*) as well as narratives and folk songs. According to Pollak (1913: 87), Seemüller introduced a practice of recording the Wenker sentences as well as a piece of spontaneous speech when making recordings of German varieties.

However, making sound recordings in these early days was hampered by a number of factors: the recording apparatus was bulky and heavy and not easy to transport, the data carriers were expensive and available only in small quantities, and the maximal duration of a recording was limited to about two (later four) minutes. As a result, what was to be recorded was often rehearsed with the informants before the actual recording was made.⁵

The advent of magnetic tape recording marked a major change in linguistic fieldwork practice as it became possible now to make recordings of longer duration and, through advancements in microphone technology, with a decent sound quality. In the *Phonogrammarchiv*, the use of magnetic tape recording to support linguistic fieldwork began in 1951 after the *Phonogrammarchiv* had received a *Philips Maestro* tape machine as a gift (Lechleitner 1999: 18), which was immediately employed in dialectological research.

³ Fritz Hauser's recordings of the dialect of Attergau (Ph 105) and of Silesian German from Czechia (Ph 106), see <http://catalog.phonogrammarchiv.at/>.

⁴ For these and other early linguistic recordings in the *Phonogrammarchiv* see the online catalogue at <http://catalog.phonogrammarchiv.at/sessions.php?onlineSessionSuchmodus=strukturierteSuche>.

⁵ The early recordings were made by means of the *Archiv Phonograph* designed by Fritz Hauser, of which over time five versions were devised. For the development of audio recording technology in the *Phonogrammarchiv* and the use of electrically recorded disc technology (which did not allow recording in the field) in the 1930s and 1940s see Lechleitner (1999), for an overview of the *Phonogrammarchiv*'s language recordings from the 20th century see Schabus (1999).

2.1 The fieldwork

Initiated by Eberhard Kranzmayer (who had been involved in dialectological fieldwork since the 1920s) and later directed mainly by Maria Hornung, the recording activities (see Figure 1) commenced in 1951 as a cooperation of the *Phonogrammarchiv* and the so-called “Wörterbuchkanzlei” (i.e. *Kommission zur Schaffung des Österreichisch-Bayerischen Wörterbuches*, from 1969 on *Kommission für Mundartkunde und Namenforschung*) with the express purpose of creating “ein umfassendes Tonarchiv echter, lebensgetreuer Mundartaufnahmen [a comprehensive sound archive of genuine, true-to-life dialect recordings]” (Hornung 1961: 185) from Austria and linguistic enclaves in neighbouring countries as a continuation of the tradition of phonographic dialectological documentation that had started in the beginning of the 20th century (Kranzmayer 1965: 129). Throughout the 1950s and 1960s, large-scale fieldwork on dialectal varieties of German was conducted by Kranzmayer and Hornung as well as their students and other friendly researchers (e.g. Leopold Kretzenbacher, Hans Hönigschmied, Erwin Koch-Emmery, Eugen Gabriel, Alois Brandstetter, Friedrich Kainz, Oskar Pausch and Hans Pusch) in numerous places in all provinces of Austria and relevant adjacent areas.

As is not uncommon in linguistic fieldwork, the recordings were not always made in the place where a particular variety was spoken. Especially in the early years, speakers from different villages were often not recorded in the villages where they actually lived but were assembled in a nearby place that offered proper recording conditions. Therefore, the village where the recording session took place is oftentimes not the village from where the sampled dialect originates. In accordance with the actual demographic situation, beside varieties of German, samples of other Austrian regional autochthonous languages were recorded (Croatian, Hungarian, Romani, Slovenian, and in addition a few samples of Hebrew) as well. Some 10 to 20 years later, selected individuals were recorded again for comparative purposes, and Hornung carried out a systematic post-survey in the province of Burgenland in 1975, up to 23 years after the initial fieldwork had taken place (see also Hornung 1999: 89).⁶

In the 1970s and 1980s, a large number of recordings were generated by Werner Bauer and Herbert Tatzreiter. Additional recordings stem from fieldwork by Wilfried Schabus, Heinz-Karl Stark, Ingeborg Schönhuber (married Geyer), Günter Lipold, Barbara Jocher, Helga Hiermanseder (married Ebner) Hermann Scheuringer, Roland Girtler, Friedrich Bouterwek or Franz Patocka. Beside her post-survey of Burgenland, Hornung now concentrated on fieldwork in South Tyrol and other German-speaking enclaves in upper Italy (Kranzmayer, who died in 1975, had withdrawn from phonographic fieldwork after the 1960s). Likewise covering all Austrian provinces (with the exception of Vorarlberg) and language islands in neighbouring countries, the recordings from the 1970s and 1980s offer comparative materials to the recordings from the 1950s and 1960s at a temporal distance of 20 to 30 years. Later, the dialectological documentation focus largely shifted to language islands in Romania and the Americas.

The recordings represent the first attempt at a comprehensive nation-wide audio documentation of linguistic varieties in Austria. They provide an exceptional picture of Austria’s dialectal landscape after the middle of the 20th century across all regions. Furthermore, they constitute a source of historical dialect audio data that is unique in the German-speaking area and is a valuable primary source for various areas of research.

⁶ According to Schabus (1999: 28), Hornung had projected another post-survey in Burgenland for the late 1990s or early 2000s, which apparently did not take place. Her last recordings in the *Phonogrammarchiv*’s holdings date from 1991 (South Tyrol). Hornung (1999: 90) says it was about time to conduct another post-survey and suggests to the “Hianzenverein” (an association to promote the preservation the Hianzen dialect) to become active in the matter, but no further information is available to us at this point.



Figure 1. Fieldwork in Timau/Tischelwang, Italy (1959, left) and Neusiedl, Austria (1975, right) (©Phonogrammarchiv)

2.2 Field methods and data analysis

In their documentation work, Kranzmayer and Hornung laid emphasis on geographical rather than social variation, envisaging the reconstruction of a historical dialect situation before the intrusion of modern influences.⁷ Their ideal informant was of older age, nonmobile, and with minimal formal education (see e.g. also Bauer 1967: 22, Tatzreiter 1978a: 133–134), i.e. a NORM or NORF.⁸ True to the dialectological tradition of their day, they aimed at capturing “die älteste erreichbare Schichte mundartlicher Ausdrucksweise [the oldest accessible stratum of vernacular expression]” by targeting old speakers who were also “auf ihre Bodenständigkeit geprüft [checked for their being rooted in their environment]”.⁹ The emerging collection of sound recordings was considered “ein Denkmal altbäuerlich-österreichischer Sprach- und Volkskultur [a monument of old rural Austrian language and folk culture]” (all quotes from Hornung 1961: 186) and “ein unersetzliches Denkmal österreichischer Wesensart [an irreplaceable monument of Austrian character]” (Hornung 1967: 50). Thus, the

⁷ For a different approach adopted in Bausinger and Ruoff’s recordings of Austrian Alemannic varieties in the mid-1950s see Gabriel (1972a: 185).

⁸ *Nonmobile, older, rural males or females*, see Chambers and Trudgill (1998: 29), Schallert (2014: 140 footnote 12). The gender ratio in our corpus is roughly 64% male vs. 36% female informants, for details see below section 4.1.

⁹ Over time, however, they of course also recorded speakers from various age groups and social backgrounds.

recordings were made not only in pursuit of purely linguistic questions but also with an eye (or two) to folklore and folklife studies.¹⁰

Research into German dialects had formerly prominently employed the so-called *Wenker sentences*, i.e. 40 preformulated sentences in standard German designed by Georg Wenker in the late 19th century that had to be translated by the language consultants into their respective native dialects (see Fleischer 2017 for some discussion). While Hornung (1961: 184–185) acknowledges the usefulness of the Wenker sentences for the German Linguistic Atlas and for early comparative dialectology, she criticises them as inadequate and problematic because of the priming effects in their elicitation and also reports of informants criticising them because of their contents. She estimates that 90 percent of such recordings do not represent actually spoken natural language (“In 90 von 100 Fällen haben solche Aufnahmen etwas Gesuchtes und entsprechen nicht der Realität [In 90 out of 100 cases such recordings have something sought-after and do not correspond to reality]”).¹¹

Rather than eliciting comprehensive wordlists or questionnaires, therefore, Kranzmayer and Hornung chose instead to rely on recordings of spontaneous language, and so they stimulated their informants to speak freely about a topic suggested by the interviewer or of their own choice, or recorded them in (equally stimulated) conversation during a session. In this way, Hornung and Kranzmayer sought to document dialectal varieties in their actual use as well as syntactic or prosodic phenomena that would not readily emerge from answering a questionnaire, capturing on tape “in gänzlich unbeeinflussten naturgetreuen Sprachproben ohne jede Vorbereitung die ureigene Sprechweise eines Ortes, einer Altersstufe [the very own way of speaking of a place, an age group, in completely uninfluenced, lifelike speech specimens without any preparation]” (Kranzmayer 1965: 129). In addition, Hornung (1964a: 10) acknowledges the “stete Kontrollmöglichkeit der schriftlich aufgezeichneten mundartlichen Lautungen [constant possibility of checking the dialectal speech sounds recorded in writing]” beside the “unfehlbare Wiedergabe des natürlichen Wort- und Satzakkentes verschiedenaltiger Sprecher aus zahlreichen Orten [infallible reproduction of the natural word and sentence accent of speakers of different ages from numerous villages]” and concedes that such audio materials collected directly from the speakers make it possible to apply “neuartige, exakte Arbeitsmethoden [novel, exact working methods]”.

Informants were approached using what Hornung refers to as the “psychologische Methode [psychological method]” (Hornung 1961: 186), which asked for a certain sensitivity on the fieldworker’s part. Basically, this method consisted of creating a relaxed session atmosphere (perhaps

¹⁰ In fact, Hornung frequently stresses the importance of folklore and folklife studies within the framework of her dialectological research as well as in her fieldwork. For example, in Hornung (1960: 1) the excursion undertaken in 1959 by her and her assistants to the German language islands of Pladen (Sappada) and Zahre (Sauris) in Upper Carnia, Italy, is referred to as a “eingehende volks- und dialektkundliche Forschungsfahrt [thorough folkloristic and dialectological research trip]”. Likewise, when reporting on the first decade of recording audio samples of Austrian dialects, Hornung (1961: 190) states that for gathering lexical data, “various folkloric subject areas were systematically elicited [verschiedene volkskundliche Sparten wurden systematisch gefragt]”. Publications on folkloric topics include Hornung (1958) and (1964b).

¹¹ A similar judgement is found in Gabriel (1965: 49), who remarks that “[Wenker] stellte 40 Sätze zusammen, die er zwar nach vorwiegend lautlichen Gesichtspunkten zusammenstellte, wodurch sich häufig genug gekünstelte und mundartfremde Konstruktionen ergaben [(Wenker) assembled 40 sentences, which he arranged according to predominantly phonetic aspects, which often enough resulted in artificial constructions foreign to the dialect]” and opts for a more elaborate questionnaire and direct speaker elicitation as opposed to sending out questionnaires by mail. The idea of making audio recordings of Wenker sentences has been met with criticism also in the contemporary German dialect recording projects. Zwirner refrained from recording Wenker sentences because he wanted to reserve audio recordings only for investigations that could not be done without such recordings (Zwirner 1956: 9). Bellmann (1964: 73–74) considers the Wenker sentences methodically questionable and reports of the necessity to supplement the sentences and to adapt them to the respective linguistic variety. Riemann (1964: 85–86) mentions various pitfalls and problems in eliciting the sentences and speaks of a high number of erroneous data in the resulting recordings due to priming effects. Some harsh criticism of audio recordings of Wenker sentences is also found in Ruoff (1965), discussing problems posed by the sentences in the fieldwork and providing many examples of them producing bad data.

a novelty in those days) and trying to guess from factors such as a consultant's behaviour, clothing, or manner of speaking, what topics she or he might be inclined to talk about.¹² Therefore, the recordings often deal with aspects of the informants' daily lives (e.g. agriculture, crafts), local customs and folk culture, festivals or fairs, traditional narratives or the informants' memories of earlier times. Occasionally, lexical items are elicited in a conversational style (later recordings often conclude with elicitations of the days of the week, the names of months and seasons, or the numbers 1 to 20).¹³

The informants' social data appear to have been recorded prior to the audio recording sessions and are preserved in written form in the archival documentation (see Figure 2). It is currently unknown if the procedure of collecting the social data has generally never been taped or if this section of the sessions was edited out of the audio recordings later. Only in a few rare cases it is preserved at the beginning of a recording.

Interestingly, audio recordings of free speech supplementing the traditional field records (questionnaires, etc.) are introduced around the same time also in North American dialect surveys (e.g. *Survey of the German Spoken in Wisconsin* and *Linguistic Atlas of the Upper Midwest [LAUM]*, see Seifert 1949, 1951, Allen 1952 and Wilson 1956). Much as in the case of Hornung and Kranzmayer, there as well "a recording of the unrehearsed conversational speech" (Allen 1952: 93) was made where "the informant is questioned only enough to get him launched on some topic that strikes his fancy. The sample so obtained is used in the search for forms and structures not readily obtainable in the interview [...]" (Wilson 1956: 17, on *LAUM*) and to "provide accurate means of checking the fidelity of the field transcription of the various interviewers" (Allen 1952: 93). It is currently not known if there was a flow of ideas of some sort between Kranzmayer and the American fellow fieldworkers or if this is a mere coincidence. It might have been facilitated by the fact that after the end of World War II, inexpensive and light-weight electrical recording machines became available on the market, which better allowed for such a type of recordings.¹⁴ Hornung (1961: 185) attributes the idea of audio-taping spontaneous language without prior rehearsals to Kranzmayer and makes no reference to other dialect survey teams.

In Germany, dialect recordings on magnetic tape by the *Deutsches Spracharchiv* in Germany (since 2004 *Archiv für gesprochenes Deutsch*) began in 1955 under the direction of Eberhard Zwirner, however with somewhat different objectives ("Zwirner-Korpus", see e.g. Wagener and Bausch eds. 1997: 112–14, Stift and Schmidt 2014: 361). What Zwirner's approach has in common with Kranzmayer and Hornung's approach is that Zwirner also opted for recordings of informants speaking freely (e.g. in a conversation or narrative) rather than recording elicitations of word lists or the Wenker

¹² A brief statement about the division of work in the field between Kranzmayer and Hornung is found in Hornung (1968: 97 footnote 1, about their fieldwork in Burgenland 1951–1959), where Hornung says that during recording sessions Kranzmayer usually directed the protocolling (on paper) of the interviews while Hornung was the person at the microphone who did the interviews.

¹³ However, Kranzmayer and Hornung may well have also done questionnaire elicitations beside recording spontaneous speech that were not captured on tape. Ruoff (1997: 11) reports that in a fieldwork session at the local museum of Bludenz in 1958, Kranzmayer was taping dialect samples from informants (with Eugen Gabriel functioning as an interpreter) while Hornung elicited a word list in the antechamber.

¹⁴ Seifert, who did his fieldwork in 1945–1946, used a SoundScriber dictation machine (<https://en.wikipedia.org/wiki/SoundScriber>). According to Allen (1952: 93, 1977: 180), *LAUM*, where most interviews took place between 1949 and 1953, was the first American linguistic atlas project to systematically supplement information gathered in the field with on-the-spot recordings on magnetic tape of informants' free conversation for checking and obtaining additional grammatical data (in addition, some wire recordings had been made that were later rerecorded on reel-to-reel tape, see Allen 1977: 180, Linn 1993: 443). Occasional sound recordings (on aluminum discs) had already been made in the course of the *Linguistic Atlas of New England* between 1933 and 1939 (see <http://www.lap.uga.edu/Site/LANE.html>). After the 1950s, tape recording became frequently employed in American dialect atlases, see Allen (1977) and Linn (1993). For an early discussion of the pros and cons of tape recording in dialectology see McDavid (1957). An overview of the history and use of audio recordings in the *Dictionary of American Regional English* (DARE) and various American linguistic surveys and the approaches chosen there is given in Purnell (2013).

sentences. However, Zwirner did not strive to document the oldest accessible stratum of dialectal expression but rather the speech of different generations, recording (ideally in each location) one autochthonous speaker from the younger, the middle and the older generation (about 20 years, 40 and over 60 years, respectively). Additionally, in order to take into account the special demographic situation prevailing in Germany at the time, refugees or expellees that had come to settle in the respective place (see Zwirner and Bethge 1958 and http://agd.ids-mannheim.de/ZW--_extern.shtml) were recorded as well. For other early German corpora of dialect recordings see Wagener and Bausch (eds. 1997).


Bd. Nr. 303		Band Nr. 27	Bandgeschwindigkeit 19 mm/sec
		Aufnahme Nr. 9	Umdrehungszahl
Des Phonographierten		Der Aufnahme	
Vor- und Zuname <u>Thelma BINDER</u>		Datum, Ort, Land <u>Nöflach, 21.9.53</u>	
Geschlecht <u>fr</u> Volk, (Stamm, Muttersprache)		Sprache, Mundart <u>o. Gsoden, Steiermark</u>	
Alter <u>86</u> Beruf <u>Haushalt</u>		Gegenstand <u>Dienstbotenleben 2'45"</u>	
Geburtsort, Land <u>Gsoden</u>			
Aufgewachsen in <u>"</u>		Musik , vokal oder instrumental	
Längere Aufenthalte in		ein- oder mehrstimmig	
Städtiger Wohnort <u>Gsoden</u>		Stimmgattung oder Instrumente	
Heimat des Vaters _____ der Mutter <u>Wainach</u>		Technische Daten <u>Dyn. Mikrofon, Zehner-Vst., Magnetophon: Philips L. B. P. Band</u>	
		Aufgenommen durch <u>Doz. Ruth G. Hermann</u>	
		wissenschaftl. Kontrolle <u>Doz. Kretzenbacher</u>	
<p><i>hnd geön- davon gehen. Adä höchst für Dienst. " witeke der Witterer witekerin der Witterer</i></p> <p>303 Mauer Vorspann 2'45"</p>			
			

Figure 2. Archive protocol of recording B 303 (excerpt) (©Phonogrammarchiv)

Despite employing audio recording in their fieldwork, American dialect survey projects appear to have always relied on extensive questionnaires. Seifert (1949, 1951), who estimates the scope of his questionnaire at “better than a thousand items” (1951: 203), describes an approach and working procedure of enhancing questionnaire-based fieldwork with recordings of spontaneous, non-elicited speech that is also encountered in most subsequent survey projects:

[1951: 201–202] The survey of the German spoken in Wisconsin has proceeded according to the following plan. The first step is the collection of a body of linguistic data [...]. The second step consists of the analysis of the records. In the third step, conclusions are then drawn from the collected data.

[1949: 128–129] Two methods of recording were used — the one by hand in a rather close phonetic transcription, the other by machine with a small but good portable recording device.

The attempt was made to do both hand and machine recording with each informant [...] I used a questionnaire in which the items [deal] chiefly with situations out of everyday life [...] To go through the entire questionnaire takes from four to eight hours [...]. To check the over-all reliability of the responses, I tried to record with the machine a considerable piece of free conversation for each informant.

In contrast to Hornung's and Kranzmayer's approach, recording an audio sample of free speech always remained an appendix to questionnaire-based fieldwork (which was later also accompanied by audio taping) for the American survey teams, and they never employed it to replace traditional field records. Hornung and Kranzmayer seemingly never published any more detailed statements of their field methods. While there may be reasons to assume that, at least occasionally, they also elicited short wordlists or questionnaires (see above, footnote 13), and elicitations of selected terms are also found on their recordings, the sheer number of speakers that were often recorded on a single day makes regular hour-long elicitations beside these recordings unlikely. To our knowledge, Kranzmayer or Hornung never compiled (or published) a standardised wordlist or questionnaire of their own, yet they seem to have drawn from some informal catalogue of topics (mainly to do with the rural environment) in their fieldwork. The latter assumption is supported by a rare yet vague description of Hornung's methods and objectives, found in Hornung (1964a: 9), with respect to gathering lexical data on her trips to East Tyrol in 1957, 1958 and 1959:¹⁵

In erster Linie ging es mir darum, möglichst viel verkehrsfernes Wortgut des alten Bauernlebens ortsweise zu sammeln, das eine brauchbare Grundlage zur Erstellung einer Wortgeographie ergeben würde. Dabei wurden nach einem von E. Kranzmayer entworfenen, in der Wiener Wörterbuchkanzlei geübten System ganz bestimmte Wörter gefragt, die in Handel und Wandel keine Rolle spielen, Bauernwörter also, an denen sich die lautkundlichen und wortschatzmäßigen Differenzierungen kleinster Dialekträume auffinden lassen. [...] Es gelang mir, [...] das alte Arbeitsgerät in seinen örtlichen Sonderformen in bäuerlichen Höfen selbst [...] kennenzulernen und zu studieren. An Hand dieses Sachgutes konnte weiteres Wortmaterial, das bisher von der Dialektgeographie wenig oder gar nicht beachtet wurde, für die Wortgeographie verwertet werden.

[My primary concern was to collect as much as possible of the heritage vocabulary of the old peasant life, which would provide a useful basis for producing a word geography. Following a system designed by E[berhard] Kranzmayer and practiced at the Viennese Wörterbuchkanzlei, very specific words were elicited that do not play a role in trade and commerce, i.e. peasant words by means of which the phonetic/phonological and vocabulary-related differentiations of the smallest dialect areas can be detected. [...] I succeeded [...] in getting to know and studying the old working tools in their respective local forms [...] directly in the rural farms. On the basis of these items, further lexical material could be used for word geography that had hitherto received little or no attention at all in dialect geography.]

Other dialectologists, however, whose field recordings are likewise contained in our corpus, also used elaborate questionnaire materials ("Fragebücher [question books]") in their fieldwork and often employed a variety of questioning techniques to elicit the desired term (naming, completing, circumlocution, etc., see Chambers and Trudgill 1998: 22–25). For example, Eugen Gabriel in his

¹⁵ In this context see also Gabriel's (1972b) review of Hornung (1964), where he criticises (among other things) at some length the apparent lack of systematic lexicographical fieldwork and accurate data analysis.

work for the linguistic atlas of Vorarlberg and Liechtenstein pursued an approach similar to that of the American survey teams. He used an adapted version of the questionnaire of the Linguistic Atlas of German Switzerland¹⁶ for fine-grained elicitations (which took four to six days at seven hours of work per day to be completed, see Gabriel 1965: 60, similarly 1972a: 160) but considered complementary audio recordings of spontaneous speech indispensable e.g. for the study of syntactic phenomena and as a source of ancillary data (1965: 60–61).¹⁷

Gabriel's questionnaire for the Alemannic varieties of Austrian German was then in turn adapted to serve as the questionnaire for fieldwork on the Bavarian varieties of Austria (or linguistic enclaves in neighbouring countries) from the late 1960s onwards (in its latest incarnation as Patocka and Scheuringer 1988) and was used in research such as Schabus (1971), Geyer (1976), Tatzreiter (1978b), Lipold (1979), Bouterwek (1988) or Scheuringer (1982, 1990). Depending on the research objective, other questionnaires were also used. Marschall's (married Bigler-Marschall) (1975) work on the lexicon of the Viennese dialect employs a questionnaire of terms drawn from the works of the controversial Austrian writer Josef Weinheber. In his 1982 fieldwork, Herbert Tatzreiter used the questionnaire for the word atlas of continental Germanic wine-making terminology ("Questionnaire zum Wortatlas der kontinentalgermanischen Winzeterminologie"). Extensive discussion of field methods, the ethics of fieldwork and the design of the questionnaire is found e.g. in Scheuringer (1990 Chs. 1.9, 2.2, 2.4) and especially in Bouterwek (1988 Ch. 4). Works such as Scheuringer (1990) or Bouterwek (1988) also show a departure from Kranzmayer and Hornung's historical-etymological approach with its strong ties to folklore studies and evidence the growing influence of sociolinguistics and the social sciences on dialectal research as well as the incorporation of statistical methods.

It is often not fully clear yet to what extent, or how, the various fieldworkers evaluated and analysed the language data they collected on tape. Kranzmayer, in his writings, seemingly never referred to any particular recordings he made but only mentions the corpus itself on a few occasions (e.g. 1954, 1963¹⁸, 1965). Hornung published four short articles that are directly concerned with the recording enterprise as such or with particular recordings (1961, 1968, 1970, 1983), and Wilfried Schabus (1993) published the transcription of a recording from 1951, with extensive discussion. Otherwise, the corpus and its recordings are only referred to more or less in passing, e.g. in Hornung (1964a, 1972) or Hornung and Roitinger (2000) and Hornung and Grüner (2002), and in a handful of articles such as Hornung (1958, 1977, 1981, 1999). Tatzreiter (1965) reports on a field trip under Kranzmayer's supervision.

It can be assumed that data from the recordings also found their way into the card files of the *Wörterbuch der bairischen Mundarten in Österreich* [Dictionary of Bavarian dialects in Austria] or general articles such as Hornung (1960) as well as Hornung's dictionaries. In Hornung and Grüner (2002: 14, dictionary of the Viennese dialect) it is stated that the recordings helped in determining the representation of speech sounds but no particular recordings are mentioned. In her dictionary of the dialect of Pladen/Sappada, one of the field recordings is given as a reference for a certain phonetic

¹⁶ The Swiss questionnaire was itself inspired by the questionnaires of Gilliéron and Edmont's *Linguistic Atlas of France* and (Gilliéron's students') Jaberg and Jud's subsequent *Linguistic Atlas of Italy and Southern Switzerland*, both of which also served as blueprints for American dialect atlases such as the *Linguistic Atlas of New England* and the *Linguistic Atlas of the Upper Midwest*, for some discussion see Chambers and Trudgill (1998: Ch.2) and Hotzenköcherle (1939, 1962).

¹⁷ Gabriel (1965: 60, 1972a: 185) says he chose not to record complete elicitation sessions because meticulously transcribing them would have been too time-consuming. Another factor may have been the considerable costs of sufficient quantities of high-quality magnetic tape, since selectively transcribing longer recordings, or using recordings merely as a reference and for checking in cases of doubt, would certainly have been an option, too. An anonymous reviewer points out that recent or contemporary German dialectological research projects such as *Syntax hessischer Dialekte* (SyHD) or *Dialektatlas Mittleres Westdeutschland* (DMW) have also adopted the practice of documenting the whole elicitation procedure by means of accompanying audio recordings.

¹⁸ Hornung (1999: 89) attributes the authorship of Kranzmayer (1963) to herself.

feature (Hornung 1972: xvi) but no further details are available. Hornung (1964a: 10, on East Tyrol) concedes that a full analysis of the relevant tape recordings must be reserved for future studies, which seemingly never materialised.

A good number of recordings in the corpus were made by students or postdoctoral researchers as accompanying the fieldwork for their theses or *Habilitations* and were exploited in this context, mostly as a means for studying or documenting phonological or phonetic issues in the respective dialect or also folkloric issues, e.g. Tatzreiter (1964, 1978b), Bauer (1967), Geyer (1976), Lipold (1979). Sometimes, the recordings provided a corpus for studying occurrence frequencies of linguistic variables, e.g. Scheuringer (1990). Tatzreiter (1964) and Lipold (1979), also present some recorded narratives together with transcriptions and translations as a demonstration of the syntax in continuous speech. Bouterwek (1988: 13, 630) regrets that despite their being an integral part of his research, his sound recordings could not be analysed within the limits of his thesis.

On the whole, it appears that these authors' findings are primarily based on traditional elicitations or questionnaire work and that the recorded audio materials played a minor role here as well (see e.g. Scheuringer 1982: 13). Moreover, in-depth research on syntax was clearly not a focus in their research. In contrast, Franz Patocka (1997) took a different approach and compiled a corpus of nearly 6000 sentences from the *Phonogrammarchiv's* dialect recordings from 1952 to 1987 (including his own recordings), on which he based his study of the syntax of spontaneous dialectal speech.

Eugen Gabriel's field recordings constitute another exception. They were likewise made in addition to interviews involving an extensive questionnaire (see above), but nevertheless played an important role in researching and documenting the German varieties of Vorarlberg and Liechtenstein. In his cooperation with the *Deutsches Spracharchiv, Tübinger Arbeitsstelle*, Gabriel's recordings from Vorarlberg and Liechtenstein were transcribed and analysed (see e.g. Gabriel 1972a, Ruoff 1973) and were partly also published (Ruoff 1985, 1993).¹⁹ From 1998 on, selected excerpts from recordings representing all dialectal varieties of Vorarlberg have been published in the series *Die Mundarten Vorarlbergs* [The Dialects of Vorarlberg] (Ruoff and Gabriel 1998–2004, Berchtold and Schallert 2013–).

In order to obtain a better knowledge in which way different dialectological researchers made use of their field recordings and what impact they had on their findings, we plan to conduct interviews with available researchers.

However, the fact that field data were used in one way or another in a thesis or publication does not necessarily mean that the corresponding recordings are preserved. For example, in his dissertation on the dialect variety of Pichl bei Wels (Upper Austria), Alois Brandstetter (1961: 59–60) mentions that he had also made phonographic recordings and presents a narrative in phonetic transcription together with a translation into Standard German, yet the recordings have never been part of the *Phonogrammarchiv's* holdings (and up to now they have not surfaced elsewhere), whereas his recordings from Lower Austria from later in the same year are available. Similarly, Hornung (1970) published a partial transcription with a commentary of a recording made in Vienna (introduced as "Probe [test]"). A tape reel with a label referring to that recording was found in her estate but it contains a different (hitherto unknown) recording and has audibly been taped over several times.

¹⁹ Gabriel's recordings from the Winter of 1964/65 were archived not only in the *Phonogrammarchiv* but also in the *Deutsches Spracharchiv, Arbeitsstelle Tübingen*, see Gabriel (1972a: 190–191) and Schallert (2014: 132) for a concordance of the archive numbers in either institution. According to Ruoff (1993: 24), the *Bregenzer Landesbibliothek* also hosts copies of the recordings (besides copies of all other recordings of the *Deutsches Spracharchiv* relevant to Vorarlberg). For more on the *Tonarchiv der Mundarten Vorarlbergs* (sound archive of Vorarlbergian dialects) see Schallert (2014: 128–136). Schallert (2014: 133 footnote 8) states that the *Bregenzer Landesbibliothek* now hosts transcripts of all recordings as well. Some dialect points covered by Gabriel have also been covered by the *Deutsches Spracharchiv*, see Schallert (2014: 139) for an overview.

Another recording treated in the same article (likewise called “test”) is archived in the *Phonogrammarchiv*. Apparently, one cannot exclude that recorded tapes were not always returned to the *Phonogrammarchiv* after a field trip was over, to the effect that the tapes are now lost, or that fieldworkers sometimes deleted recordings after analysing the data and reused the tape.

Additionally, Tatzreiter presents three narratives transcribed from tapes in his dissertation (1964: 40–46) that are not among the *Phonogrammarchiv*’s holdings. However, since Tatzreiter (1964) does not mention the *Phonogrammarchiv* and Hornung (1972: viii) states that he had received training in dialectological fieldwork from her, there is still a slim chance that these tapes might be found among the hitherto unidentified tapes in Hornung’s estate.

In 1974, Kranzmayer was granted two Austrian Science Fund (FWF) projects concerned with dialect tape recordings, but it is currently unknown to us what became of these projects and what results they achieved.²⁰ We had already found indications that transcripts may exist of a number of recordings, and a subsequently initiated search among papers from Hornung’s estate, currently stored at the department “Variation and Change of German in Austria”, indeed retrieved c. 1200 sheets of handwritten transcripts in various stages of completion (beside working notes and other materials), created between the years of 1970 and 1974. One part of these transcripts corresponds to archived recordings, whereas the other part represents transcriptions of hitherto unknown recordings the whereabouts of which are not known. We assume that these transcripts were at least partly produced in the course of Kranzmayer’s FWF projects and hope that further inquiries will help to retrieve the missing recordings or shed some light on their fate. The 1200 sheets of handwritten transcripts have meanwhile been scanned and in part made machine-readable by means of the *Transkribus* software. While the *Phonogrammarchiv*’s archival documentation contains all social data and other metadata pertaining to the recordings, it contains only few (full or partial) transcripts.

2.3 Assembling and partitioning the corpus

Due to their heterogenous nature, the recordings featuring dialectal or other varieties of German in the *Phonogrammarchiv*’s collections do not form a coherent corpus that can be easily delimited. Dialectal German is found not only on recordings featuring spontaneous speech or elicitations that were explicitly made for linguistic purposes but also on recordings of interviews made for the purpose of folkloristic documentation, and, in addition, on recordings of e.g. poetry recitals, folk songs or folk plays. Furthermore, the recordings are geographically distributed over several (historical as well as present-day) countries. In order to distil from all these recordings a workable corpus of spoken German varieties, we had to make a number of choices regarding the inclusion or exclusion of recordings, the partitioning of the material in the workflow, and also aspects of archiving and digitisation.

Firstly, recordings of dialect poetry readings, folk songs or folk plays have not been included because of their non-spontaneous nature and in order to avoid potential legal issues in prospective later dissemination. Secondly, for the time being it was decided to limit the corpus largely to a selection of recordings from Austria and northern Italy (South Tyrol, Venetia including the Seven Communities and the Thirteen Communities, Friuli Venezia Giulia) from the 1950s up to the early-to-mid 1980s. These recordings also constitute the bulk of the recorded material. In order to keep the corpus manageable within the project’s resources, we did not include linguistic atlas materials from the 1990s and early 2000s as well as certain further recording series from Austria or other linguistic enclaves.

²⁰ “Auswertung des Tonbandarchivs der österreichischen Mundarten [Evaluation of the tape archive of Austrian dialects]” (project no. 2017, funded with ATS 120,000) and “Auswertung des mundartlichen Tonbandarchivs [Evaluation of the dialectal tape archive]” (project no. 2467, funded with ATS 300,000), see the FWF Annual Report for 1974. Since Kranzmayer died in September 1975, it is presently unclear if he even lived to witness the projects’ completion.

However, for achieving proper digitisation and archiving results, we decided to always digitise complete tapes and to archive complete recording series instead of leaving out portions that are not immediately relevant to the corpus as defined.²¹ Consequently, not all recordings in the corpus represent only spontaneous speech (or linguistic elicitations), and some recordings represent varieties of German spoken in countries such as Hungary or today's Slovenia. When the occasion presented itself, fieldworkers of course also recorded folk songs or recitations of dialect poems, or, for example when doing fieldwork in northern Italy, grasped the opportunity to cross the border to former Yugoslavia to do additional fieldwork there. Moreover, recordings made in plurilingual areas of Austria occasionally also feature other local languages beside German, e.g. Hungarian, Croatian or Romani in Burgenland, or Slovenian in Carinthia. Recording series exclusively concerned with the documentation of Hungarian, Croatian or Slovenian in these provinces, however, are not included in our corpus.

The resulting corpus comprises 2442 audio recordings (c. 539 hours) on magnetic tape (alongside associated archival materials) from the 1950s up to the 1980s (and additionally a few recordings from the 1990s, see below) documenting (mostly dialectal) linguistic varieties of German in Austria and additionally in Northern Italy, in Hungary, in former Yugoslavia and in former Czechoslovakia.

In total, samples of dialectal varieties from 1036 places were collected and 2550 speakers were recorded.²² In numerous cases, thus, two or more speakers were recorded for a particular local variety. Likewise, recordings of bilingual speakers from plurilingual areas may feature two local languages spoken by the same speaker (e.g. German and Croatian, Hungarian or Romani in Burgenland, or German and Slovenian in Carinthia). Exceptionally rich documentation is available for the province of Burgenland (730 recordings)²³, whereas comparatively few (37) recordings were made of varieties of Vienna. To counterbalance this disproportion, Wilfried Schabus' recordings of older Viennese citizens from the 1990s have also been added to the corpus. The recordings represent the first attempt at a comprehensive nation-wide audio documentation of linguistic varieties in Austria. They provide an exceptional picture of Austria's dialect landscape after the middle of the 20th century across all regions and constitute a source of historical dialect audio data that is unique in the German-speaking area and is a valuable primary source for various areas of research.

For administrative and workability reasons, work on the corpus was partitioned into (at the time of writing) two stages. The first stage in 2019 was dedicated to digitising and archiving a subcorpus of 1768 dialect recordings from 1951 to 1983, a great number of which were also included in UNESCO's national register "Memory of Austria" in 2018. The remainder of the corpus (674 recordings) is dealt with in the second stage, which started in 2020. Given the funding, a potential third stage will be concerned with further recordings of German varieties spoken in Austria and

²¹ In general, a recording series coincides with the respective field trip in the course of which the recordings were made. However, this practice has not always been strictly followed in the *Phonogrammarchiv's* archival documentation, so that occasionally two unrelated field trips by different researchers were collapsed into a single recording series, or a field trip was split into two or more recording series. Due to the long-time practice of copying original field recordings onto archival tapes for long-term preservation and storage, an archival tape reel may also contain recordings of totally unrelated field research projects (e.g. the end of one and the beginning of another recording project). In such cases, we also digitised the complete tape reel but the additional recordings are of course not part of our corpus.

²² These and other figures may still be subject to (slight) change as data are checked for accuracy and completeness.

²³ The fieldwork in Burgenland 1952–1959 was funded by the provincial government of Burgenland. On behalf of then-provincial governor Lorenz Karall, all municipalities and all regional languages of Burgenland were covered following a presentation of the recordings from the first field trip to Burgenland in 1952 (see Hornung 1968: 97, 1999: 89), and informants were often found with the help of the provincial and municipal administrations. The latter strategy was also pursued in the Vorarlbergian enterprise of documenting the dialectal varieties there (likewise employing audio documentation), see Schwarz (1957: 114).

neighbouring countries (Hungary, Slovenia, Czechia and Germany [Bohemian Forest], Slovakia and Romania) as well as newly discovered relevant recordings from Hornung's estate. Unfortunately, the Covid-19 pandemic has severely affected the progress of the project (as it is undoubtedly the case with countless other projects) and it is not yet clear which adaptations will have to be made.

2.4 Options of analysis and data utilisation

Hornung, Kranzmayer and their fellow researchers deviated from the earlier dialectological fieldwork or recording practice, which focussed on questionnaire elicitation such as the 40 Wenker sentences, and during which even recordings of purported spontaneous speech had, in fact, often been rehearsed many times before the actual recording session took place. Their recordings may thus be less suitable for quickly filling gaps in a dialect atlas but in return exhibit structures and phenomena that are not readily obtainable in an interview and cannot be gleaned from questionnaire data. In this way, they also preserve dialectal features that cannot be captured in elicitations and may have vanished for a long time. Typically featuring long passages of spontaneous continuous speech, the recordings lend themselves, e.g., to studies pertaining to prosody, information structure and discourse prominence-related issues, pronominal reference, narrative style and other phenomena that can fully be studied only in larger non-preconstructed contexts. Their oftentimes astonishingly good sound quality makes them suitable also for phonological and phonetic investigations.

Spanning almost five decades, the corpus likewise enables us to trace linguistic change in the recorded varieties. The large variety of covered places and recorded speakers together with the considerable chronological depth allows for multi-dimensional sociolinguistic and fine-grained geolinguistic analyses. The fact that 10 to 20 years after the initial fieldwork of the 1950s and 1960s, most areas were covered again in the 1970s and 1980s, sometimes even featuring the same speakers, makes it possible to conduct inter- as well as intra-speaker real time analyses.

Since the speakers often talked about their immediate living environment and aspects of their culture, or offered specimens of their narrative traditions, the corpus is also of great socio-historical and historic-cultural importance. Thus, it not only provides data for linguistic research in the narrower sense, but, in addition, offers materials relevant to disciplines such as social and cultural anthropology, folklore and folklife studies, research in oral history and oral traditions, or ethnomusicology. For these reasons, it is planned to make the recordings accessible via a multidisciplinary platform that serves scholarly purposes across disciplines as well as interests of the non-academic public. For example, by way of the subject entries as well as the descriptive timeline of a recording's contents, both taken from original archive protocols and synchronised with the digital audio files, the recordings will be searchable, e.g. for their cultural or folkloristic content or other content-related features. In addition, machine-readable (tagged, glossed) transcriptions plus translations of large portions (ideally, of all recordings), will make the recordings exploitable for corpus linguistic tasks, and will also serve as the basis for a comparative dictionary and the depiction of lexical variation. Precise geodata of all dialect points support studies on areal variation. In this way we hope to make the corpus usable and attractive to a broader academic as well as non-academic audience.

3 Preparing the corpus

In this section we address digitisation and metadata-related issues and procedures in preparing the corpus for further use, and discuss problems encountered therein.

3.1 Digitisation

Out of the presently envisaged 2442 recordings (c. 539 hours) in the corpus, only 1219 recordings (c. 130 hours) had already been digitised when the project started in January 2019.²⁴ With the exception of a few DAT tapes, the remaining recordings consisted of c. 400 reels of magnetic audio tape. Despite the pandemic, their digitisation could be completed in 2020. Following their digitisation to 24 bit/96 kHz WAV files, the digital copies of these tapes were segmented so that each recording is available as a(n accordingly labelled) separate file. Most recordings are mono recordings. We also discovered that among the previously digitised materials, a considerable number of digital copies of archive tapes had not been segmented, or only incompletely so, and other archive tapes had been digitised only partially. We therefore had to include the completion of these tasks in our workflow. Also, the recordings from the 1990s on digital audio tape (DAT) cassettes have meanwhile been transferred and segmented.

3.2 Tape generations

In digitising the audio tapes, we are concerned with first-generation (i.e. original) and second-generation tapes (i.e. analogue copies), which has to do with the history of analogue audio archiving in the *Phonogrammarchiv* in the analogue era. In the years 1951–1958, when the *Phonogrammarchiv* possessed only a single tape machine (Lechleitner 1999), the original tapes recorded in the field were included in the *Phonogrammarchiv*'s holdings after they were cut and interspersed with sections of leader tape to separate individual recordings from one another. The digital copies of these tapes are therefore sourced from the first tape generation. It cannot be excluded that in the course of archiving, parts of recordings that were deemed unfit or not worthy of being archived were cut away and are now lost.

Later, when battery-powered, portable tape recorders could also be acquired and were frequently lent to field researchers, there was a policy of copying the resulting recordings onto so-called archive tapes, and only the latter were fully included in the *Phonogrammarchiv*'s holdings (in the mid-1970s, leader tape between recordings was also replaced by reference tones). This was done to ensure that the contents of the original field recordings were stored on high-quality magnetic tape kept under proper storage conditions for long-term preservation. These archive copies were given priority over the original tapes, and all archival recording-specific documentation refers to them. Digital copies of these archive tapes are therefore sourced from the second tape generation only. Unfortunately, the *Phonogrammarchiv* never kept records of the original tapes so that it is often unclear nowadays what

²⁴ The figure of 2442 refers to the total of archive numbers that comprise our corpus but does not actually reflect distinct, individual recordings in all cases. Likewise, the figure of 1219 or other figures pertaining to recordings refer to archive numbers only. In some cases, recordings on archive tapes with different archive numbers have turned out to be adjacent, overlapping parts of one continuous original recording. In some other cases, (parts of) potentially different recordings evidently had been joined together on the archive tape and filed under the same archive number (see section 3.2 below for original tapes and analogue copies on archive tapes). For the most part we can only speculate why things had been handled this or that way by the respective archivists in the past. In a few cases an original recording had to be split over two archive tapes and the continuation was not identified as such but given a new archive number. In yet other cases, seemingly a new archive number had been introduced (with or without splitting the original recording when copying it onto the archive tape) when a change of topic occurred or another speaker became dominant, or (parts of) recordings had been joined together to unite parts featuring the same speaker or topic under the same archive number. Sometimes perhaps also a part (or parts) of an original recording not deemed fit to be archived had been edited out in the archiving process, leaving discontinuous pieces of the original recording. Where possible and reasonable, we restored the original shape of split recordings following their digitisation. For reasons of consistency, we did not change archive numbers. WAV files resulting from restoring the original shape of a recording involving two or three archive numbers were labeled accordingly by providing all the respective archive numbers in the file name.

happened to them. Some are known to have been reused or returned to the fieldworkers whereas others remained in the *Phonogrammarchiv* but were not catalogued.²⁵

However, since analogue tape copies suffer a loss in audio fidelity, it is of course preferable to digitise the original tapes instead of the 2nd generation archive tapes if the original tapes can be retrieved and properly played. Of some dialect recordings, the original tapes were discovered in the *Phonogrammarchiv*. Unfortunately, their actual contents are often not clear because the tape boxes are only sparsely labelled and no documentation is available as to what recordings are found on what tape reel, and their condition and playability is yet to be determined. Among those original tapes that have so far been digitised, most were still perfectly playable. Only five tapes with recordings from 1991 could no longer be played (we do not know, however, how long these tapes had already been in use when the 1991 recordings were made). Also, it is not known if the recordings were always archived according to their order on the original tapes, and, again, it cannot be excluded that certain parts were edited out when copying the originals onto the archive tapes. From all this it becomes clear that in order to use the original tapes, the archive tapes must be digitised so as to enable us to identify matching recordings on the original tapes. We have decided to fall back on original tapes only to the extent feasible within the project's resources.

After Hornung's passing in 2010, some 40 tapes from her estate that had been found in her flat were handed over to the *Phonogrammarchiv*. Again, no documentation is available as to their precise contents and labelling is sparse, but cursory checks have shown that some tapes contain the original versions of recordings found on 2nd generation archive tapes, and the better part of them has already been digitised. In some cases, the tape boxes were mislabelled and the tapes contained new dialect recordings that had not previously been known to exist. We plan to digitise the remainder of these tapes as well.

A similar state of affairs must be expected also regarding other parts of the *Phonogrammarchiv's* tape holdings. Therefore, we plan to explore the possibilities of audio fingerprinting for identifying original versions of archived recordings sourced from second-generation tapes on digital copies of original tapes for which no documentation is available.

3.3 Granularisation, enrichment and re-assigning of metadata

Ruoff (1973: 26) remarked that “Selbst die überaus gewissenhaften und erstaunlich umfassenden Protokolle zu den Sprachaufnahmen des Wiener Phonogrammarchivs haben deren spätere sprachwissenschaftliche Auswertung nicht gefördert [Even the extremely diligent and astonishingly comprehensive protocols on the language recordings of the Viennese Phonogrammarchiv did not promote their later linguistic evaluation].” In order to encourage any further evaluation or analysis of the recordings, these protocols must be transferred into a well-structured and searchable electronic form of documentation.

In the original historical archival documentation (data sheets on paper, so-called protocols) there is a separate data sheet (for a long time handwritten, later typewritten) for each recording, which contains the metadata pertaining to that recording (for an example from 1953 see Figure 2 above). Metadata include i.a. the archive signature (the recording's catalogue number), the date and place of the recording, its duration, the recordees' names and social data, the involved fieldworker(s), recorded languages/varieties or musical forms, topics and other content-related indications, a time protocol, and

²⁵ For example, according to Wagener and Bausch (eds. 1997: 221), the original tapes of Hermann Scheuringer's recordings for the linguistic atlas of Upper Austria (*Sprachatlas von Oberösterreich*) are housed by the Adalbert-Stifter-Institut des Landes Oberösterreich (see also the web page of the Stifter-Haus, <https://stifterhaus.at/index.php?id=20>). However, no mention of this fact is found in the documentation accompanying the archival copies of these tapes in the *Phonogrammarchiv*.

technical metadata (e.g. technical equipment involved, original and archive format, analogue-to-analogue or analogue-to-digital transfer procedures), and so forth.²⁶

When the *Phonogrammarchiv* introduced the electronic documentation of recordings by means of a database around 1990 there were already tens of thousands of recordings with an analogue archival documentation. In order to have all recordings represented in the database quickly, in most, if not all, cases only some basic metadata had been entered. Therefore, an important task in our project is to enrich the electronic metadata pertaining to our corpus on the basis of the available analogue documentary materials, and, on this occasion, also to correct possible errors.

However, when switching to electronic documentation, also the fatal decision had been made to set up the database in such a way that it does not document individual recordings but only bundles of recordings: the metadata of the individual recordings made by a fieldworker on the same day were collapsed and lumped together into a single general bundled entry composed of the metadata of all recordings in the bundle, thereby dissociating the metadata from the actual recordings to which they pertain. As schematically illustrated in Figure 3, in bundle entries the metadata are no longer associated with individual recordings but only with the bundle as such.

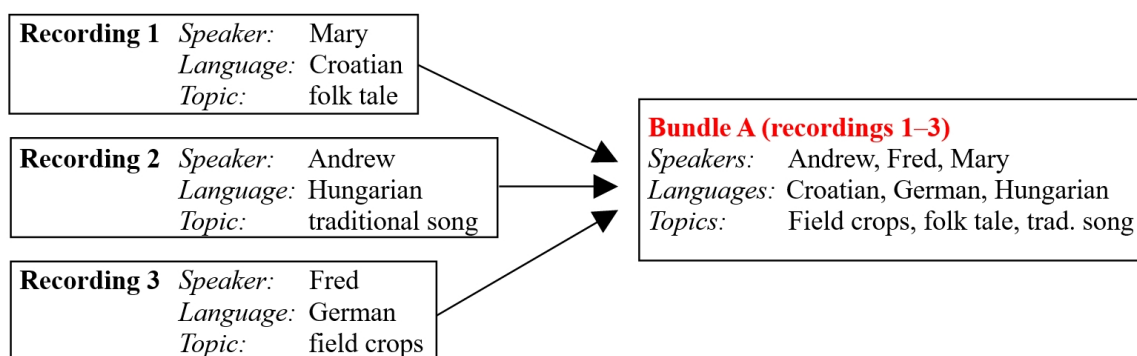


Figure 3. Lumping together metadata in a bundle entry

As a consequence, a search for recordings according to a particular search criterion does not return individual recordings in the search results but only bundles of recordings that contain one or more recordings to which the search criterion applies. In addition, the search results cannot specify which recordings exactly these are. The search may thus return also quite a number of recordings to which the search criterion does not apply. In Figure 3, for example, a search for *Croatian* will not only return Recording 1 but also Recordings 2 and 3 because they are included in the bundle, despite the fact that they do not contain *Croatian*. Similarly, a combined search for *Croatian* and *Traditional Song*, e.g. when searching for traditional songs in Croatian, will return Bundle A with its three recordings because the bundle contains the metadata entries *Croatian* as well as *Traditional Song*, although not a single recording in Bundle A features a traditional song in Croatian since each of the two criteria applies to a different recording in the bundle. Unambiguous and accurate search results are only possible in cases where a bundle contains just a single recording, as in Figure 4.

Since in the case of bundles that contain more than one recording, the search criterion may apply to minimally one and maximally all recordings (and any number in between) contained in a given bundle in the search results, the original protocols on paper must be consulted to determine the precise recording(s) to which the search criterion applies. Thus, it is evident that a huge number of recordings cannot be unambiguously found by a search in the database, and conversely that the

²⁶ Interestingly, the protocols often make no explicit mention of who the actual investigator in the field was.

database often returns search results that do not conform to the search criteria as far as recordings are concerned.

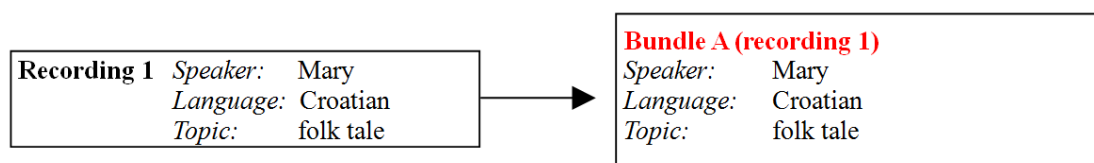


Figure 4. Single recording bundle

In our corpus, roughly 50% of the recordings are included in such metadata bundles. Since sometimes up to 20 speakers (each representing the local dialect of a different village) were recorded on a single day in the field, we are faced with a number of very complex bundles. In order to make the electronic documentation usable for any search-related purposes and corpus exploitation tasks, it was therefore necessary to granularise all metadata bundles and re-associate all pieces of metadata with those individual recordings to which they actually pertain. Since the problem of metadata bundles is not restricted to our corpus but extends across the *Phonogrammarchiv's* database, we decided that the procedures to achieve this must be applicable to the database in general.

First, we created an excerpt of the *Phonogrammarchiv's* database that contains only the data sets that are relevant to our corpus. Later, these data sets will be re-transferred and will replace the original entries. This was done in view of the risk of inadvertently compromising the complete database by applying experimental procedures to the data in our corpus. In the next step we granularised all bundle entries that were composed of the metadata of several (i.e. more than one) recordings into as many single-recording bundles as there were recordings in the bundle, together with extending the bundle signature by a delimiter followed by internal consecutive numbering (see Figure 5).

<i>multi-recording bundle signature</i>	<i>single-recording bundle signatures</i>	<i>(original recordings)</i>
19520919.N001 (= B 181–B 197)	=> 19520919.N001#001 19520919.N001#002 19520919.N001#003 <i>(etc.)</i>	(= B 181) (= B 182) (= B 183)

Figure 5. Granularisation of multi-recording bundles

Thus, a bundle entry comprising the metadata of 17 different recordings such as *19520919.N001* in Figure 5, which is made up of the metadata of the recordings B 181 to B 197, is granularised into 17 single-recording bundles (initially, empty copies of the original bundle), each of which in the end contains the metadata of just the respective single recording. This was done in order not to alter the basic architecture of the database and to enable us later to re-transfer the granularised entries. With the help of a matrix tool, each piece of metadata from the original bundle entry is now assigned to the single-recording bundle to which it pertains, as schematically shown in Figure 6.

Since in the original multi-recording entries all links between the metadata and the respective recordings were lost, this reassignment of metadata has to be done manually by falling back on the original hand- or typewritten documentation. By granularising a bundle of many recordings into many bundles of single recordings, the existing architecture of the database is left intact but it becomes possible for a search to find all and only the recordings to which a search criterion applies while excluding recordings to which the search criterion does not apply.

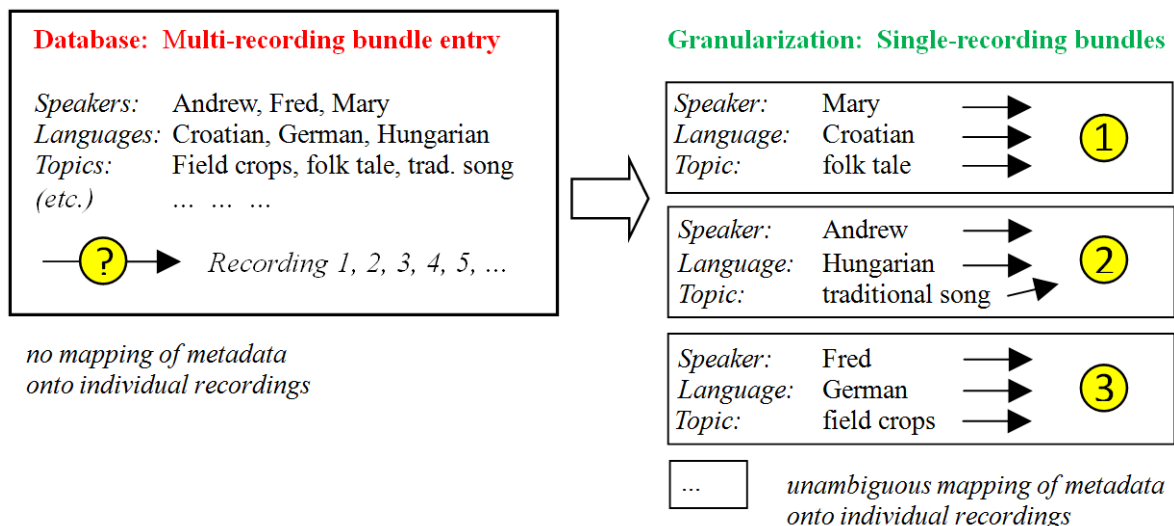


Figure 6. Granularisation, (re-)assignment of metadata

In the course of enriching the metadata, we also add the geodata of the villages or towns whose dialectal varieties are represented in the recordings, or where recording sessions took place, with the prospect of creating an interactive map of the documented dialect points in Austria and the adjacent regions. Since, more often than not, the place where a recording was made is not the town or village where the sampled variety is spoken, a feasible solution had to be found to include the relevant geodata in the dataset, as the database originally allowed only for entering the geodata of the place of recording. We offer a discussion of these issues in the next section.

4 The corpus – more details, some problems

In this section, we present a few more details on the composition of the present corpus and discuss our approach to certain issues arising in properly representing it in our electronic documentation. First, we present some statistical data, then turn to problems in including geographical metadata in various places in our database, and finally we also address some ethical and legal questions relating to the recordings in our corpus.

4.1 Some descriptive statistics

As mentioned above, the corpus in its present state comprises samples of dialectal varieties from 2550 recorded speakers and 944 places in Austria as well as 92 places in adjacent German-speaking areas (also see footnote 22), with a total recording volume of 539 hours. The exact location of these places is shown in Figure 7 below (for the over-representation of Burgenland see footnote 23 above), whereas Figure 8 shows the amount of audio recordings in hours per year and region for the nine Austrian provinces and South Tyrol/North Italy, from where the great majority of recordings in the corpus originate.

We can now also make accurate statements about proportions of recorded speakers within the corpus. The gender split is fairly even, at around 64% male and 36% female. Sadly, the original protocols remain largely silent about the educational history of the speakers. Out of the 2550 speakers in the corpus, for 1982 speakers no explicit information of this kind is provided. Especially during the earlier years of the fieldwork, such information was not systematically collected. Rather, educational data seem to have been recorded only in those cases where a speaker clearly did not fulfil the profile

of an uneducated, rural speaker of a local dialect but had reached a higher level of education (also compare section 2.2 above and footnotes 8 and 9). This gap was later rectified, so that later recording series often specify such information for all speakers, or provide more detail.

Among those speakers whose educational level is recorded in the documentation, 76% are specified as having undergone no more than the mandatory minimum schooling (*Volksschule* [primary school]). The rest have undergone also secondary schooling up to the *Matura* (maturity exam) level or attended a variety of vocational schools (20%), or received education at a university (4%). The last figure is partially explained by a small number of people who were recruited as speakers of their home dialect while they attended the university in Vienna, and whose interviews were also recorded there.

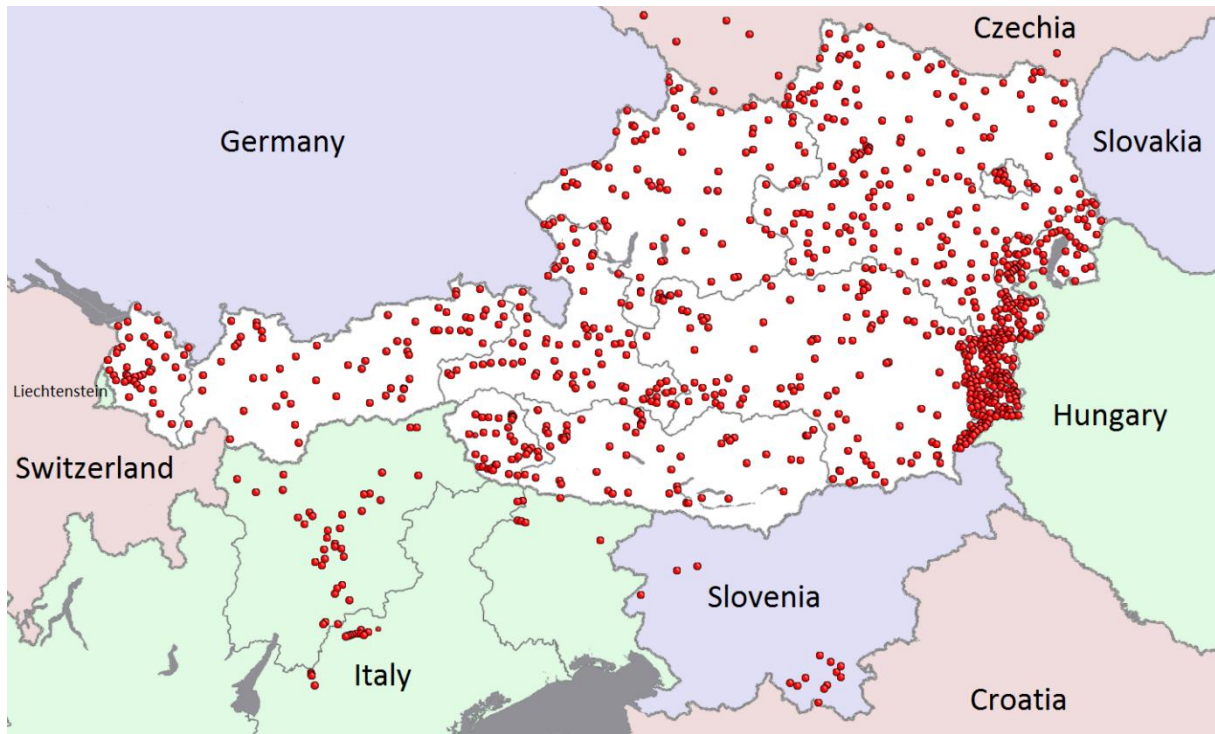


Figure 7. Locations of documented dialect points (audio recordings)
(© OpenStreetMap contributors)

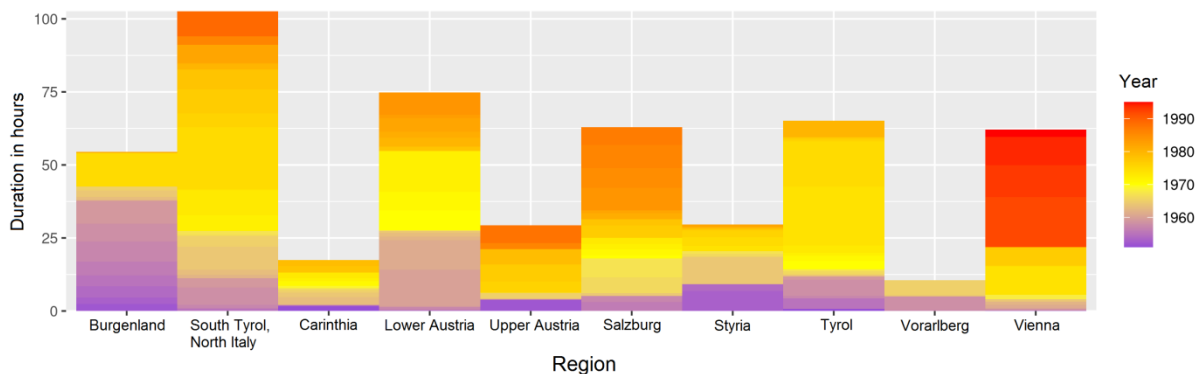


Figure 8. Recordings (in hours) per year and region

Due to the objective of capturing the oldest accessible dialectal layers (especially in Hornung and Kranzmayer's fieldwork, see Hornung 1961: 186), the speakers were also largely at an advanced age at the time of recording. In

Figure 9 below, we provide a graph of the birth years of our recorded speakers.

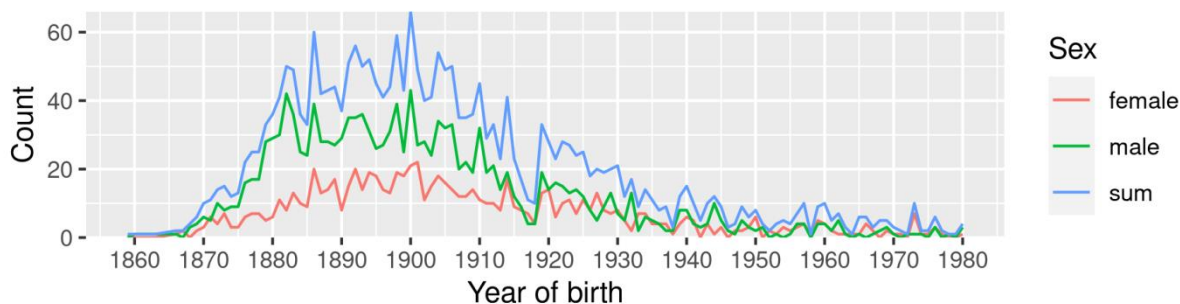


Figure 9. Number of recorded speakers and their years of birth

From the graph, one can observe that the vast majority of speakers in our corpus is very unlikely to be still alive today, which also has (legal) implications for our plans to make the corpus publicly accessible. Yet it might still be possible to conduct follow-up studies involving some of the informants from field trips in later years.

However, until we were able to present corpus-related data in such a form, we had to cope with a number of problems regarding the representations of specific types of metadata in the database. In the next section, we will discuss some issues with the representation of geographical data, which constitutes a crucial factor in a corpus documenting areal linguistic variation.

4.2 Some issues concerning geographical metadata

While the project was originally intended to make use only of the existing capabilities of the *Phonogrammarchiv's* database, we soon noticed multiple mismatches between the extant technical infrastructure and standards of description and the desiderata of a digital corpus that must allow for a machine-aided analysis of its contents.

A crucial issue was the manner in which geographical information is stored in the database (i.e. information relating to towns or villages, municipalities, regions, etc.). In the *Phonogrammarchiv's* database, geographical references are stored independently of each other in the metadata set of each individual recording (or bundle of recordings), or in the metadata set associated with each informant. Moreover, the *Phonogrammarchiv's* database supports hierarchically organised geographical information only with respect to the place where a recording was made but not in the case of other geographical indications, such as a speaker's birthplace or place of residence, or, very often, the place of provenance of the sampled dialect (recall from above that many recordings were not made in the place the dialect of which they represent but somewhere else). In these latter cases, geographical information is entered only in free text fields, that is, fields where information is given as arbitrary text with no restrictions on form or content, and without any relational structure. This mode of representation had initially been adopted by the *Phonogrammarchiv* for its electronic documentation because it allowed to accurately replicate all respective information as it is given in the original archive protocols, and also provided a quick and uncomplicated way to enter such data. However, the (undesirable) consequence of this approach is that in many cases, it is not readily possible to determine in a reliable way which entries refer to the same location and which ones do not, and what their other relations might be, or to use supplemental geographical indications to perform searches in a systematic and exhaustive way.

For example, since out of the 17271²⁷ towns in Austria, 36% do not carry a unique name, and reversely, 14.5% of the names of towns (or villages) in Austria are used by more than one town, it is not possible to unambiguously identify a town by its name alone. While many towns can be disambiguated by their epithets (e.g. *Bruck an der Mur* vs. *Bruck an der Leitha*) and others by adding to their entries disambiguating information such as the municipality and/or district within which they are situated, even this is still not sufficient in certain cases to disambiguate them without a doubt and to identify unique locations.

Another issue was a mismatch between the way geographical data were recorded in the original archive protocols (or in the free text fields) and a representation that is useful for future analysis. Fortunately, we found that on the whole, the data have been entered diligently and in great detail. However, since the original documentation as well as the entries in the free text fields in the database were created by dozens of people over several decades in a variety of settings, the data are formally inconsistent. Names of towns could be written in one of multiple alternative forms (e.g. *Steinbrunn* or [formerly] *Stinkenbrunn*), with or without an epithet (*Bruck an der Mur* vs. solely *Bruck*) or by using various ways of abbreviating the epithet (e.g. *Bruck an der Mur* vs. *Bruck a.d. Mur* vs. *Bruck/Mur*). Likewise, an entry may include or omit disambiguating additions such as a reference to the province or cultural region in which the respective place lies. In plurilingual areas, the name of a place could also be given in one of several languages (e.g. German *Großwarasdorf* vs. Croatian *Veliki Borištof*, situated in the province of Burgenland). On top of that, regions are more likely to be provided in the form of colloquial, imprecise designations (e.g. “Bucklige Welt” or “Salzkammergut”) instead of well-defined official administrative terms.

Of course, these imprecisions mean that identification sometimes requires interpretation, and they have previously led to confusion.²⁸ The non-Austrian locations of German dialect points, for example, are mostly given by their German names (e.g. German *Pladen* vs. Italian *Sappada* in northern Italy), and at times these are names that are no longer in use today and thus are not easy to track down. Certain small hamlets, such as *Boskangruabe*, cannot be found by any search engine under their German or local names. In the case of *Boskangruabe*, which belongs to the community of Selva di Progno con Giazza in northern Italy (one of the Cimbrian German-speaking *Thirteen Communities*), the precise location could only be determined by following the road leading from Giazza/Ljetzan northwards in *Google Maps Street View* until the village boundary signs reading *Boscangrove*, *Boskangruabe* and *Boscangrobe* were found (see below Figure 10). So far, the only hitherto unsolvable cases in our corpus are posed by a handful of small Bohemian towns with, within their region, non-unique names, whose dialects are represented by speakers who had moved over larger distances into Austria so that not even an educated guess based on proximity or other supplemental information is possible.

For all these reasons we decided to change the way geographical information is stored in the database, and to append geographic coordinates to all our place names. However, a workflow of manually managing dozens of such entries of plain text per individual recording (in addition to the danger of introducing errors in the course) was judged unreasonable. We therefore chose to do away with specifying geographical information in free text fields and instead to move to a modern relational structure, using a controlled list of places and converting mentions of place names into references to entries in the list of places.

²⁷ Based on the dataset provided by *Statistik Austria*, see below.

²⁸ Patocka (1997: 72 footnote 127) for example notes, “Der im Protokoll des Phonogrammarchivs angegebene Geburtsort ‘Edelbach’ existiert nicht [The place of birth specified in the Phonogrammarchiv’s protocol as ‘Edelbach’ does not exist]”. There is, however, a town by the name of Edlbach (without a second “e”). Comparable mistakes in the spelling of place names are extremely common, likely due to the fact that speakers only verbally communicated those names.

First, we obtained a solid basis for the standardisation of the town names and a source of reliable geographic coordinates. We approached a local authority, *Statistik Austria*, which provided us with an up-to-date and official dataset of all towns in the country of Austria, including their official administrative names and geodata as well as the larger administrative units (municipalities, districts, provinces) to which they belong. With the help of this data set, it was possible for us to set up a representation of place names in such a way that they are not only identified by their official designations (and geographical coordinates) but also are embedded in the hierarchy of the respective administrative units, where each level is embedded under the next higher level (i.e. PLACE < MUNICIPALITY < DISTRICT < PROVINCE < STATE), with the option of also adding alternative names of a place (e.g. potential historical names of a place or its name in other languages). In addition, we created a list of only loosely defined cultural regions outside the administrative hierarchy (e.g. *Bucklige Welt*, *Salzkammergut*) and, where relevant, linked the name of a place also to the respective region in that list. This makes it possible not only to unambiguously identify a given place but also to search for places in a structured, systematic way. We then manually added all places outside of Austria that are relevant to our corpus, and finally matched up the resulting dataset to the list of text entries in our database.



Figure 10. Village boundary signs of Boscangrove/Boskangruabe/Boscangrobe
(from Google Maps Street View)

In this new system, any place entry in the list is a complex data set of its own, containing e.g. potential alternative names as aliases, geographical coordinates, historical notes or other indications. This data set is automatically associated with each individual mention of that place. Thus, it is now easy to support e.g. place names in multiple languages, or to add a note that a town does not exist anymore, directly to the basic entry of that town, without having to repeat the procedure for each respective free text field in the data set of each relevant recording. Additionally, automated queries can be made on the basis of reliable, unique identifiers within our location data and do not need to be based on comparisons of text that are subject to all the imprecisions mentioned above.

We hope that this change will improve the usability of the corpus in a number of ways: by enhancing discoverability, through increasing the options for making queries on the data, and by optimising the reliability of those queries and their results. In this sense, this work also constituted a necessary step towards our aim of making the corpus publicly accessible, since this includes, as a basic requirement, the possibility to display any of our data on maps.

4.3 Some legal considerations

Beside questions concerning technical and IT-related aspects, the project also faces ethical and legal questions: Since the recordings in the corpus represent the Austrian dialects of German at stages of up to 70 years ago that often can no longer be encountered today, it would in principle be desirable also to make the recordings themselves available to a wider audience. However, this is where ethical and legal considerations come into play. It is evident that great care has to be taken in this matter. The legal status of such field recordings is not yet clear, and it must first be examined which legal regulations actually apply to them (also, a handful of speakers were still underage at the time of recording). Moreover, even if there are no legal objections, it must still be examined for each recording individually whether a publication is ethically acceptable or whether there are reasons to refrain from making it publicly available. When dealing with corpora of historical field recordings it must be kept in mind that at the time when these recordings were made, it was not customary to have recorded speakers undersign comprehensive legal documents regulating each and every aspect of the use of the recordings or their personal data (a comparable state of affairs holds even with present-day field recordings from remote or off-road areas of speakers who might not exhibit a Western academic level of literacy). In today's frequent calls for extensive and unconstrained dissemination of field data, these aspects are generally overlooked (or ignored).

While first legal consultations have already taken place, it will be an important future task of the project to conduct further research in this direction. In the end, these questions are relevant not only to the corpus at hand but also to historical recordings from field research in general.

5 Outlook

Until the end of the year 2021, we intend to complete the digital description of the recordings belonging to the “extended corpus” (see section 2.3 above; the digitisation of these files could be completed still in 2020). Beyond that, we are currently assessing the technical and legal options for establishing an online platform to present the corpus, and we are in the process of formulating a time schedule for its development. This being a complex and longer-term undertaking (and also depending on the funding situation), however, no realistic estimate can presently be made as to when the platform will be fully functional and can be made accessible.

Acknowledgments

This project would not exist without the contributions of both project partners. Alexandra Lenz, head of the Research Department “Variation and Change of German in Austria” (Austrian Academy of Sciences) and spokesperson of the Austrian Science Fund Special Research Programme “German in Austria: Variation – Contact – Perception” (F60), made it possible to employ a person for entering metadata and related work, and participated in covering the programming costs. The *Phonogrammarchiv*, as an in-kind contribution, took care of digitising the analogue audio tapes and of associated procedures, and covers further metadata-related work and also part of the programming costs. We would also like to thank Ludwig Maximilian Breuer for his support in the initial stage of the project, and to express our gratitude to Gernot Katzlberger of *Statistik Austria* for making available to

us the geodata of nearly 17300 places in Austria as recorded in the “Österreichisches Ortsverzeichnis [Gazetteer of Austria] 2015”. We are also indebted to the anonymous referees for their suggestions and comments.

References

- Allen, Harold B. 1952. The Linguistic Atlas of the Upper Midwest of the United States. *Orbis* 1(1), 89–94.
- Allen, Harold B. 1977. Regional Dialects, 1945–1974. *American Speech* 52(3/4), 163–261.
- Austrian Science Fund (FWF). 1975. *Bericht an das Bundesministerium für Wissenschaft und Forschung über das Jahr 1974 und die Lage der wissenschaftlichen Forschung*. URL https://www.fwf.ac.at/fileadmin/files/Dokumente/Ueber_den_FWF/Publikationen/FWF-Jahresberichte/fwf-jahresbericht-1974.pdf (last accessed 2021-02-06)
- Bauer, Werner. 1967. *Lautlehre der Mundart des Lungaus mit besonderer Berücksichtigung von Zederhaus*. Vienna: University of Vienna dissertation.
- Bellmann, Günter. 1964. Wege und Möglichkeiten der Schallaufnahme ostdeutscher Mundarten heute. Zur Tonbandaufnahme der Vertriebenenmundarten. *Zeitschrift für Mundartforschung* 31(1), 62–79.
- Berchtold, Simone & Oliver Schallert. 2013. *Die Mundarten der Walsertäler und des Tannbergs* (Die Mundarten Vorarlbergs 11; Schriften der Vorarlberger Landesbibliothek 3/11). Graz: Neugebauer.
- Bouterwek, Friedrich. 1988. *Aspekte der Dialektveränderung. Theoretische Grundlagen und eine empirische Untersuchung in der 'Buckligen Welt' in Niederösterreich*. Vienna: University of Vienna dissertation.
- Brandstetter, Alois. 1961. *Laut- und bedeutungskundliche Untersuchungen an der Mundart von Pichl bei Wels*. Vienna: University of Vienna dissertation.
- Chambers, John K. & Peter Trudgill. 1998. *Dialectology*. 2nd revised edn. (Cambridge Textbooks in Linguistics). Cambridge & New York: Cambridge University Press
- Fleischer, Jürg. 2017. *Geschichte, Anlage und Durchführung der Fragebogen-Erhebungen von Georg Wenkers 40 Sätzen: Dokumentation, Entdeckungen und Neubewertungen*. Hildesheim: Olms.
- Gabriel, Eugen. 1965. Der Vorarlberger Sprachatlas. *Montfort – Zeitschrift für Geschichte, Heimat- und Volkskunde Vorarlbergs* 17(1), 48–63.
- Gabriel, Eugen. 1972a. Der Vorarlberger Sprachatlas mit Einschluß des Fürstentums Liechtenstein. *Montfort – Vierteljahresschrift für Geschichte und Gegenwartskunde Vorarlbergs* 24(1), 159–239.
- Gabriel, Eugen. 1972b. Review of Maria Hornung, *Mundartkunde Osttirols. Eine dialektgeographische Darstellung mit volkskundlichen Einblicken in die alpbäuerliche Lebenswelt*. Graz, Wien, Köln: Hermann Böhlhaus Nachfahren, 1964 [=Hornung (1964)]. *Zeitschrift für deutsches Altertum und deutsche Literatur* 101(4), 261–268.
- Geyer, Ingeborg. 1976. *Die deutsche Mundart von Tischelwang (Timau) in Kärnten (Oberitalien)*. Vienna: University of Vienna dissertation. Printed 1984 in *Beiträge zur Sprachinselforschung* Band 3, Wien: VWGÖ.
- Hajek, Leo. 1928. Das Phonogrammarchiv der Akademie der Wissenschaften in Wien von seiner Gründung bis zur Neueinrichtung im Jahre 1927 (58. Mitteilung der Phonogrammarchivs-Kommission). *Sitzungsberichte der Akademie der Wissenschaften in Wien, philosophisch-historische Klasse* 207/3), 1–22. Wien, Leipzig: Hölder-Pichler-Tempsky.
- Hornung, Maria. 1958. Zur Verwendung des Lehms im burgenländischen Hausbau. *Burgenländische Heimatblätter* 20, 56–62.
- Hornung, Maria. 1960. Die Osttiroler Bauernsprachinseln Pladen und Zahre in Oberkärnten. *Osttiroler Heimatblätter* 28, Sonderdruck.
- Hornung, Maria. 1961. Tonaufnahmen im Dienste der Mundartforschung. Zum 60jährigen Bestehen des Phonogrammarchivs der österreichischen Akademie der Wissenschaften in Wien. *Zeitschrift für Mundartforschung* 28(2), 183–191.
- Hornung, Maria. 1964a. *Mundartkunde Osttirols. Eine dialektgeographische Darstellung mit volkskundlichen Einblicken in die alpbäuerliche Lebenswelt* (Studien zur österreichisch-bairischen Dialektkunde 3). Graz, Wien, Köln: Hermann Böhlhaus Nachfahren.
- Hornung, Maria. 1964b. *Rauchküche und Rauchstube in Osttirol*. Graz, Wien, Köln: Hermann Böhlhaus Nachfahren.
- Hornung, Maria. 1967. Eberhard Kranzmayer – Ein Siebziger. *Burgenländische Heimatblätter* 29(2), 49–50.
- Hornung, Maria. 1968. Dämonische Sagengestalten in burgenländischer Volksüberlieferung. *Burgenländische Heimatblätter* 30, 97–105.
- Hornung, Maria. 1970. Wiener Lebensbilder an Hand von Tonaufnahmen. *Wiener Geschichtsblätter* 25(87)(4), 97–105.

- Hornung, Maria. 1972. *Wörterbuch der deutschen Sprachinselmundart von Pladen/Sappada in Karnien (Italien) mit Verwertung der Sammlungen von Pietro Sartor Schlossar † (Sappada – Rom)* (Studien zur österreichisch-bairischen Dialektkunde 6). Graz, Wien, Köln: Hermann Böhlaus Nachfahren.
- Hornung, Maria. 1977. Sprachmischung im ostoberitalienischen Sprachinselraum. In Herbert Kolb & Hartmut Lauffer (eds.), *Sprachliche Interferenz. Festschrift für Werner Betz zum 65. Geburtstag*, 196–213. Tübingen: Max Niemeyer.
- Hornung, Maria. 1981. Deutsche Mundart [des Bezirks Mattersburg]. *Allgemeine Landestopographie des Burgenlandes, Band 3 Teil 1*, 266–268. Eisenstadt: Selbstverlag des Amtes der Burgenländischen Landesregierung, Abt. XII/2 (Landesarchiv-Landesbibliothek).
- Hornung, Maria. 1983. Die mündlich tradierte Volkserzählung in Österreich unserer Zeit und in den altösterreichischen Sprachinseln. In Rainer Wehse (ed.), *Märchenerzähler – Erzählgemeinschaft* (Veröffentlichungen der Europäischen Märchengesellschaft 4), 30–38. Kassel: Erich Röth.
- Hornung, Maria. 1999. Die heanzischen Mundarten des Burgenlandes im Wandel unseres Jahrhunderts. In Ferenc Szász & Imre Kurdi (eds.), *Im Dienste der Auslandsgermanistik, Festschrift für Professor Dr. Dr. h.c. Antal Mádl zum 70. Geburtstag*, 87–95. Budapest: ELTE Germanistisches Institut.
- Hornung, Maria & Sigmar Grüner. 2002. *Wörterbuch der Wiener Mundart, 2. erweiterte und verbesserte Auflage*. Wien: OBV & HPT.
- Hornung, Maria & Franz Roitinger. 2000. *Die österreichischen Mundarten. Eine Einführung*. Wien: Österreichischer Bundesverlag.
- Hotzenköcherle, Rudolf. 1939. Ein schweizerdeutscher Sprachatlas. *Zeitschrift für Mundartforschung* 15(3/4), 183–188.
- Hotzenköcherle, Rudolf. 1962. Zur Methodologie der Kleinraumatlanten. *Zeitschrift für Mundartforschung* 29(2), 133–149.
- Huber, Christian, Benjamin Fischer & Bernhard Graf. 2019. Corpus of Austrian Dialect Recordings from the 20th Century – A Cooperation Project. In Michael Pucher, Jürgen Trouvain & Carina Lozo (eds.), *HSCR 2019 – Proceedings of the Third International Workshop on the History of Speech Communication Research Vienna, September 13–14, 2019* (Studentexte zur Sprachkommunikation 94), 31–40. Dresden: TUDpress.
- Kranzmayer, Eberhard. 1954. Deutsche Mundart [des Bezirks Neusiedl]. *Allgemeine Landestopographie des Burgenlandes, Band 1*, 123–126. Eisenstadt: Selbstverlag des Amtes der Burgenländischen Landesregierung, Landesarchiv.
- Kranzmayer, Eberhard. 1963. Deutsche Mundart [des Bezirks Eisenstadt und der Freistädte Eisenstadt und Rust]. *Allgemeine Landestopographie des Burgenlandes, Band 2 Teil 1*, 266–268. Eisenstadt: Selbstverlag des Amtes der Burgenländischen Landesregierung, Landesarchiv.
- Kranzmayer, Eberhard. 1965. Die Kanzlei zur Schaffung eines bairisch-österreichischen Wörterbuchs und zur Erforschung unserer Mundarten (Österreichische Akademie der Wissenschaften). *Zeitschrift für Mundartforschung*, 32(2), 129–130.
- Lechleitner, Franz. 1999. Vertrauen ist gut, Kontrolle ist besser! 100 Jahre Technik im Phonogrammarchiv. *Das audiovisuelle Archiv* 45, 16–22.
- Lenz, Alexandra, Christian Huber, Ludwig Maximilian Breuer, Benjamin Fischer & Bernhard Graf. 2020. „Österreichische Dialektaufnahmen im 20. Jahrhundert“ – zur Genese, Aufbereitung und wissenschaftlichen Nutzung eines einmaligen Sprachkorpus. *International Forum on Audio-Visual Research – Jahrbuch des Phonogrammarchivs* 10, 128–140.
- Lipold, Günter. 1979. *System, Stil und Prozeß: Phonologie der Sprachinselmundart von Gottschie in Jugoslawien*. Vienna: University of Vienna habilitation.
- Linn, Michael D. 1993. Appendix: Resources for Research. In Dennis R. Preston (ed.), *American Dialect Research*, 433–450. Amsterdam, Bethlehem Pike: John Benjamins.
- Marschall, Ingrid. 1975. *Der mundartliche Wortschatz Josef Weinhebers im gegenwärtigen Wienerischen: Eine altersspezifische Untersuchung*. Vienna: University of Vienna dissertation.
- McDavid, Raven I. 1957. Tape Recording in Dialect Geography: A Cautionary Note. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 3(1), 3–8.
- Patocka, Franz. 1997. *Satzgliedstellung in den bairischen Dialekten Österreichs* (Schriften zur deutschen Sprache in Österreich 20). Frankfurt am Main: Peter Lang.
- Patocka, Franz & Hermann Scheuringer. 1988. *Fragebuch für die bairischen Mundarten in Österreich und Südtirol*. 4., völlig neu bearbeitete Auflage. Ms.: Wien.
- Purnell, Thomas. 2013. Hearing The American Language Change: The State of DARE Recordings. *American Speech* 88(3), 275–302.
- Pollak, Hans W. 1913. Die Aufnahme deutscher Mundarten durch das Phonogramm-Archiv der kaiserlichen Akademie der Wissenschaften in Wien. *Zeitschrift für Deutsche Mundarten* 8, 83–88.

- Riemann, Erhard. 1964. Erfahrungen bei der Tonbandaufnahme der Vertriebenenmundarten. *Zeitschrift für Mundartforschung* 31(1), 79–87.
- Ruoff, Arno. 1965. Wenkersätze auf Tonband? In: Hermann Bluhme, *Sprachen – Zuordnung – Strukturen. Festgabe seiner Schüler für Eberhard Zwirner*, 94–113. Den Haag: Nijhoff.
- Ruoff, Arno. 1973. Grundlagen und Methoden der Untersuchung gesprochener Sprache: Einführung in die Reihe »Idiomática« mit einem Katalog der ausgewerteten Tonbandaufnahmen (IDIOMATICA – Veröffentlichungen der Tübinger Arbeitsstelle »Sprache in Südwestdeutschland« 1). Tübingen: Max Niemeyer.
- Ruoff, Arno. 1985. *Alltagstexte II: Transkriptionen von Tonbandaufnahmen aus Baden-Württemberg, Vorarlberg und Liechtenstein* (IDIOMATICA – Veröffentlichungen der Tübinger Arbeitsstelle »Sprache in Südwestdeutschland« 11). Tübingen: Max Niemeyer.
- Ruoff, Arno. 1993. Das Tonarchiv der Mundarten Vorarlbergs. *Montfort – Vierteljahresschrift für Geschichte und Gegenwart Vorarlbergs* 45(1), 21–32.
- Ruoff, Arno. 1997. Eugen Gabriel und der Vorarlberger Sprachatlas. Dem einen zum 60., dem anderen zum 12. Geburtstag. *Montfort – Vierteljahresschrift für Geschichte und Gegenwart Vorarlbergs* 49(1), 9–13.
- Ruoff, Arno & Eugen Gabriel. 1998–2004. *Die Mundarten Vorarlbergs* (10 volumes; Schriften der Vorarlberger Landesbibliothek 3). Graz: Neugebauer.
- Schabus, Wilfried. 1971. *Dialektgeographie des Lesachtals*. Vienna: University of Vienna dissertation.
- Schabus, Wilfried. 1993. Die Mundart des Lesachtals. *Carinthia* 183, 543–564.
- Schabus, Wilfried. 1999. Die Bestände des Phonogrammarchivs an Sprachaufnahmen. *Das audiovisuelle Archiv* 45, 23–32.
- Schallert, Oliver. 2014. *Zur Syntax der Ersatzinfinitivkonstruktion. Typologie und Variation* (Studien zur deutschen Grammatik 87). Tübingen: Stauffenburg.
- Scheuringer, Hermann. 1982. *Sprachzustand und Sprachveränderung im westlichen Sawwald in Oberösterreich: Lautlehre*. Vienna: University of Vienna dissertation.
- Scheuringer, Hermann. 1990. *Sprachentwicklung in Bayern und Österreich*. Hamburg: Helmut Buske.
- Schwarz, Artur. 1957. Mundartaufnahmen im Lande Vorarlberg. *Montfort – Zeitschrift für Geschichte, Heimat- und Volkskunde Vorarlbergs* 9(1), 112–116.
- Seifert, Lester W. J. 1949. The Problem of Speech-Mixture in the German Spoken in Northwestern Dane County, Wisconsin. *Transactions of the Wisconsin Academy of Sciences, Arts and Letters* 39, 127–139.
- Seifert, Lester W. J. 1951. Methods and Aims of a Survey of the German Spoken in Wisconsin. *Transactions of the Wisconsin Academy of Sciences, Arts and Letters* 40(2), 201–210.
- Stift, Ulf-Michael & Thomas Schmidt. 2014. Mündliche Korpora am IDS: Vom Deutschen Spracharchiv zur Datenbank für Gesprochenes Deutsch. In Institut für Deutsche Sprache (ed.), *Ansichten und Einsichten. 50 Jahre Institut für Deutsche Sprache*, 360–375. Mannheim: Institut für Deutsche Sprache.
- Tatzreiter, Herbert. 1964. *Sprachbiologische Untersuchungen an der Mundart von Steinbach am Ziehberg*. Vienna: University of Vienna dissertation.
- Tatzreiter, Herbert. 1965. Auf Mundart-Kundfahrt durchs Mühlviertel. *Mühlviertler Heimatblätter* 5(9/10), 153–155.
- Tatzreiter, Herbert. 1978a. Norm und Varietät in Ortsmundarten. Mit Beispielen aus dem steirischen Obermurgebiet. *Zeitschrift für Dialektologie und Linguistik* 45(2), 133–148.
- Tatzreiter, Herbert. 1978b. *Beharrsamkeit, Varietät und Wandel im Dialekt des Obermurgebietes (Vokalismus)*. Vienna: University of Vienna habilitation.
- Wagner, Peter & Karl-Heinz Bausch (eds.). 1997. *Tonaufnahmen des gesprochenen Deutsch: Dokumentation der Bestände von sprachwissenschaftlichen Forschungsprojekten und Archiven* (Phonai - Texte und Untersuchungen zum gesprochenen Deutsch 40). Tübingen: Max Niemeyer.
- Wilson, Rex. 1956. The Implication of Tape Recording in the Field of Dialect Geography. *Journal of the Canadian Linguistic Society* 2(1), 17–22.
- Zwirner, Eberhard. 1956. Lautdenkmal der deutschen Sprache. *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft* 9: 3–13.
- Zwirner, Eberhard & Wolfgang Bethge. 1958. *Erläuterungen zu den Texten* (Lautbibliothek der deutschen Mundarten 1). Göttingen: Vandenhoeck & Ruprecht.

URLs

- Adalbert-Stifter-Institut des Landes Oberösterreich, Stifter-Haus, Spracharchiv: URL <https://stifterhaus.at/index.php?id=20> (last accessed 2021-02-06)
- Archiv für gesprochenes Deutsch, Zwirner-Korpus: URL http://agd.ids-mannheim.de/ZW--_extern.shtml (last accessed 2021-02-03)
- Dialektatlas Mittleres Westdeutschland (DMW): URL <https://www.dmw-projekt.de/> (last accessed 2021-02-03)

Linguistic Atlas Project: Linguistic Atlas of New England (LANE). URL <http://www.lap.uga.edu/Site/LANE.html> (last accessed 2021-02-03)
Phonogrammarchiv, Online Catalogue: URL <http://catalog.phonogrammarchiv.at/>
Syntax hessischer Dialekte (SyHD): URL <http://www.syhd.info> (last accessed 2021-02-03)
Transkribus: URL <https://readcoop.eu> (last accessed 2021-02-03)
UNESCO – Dokumentenerbe „Memory of the World“/„Gedächtnis der Menschheit“ – *Tonaufnahmen österreichischer Dialekte 1951–1983 (aufgenommen 2018)*. URL <https://www.unesco.at/kommunikation/dokumentenerbe/memory-of-austria/verzeichnis/detail/article/tonaufnahmen-oesterreichischer-dialekte-1951-1983/> (last accessed 2021-02-06)
Wikipedia s.v. *SoundScriber*: URL <https://en.wikipedia.org/wiki/SoundScriber> (last accessed 2021-02-06)