Hans C. Boas und Matthias Fingerhuth

7 Deutsche Sprachinselkorpora im 21. Jahrhundert

Abstract: Dieser Beitrag umreißt die Dokumentation von Sprachinseln des Deutschen in ihrer historischen Entwicklung, ihren gegenwärtigen Stand sowie die Herausforderungen, die sich bei der Erhaltung und Verfügbarmachung dieser Daten für ein (Fach-)Publikum stellen. Nachdem die Sprachinselforschung anfänglich auf geschriebenen Notizen basierte, benutzte sie über weite Teile des 20. Jahrhunderts analoge Medien. Die analog gespeicherten Daten sind gefährdet, sowohl durch den Verfall der Medien selbst als auch durch den der Abspielgeräte. Darüber hinaus schränken die analogen Medien Zugang zu und Vervielfältigung der Daten stark ein. Durch Digitalisierung kann nicht nur die langfristige Archivierung gewährleistet, sondern auch ein vereinfachter globaler Zugang gewährt werden. Obwohl somit technische Lösungen für die dringendsten Probleme der analogen Aufnahmen existieren, gibt es weitere Hürden für die Erstellung zugänglicher Archive. Die Sprachinselforschung hat sich über lange Zeit weitgehend dezentral entwickelt, und einzelne Forschungsprojekte haben jeweils eigene Methodologien und Forschungsinfrastrukturen verwendet. In manchen Fällen sind erhobene Daten verschollen, auf der anderen Seite sind nicht sämtliche erhaltenen Daten dokumentiert. Darüber hinaus geben rechtliche Vorgaben gegenwärtig einen engen Rahmen für die Archivierung und Nutzung von Sprecherdaten ohne die explizite Einwilligung der Sprecher vor, die jedoch nicht bei allen existierenden Aufnahmen vorliegt. Trotz dieser Hürden gibt es Bemühungen, die Archivierung und Veröffentlichung historischer Sprachinseldaten zu verbessern, und die Erhebung neuer Daten zu koordinieren. Nennenswerte Beispiele dafür sind die *Datenbank für Gesprochenes Deutsch* am Institut für Deutsche Sprache in Mannheim sowie das *Sprachinselarchiv* an der University of Texas at Austin.

Anmerkung: Für hilfreiche Kommentare bedanken wir uns bei Csaba Földes, Nicole Eller-Wildfeuer, Claudia Riehl, Peter Wagener und den Herausgebern dieses Bandes.

Hans C. Boas, Department of Germanic Studies, The University of Texas at Austin, 1 University Station, C3300, 2505 University Blvd., Austin, Texas 78712, U.S.A., E-Mail: hcb@mail.utexas.edu
Matthias Fingerhuth, Institut für Germanistik, Universität Wien, Universitätsring 1, A-1010 Wien, Österreich, E-Mail: matthias.fingerhuth@univie.ac.at

Open Access. © 2018 Hans C. Boas und Matthias Fingerhuth, publiziert von De Gruyter. CC BY
Dieses Werk ist lizenziert unter der Creative Commons Attribution 4.0 Lizenz.
https://doi.org/10.1515/9783110538649-008

Keywords: Dialektologie, Spracharchiv, Sprachdokumentation, Sprachinsel, Sprachkontakt

1 Einleitung

In diesem Beitrag wird über Korpora berichtet, mit denen deutsche Sprachinseldialekte untersucht werden können. Der Schwerpunkt liegt auf digitalen Sprachinselkorpora, die seit Ende des 20. Jahrhunderts in der sprachwissenschaftlichen Forschung immer mehr an Bedeutung gewonnen haben, gerade auch im Rahmen einer intensiveren Beschäftigung mit empirischen Methoden und Einsichten der Korpuslinguistik (Lüdeling & Kytö 2008/2009). Dieser Beitrag ist wie folgt gegliedert: Abschnitt 2 gibt einen historischen Überblick über die Dokumentation und Archivierung von Sprachinseldaten.[1] Abschnitt 3 diskutiert zunächst technische und methodologische Aspekte der Erstellung von digitalen Sprachinselkorpora. Anschließend wird ein Überblick über eine Reihe unterschiedlicher digitaler Sprachinselkorpora gegeben. Abschnitt 4 diskutiert methodologische und technische Aspekte einer vergleichenden Sprachinselforschung mit digitalen Sprachinselkorpora und zeigt, wie gewisse sprachwissenschaftliche Probleme mit neuen technischen Methoden gelöst werden können. Abschnitt 5 fasst den Beitrag zusammen.

2 Historischer Überblick

2.1 Datenerhebung

Das im Laufe des 19. Jahrhunderts aufkommende Interesse an der Erforschung deutscher Dialekte führte zu unterschiedlichen Arten der Datenerhebung. Die womöglich bekannteste Art der damaligen Datenerhebung lässt sich auf Georg Wenker zurückführen, der 1878 den weltweit ersten Sprachatlas, den *Sprach-Atlas der Rheinprovinz nördlich der Mosel sowie des Kreises Siegen* veröffentlichte. Dieser Sprachatlas bildete die Grundlage für den späteren *Sprachatlas des Deutschen Reichs*, bestehend aus 1.668 von Hand gezeichneten Karten.

Die Datenerhebung selbst bestand aus der Versendung von Fragebögen mit 40 volkstümlichen Sätzen (die sogenannten Wenkersätze), die Wenker mit der

1 Zur Begriffsbestimmung des Begriffs „Sprachinsel“ siehe Wildfeuer (2017a).

Hilfe von Lehrern in die jeweiligen Ortsdialekte übersetzen ließ (indirekte Erhebung). Die zurückgeschickten Fragebögen dienten als Datenbasis für den Sprachatlas, in dem auf Teilkarten die erhobenen Daten eingetragen wurden. Später wurden Nacherhebungen für die Gebiete außerhalb des Deutschen Reiches vorgenommen, so dass für Europa eine nahezu vollständige Erhebung, einschließlich einiger deutscher „Sprachinseln", vorliegt. Insgesamt konnten 51.480 Bögen aus 49.363 deutschsprachigen Orten gesammelt werden, zusätzlich gingen 2.050 fremdsprachige Bögen ein. Alle Originalbögen sind im Forschungsinstitut „Deutscher Sprachatlas" in Marburg archiviert (Herrgen & Lenz 2003). Die von Wenker verwendeten 40 Sätze bildeten im 20. Jahrhundert auch die Basis für viele Sprachatlanten der sogenannten Marburger Schule, so dass es schrittweise gelang, die deutschen Dialekte mittels der Marburger Erhebungsmethode in einer gewissen Vollständigkeit zu erfassen (Knoop, Putschke & Wiegand 1982; Goebl & Schiltz 2006). Der Deutsche Sprachatlas bildete die Grundlage für zahlreiche Publikationen und inspirierte im Laufe des 20. Jahrhunderts auch viele Ortsgrammatiken sowie Sprachatlanten deutscher Sprachinseln (Reed & Seifert 1954; Klein & Schmitt 1961/1965; Gilbert 1972; Brenner, Erb & Manherz 2008).

Die von O. Bremer vorgebrachte Kritik an der Methodik der Marburger Schule führte dazu, dass auch direkt erhobene, in kleinräumigen Untersuchungen zusammengestellte Daten erhoben wurden. Als solches ist die direkte Befragung von Gewährspersonen die häufigste Form der dialektologischen Datenerhebung. Sie erlaubt es, kompetente Sprecher als Experten heranzuziehen und ihnen bedachtsam ausgewählte Fragen zur Beantwortung vorzulegen (Wagener 1988: 103–104). Obwohl es an Kritik zum Vorgehen des Sprachatlasses nicht mangelt (dazu mehr im Folgenden), und zahlreiche Unsicherheiten bezüglich der so erhobenen Daten verbleiben, kann man es doch als einen methodischen Fortschritt gegenüber vorhergehender Arbeit sehen.

Als Beispiel kann man etwa Schmellers (1838) Arbeit zum Zimbrischen anführen, über deren Entstehung das Skript eines Vortrags Schmellers Hinweise birgt, jedoch keine letztliche Aufklärung verschafft. So schreibt Schmeller, dass er während einer Reise in das zimbrische Gebiet Gelegenheit hatte, „verschiedene Personen methodisch nach den Hauptrubriken ‚meines Versuches über die Mundarten Bayerns' zu vernehmen" (Schmeller 1838: 588). Die Ausführungen zur Methodologie sind in der erwähnten Schrift *Die Mundarten Bayerns* jedoch bemerkenswert kurz und beschränken sich auf den Hinweis, dass er viele der zugrunde liegenden Notizen selbst bei Wanderungen durch das Land und durch „planmäßige Vernehmung neu eingereihter Conscribierten" (Schmeller 1821: XI) erhoben hat. Aus dem Weiteren in Schmellers Vortrag Erwähnten geht jedoch auch hervor, dass Schmeller bei seiner Reise auch bemüht war, schriftliche Quellen zu kaufen oder zumindest zu kopieren.

Bei der Erhebung sprachlicher Daten wurden im vortechnischen Zeitalter traditionell nur Papier und Stift für die Erfassung sprachlicher Daten (und Metadaten) benutzt, erst später ist die Ton- und Bildaufzeichnung zu einem gängigen Werkzeug der Forschung geworden. So war z. B. in der ersten Hälfte des 20. Jahrhunderts das wissenschaftliche Interesse an der Ergründung menschlicher Sprache eine treibende Kraft für die Entwicklung der Technik für Tonaufnahmen (Schüller 2008: 4). Ungeachtet dieses Umstands und des regen Interesses für Dialektologie um die Jahrhundertwende mag es ein wenig überraschen, dass die noch junge Technik bereits Anfang des 20. Jahrhunderts zur Aufnahme von Sprachinseln eingesetzt wurde. So finden sich im Wiener Phonogrammarchiv bereits 1912 von Anton Pfalz in den Sieben Gemeinden gemachte Aufnahmen des Zimbrischen.[2]

Die Methodik der Datenerhebung wird in der frühen Sprachinselforschung teilweise nur wenig diskutiert und ist deshalb auch nur bedingt nachvollziehbar. Die Einblicke, die sich ergeben, suggerieren ein (aus heutiger Sicht) teils wenig systematisches Vorgehen. Über die von Anton Pfalz durchgeführten Aufnahmen der Zimbern in den Sieben Gemeinden erfahren wir etwa Folgendes: „Der Sprecher, [...] ein etwa fünfzigjähriger Zimber, hat die Aufnahme A und B a, b frei erfunden, B c–h nach der dort angegebenen Quelle in den Apparat gesprochen." (Lessiak & Pfalz 1918: 59) A verweist dabei auf eine Reihe von nur teilweise zusammenhängenden Sätzen, deren thematischer Schwerpunkt sich vielleicht mit ‚folkloristisch' beschreiben lässt, B auf zimbrische Sprichwörter, die meisten derer aus einer philologischen Zeitschrift entnommen sind.

Daten jüngeren Datums sind dabei nicht automatisch zeitgemäßer, wie sich, um beim Zimbrischen zu bleiben, an den von Bruno Schweizer erhobenen Daten aus den Dreizehn Gemeinden zeigt. Zur Datenerhebung schreibt er etwa:

> Ich beginne mit dem Abdruck ganz einfacher Zusammenstellungen und Redensarten, wie sie mir von meinen zimbrischen Freunden zur Einführung ihre Sprache vorgetragen oder diktiert wurden. Die Fragen und Antworten unter Nr. 9 [Gespräche] sind von meinem Herbergsvater Stefano Nordera aus einem italienisch-deutschen Sprachführer übersetzt. Leider war dieser wackere Mann von seinen täglichen Arbeiten und von der Bewirtung seiner Gäste immer so in Anspruch genommen, daß er nur selten Zeit fand, sich ein wenig für die zimbrische Sprachforschung zur Verfügung zu stellen. (Schweizer 1939: 18–19)

Für die große Mehrheit der Sprachproben ist jedoch die Quelle nicht nachvollziehbar. Zur Methode bemerkt Schweizer weiter, dass es sich „größtenteils um

2 Zu Anfang des 20. Jahrhunderts wurden an mehreren Forschungsinstitutionen Tonarchive zu Forschungszwecken gegründet: Wien (1899), Berlin (1900), St. Petersburg (1908) und Zürich (1909), vgl. Schüller (2008).

von mir selbst abgehörte und nach dem Wortlaut vom Munde des Erzählers weg in Lautschrift festgehaltene Texte“ (Schweizer 1939: 9) handelt.

Die Methodenvielfalt lässt sich auch im weiteren Verlauf des 20. Jahrhunderts beobachten. Ein kurzer Einblick in die nordamerikanische Forschung kann dies, wenn nicht repräsentativ, so doch exemplarisch verdeutlichen. Die ersten sprachwissenschaftlichen Daten zum Texasdeutschen wurden in den 1930er und 1940er Jahren von Fred Eikel auf Grundlage von Übersetzungsaufgaben erhoben, bei der 191 Sätze aus dem Englischen ins Texasdeutsche übersetzt wurden (Eikel 1949). Als nächstes erhob Glenn Gilbert Anfang der 1960er Jahre Daten zum Texasdeutschen, abermals ausgehend von Übersetzungsaufgaben, jedoch auf Grundlage von Wortlisten, die er von Reed & Seifert (1954) und Atwood (1962) abgewandelt hatte (Gilbert 1963: 28–63). Dies ist insofern bemerkenswert, da auch Firchow (1991: 260) berichtet, in Minnesota auf Grundlage von Materialien zu arbeiten, die sie von Moellecken erhalten hatte, welcher sie wiederum von Reed und Seifert adaptiert hatte (cf. Moelleken 1988: 109). Hier gibt es also methodologische Überschneidungen zwischen verschiedenen Sprachinseln. Der in Fortsetzung dieser Arbeit von Gilbert (1972) erstellte *Linguistic Atlas of Texas German* basiert dagegen abermals auf der Übersetzung ganzer Sätze, die jedoch nicht mit denen Eikels übereinstimmen (Boas 2009: 9–10), und auch nicht mit der Methode von Reed und Seifert. Eine dritte Welle der Dokumentation erfolgt schließlich seit 2001 mit dem *Texas German Dialect Project* (TGDP).[3] Wie unten noch detaillierter beschrieben wird, greift das TGDP die Methode von Eikel und Gilberts Sprachatlas auf, um einen historischen Vergleich zu ermöglichen. Nimmt man jedoch weitere Projekte hinzu, so tun sich weitere Unterschiede auf. Für den *Linguistic Atlas of Kansas German*[4] wurden beispielsweise Wenkersätze sowie ein eigens konzipierter Fragebogen verwendet.

Trotz teilweiser Überschneidungen summieren sich damit die Unterschiede zwischen den verschiedenen Korpora. Eine methodologische Vielfalt ist sicher nicht grundsätzlich beklagenswert, da unterschiedliche Ansätze unterschiedlichen Erkenntniszielen entsprechen können und ein Festhalten an einer einzelnen Erhebungsweise schwer mit methodologischem Fortschritt vereinbar scheint. Es ist sicher auch so, dass sich die Verhältnisse innerhalb der verschiedenen Sprachinseln deutlich unterscheiden können, so dass Fragebögen, die auf eine Erhebung in Texas abzielen, für eine Erhebung in Wisconsin oder Sibirien ohne eine Anpassung problematisch sind. Auch das Interesse für bestimmte sprachliche Phänomene, auf deren Erkundung etwa Übersetzungs-

3 http://www.tgdp.org (letzter Zugriff: 25. 6. 2017).

4 http://www2.ku.edu/~germanic/LAKGD/Atlas_Intro.shtml (letzter Zugriff: 25. 6. 2017).

aufgaben gesondert abzielen können, kann in direkter Abhängigkeit von der Kontaktsprache stehen. Dennoch ist die bewusste Abstimmung zwischen Projekten, wie man sie in der Arbeit der letzten Jahre vermehrt beobachten kann, eine Entwicklung, die den Vergleich zwischen Erhebungen erleichtern dürfte.

2.2 Analoge Datenarchivierung und -darstellung

Ab den 1930er Jahren gab es dank technischen Fortschritts preiswertere und leichter zu benutzende Aufnahmegeräte und Mikrofone, die es den Forschern ermöglichten, mehr Tonaufnahmen durchzuführen. Neben den Tonaufnahmen fertigten Forscher auch weiterhin Fragebögen und andere schriftliche Beobachtungen an, die als Metadaten und begleitendes Forschungsmaterial mit den Tonträgern archiviert wurden. Viele Tonaufnahmen und ihre Begleitmaterialien wurden an Forschungsinstituten, Universitäten und Bibliotheken archiviert. Das wahrscheinlich bekannteste Archiv ist das Deutsche Spracharchiv, das bereits 1932 von Eberhard Zwirner in Berlin gegründet wurde, um die Analyse konstitutiver Faktoren und die Struktur gesprochener Sprache zu analysieren (Knetschke & Sperlbaum 1983). Die Tonaufnahmen wurden auf Schallplatten aufgenommen, dazu wurden Transkripte sowie Kurven und Messwerte zur phonometrischen Auswertung angefertigt (siehe Zwirner 1983 und Stift & Schmidt 2014).

Im Bereich der Sprachinselforschung gibt es das von Viktor Schirmunski an der Universität Leningrad 1927 gegründete Volksliedarchiv, dessen Daten für die Untersuchung sprachlicher Strukturen zwar nur begrenzt nützlich sind, aber dennoch ein interessantes Zeitzeugnis russlanddeutscher Sprache darstellen. Die Aufnahmen wurden von 1924–1931 aufgezeichnet und in Leningrad (heute St. Petersburg) archiviert (s. John & Swetosarowa 2005).[5] Einige Tonaufnahmen deutscher Sprachinseln finden erst Jahrzehnte später ihren Weg zu wissenschaftlichen Institutionen, wie z. B. die von Glenn Gilbert in den 1960er Jahren gemachten Tonaufnahmen, die 2006 an das Max Kade-Institut an der University of Wisconsin at Madison gegangen sind, oder die von Michael Clyne über mehrere Jahrzehnte geschaffenen Tonaufnahmen des Australiendeutschen, die 2008 an das Institut für Deutsche Sprache (IDS) in Mannheim kamen.

In den Jahren von 1991 bis 1993 hat sich das IDS um die Dokumentation der zu diesem Zeitpunkt in germanistischen Instituten, Archiven und Forschungsprojekten vorliegenden Aufnahmen des gesprochenen Deutsch bemüht,

5 http://www.liederlexikon.de/ueber_liederlexikon_de/projekte/bkm_projekt (letzter Zugriff: 25. 6. 2017).

indem es eine entsprechende Umfrage durchgeführt und die Ergebnisse in Folge veröffentlicht hat (Wagener & Bausch 1997). Die dort vorgenommene Erhebung weist nach Angabe der Autoren Lücken auf und war aufgrund des zeitlichen Abstandes zwischen Datenerhebung und Publikation in manchen Fällen bereits zum Veröffentlichungszeitpunkt nicht mehr aktuell, was aus heutiger Sicht umso mehr gelten muss. Dennoch stellt sie die bislang wohl vollständigste Dokumentation von Aufnahmen des Deutschen außerhalb des binnendeutschen Sprachraums dar. Einige der von Wagener & Bausch (1997) dokumentierten Aufnahmen sind mittlerweile ins Archiv für gesprochenes Deutsch aufgenommen bzw. über die Datenbank für Gesprochenes Deutsch (DGD) des IDS zugänglich gemacht worden. Obwohl analog aufgenommen, können diese Korpora damit inzwischen als digital angesehen werden, weshalb Details zum Fortschritt dieser Arbeit entsprechend weiter unten besprochen werden.

Außer solch größeren Sammlungen archivierter Tonaufnahmen ist jedoch nicht klar, wie viele Aufnahmen Forscher in ihren Privatbeständen belassen haben und diese nicht den Weg in wissenschaftliche Archive gefunden haben. Als Beispiel seien hier die Arbeiten von Fred Eikel (1949, 1954, 1966) zum New Braunfels Texas German genannt, welche auf Tonaufnahmen beruhen, die in den frühen 1940er Jahren aufgenommen wurden. Diese Tonaufnahmen wurden weder an der Johns Hopkins University, wo Eikel promoviert haben soll, noch an einer anderen Institution, wie z. B. dem New Braunfelser Sophienburg Museum archiviert. Per Zufall stießen Mitarbeiter des Texas German Dialect Projects im Herbst 2006 bei einem Interview auf eine alte Schallplatte, auf der ein von Fred Eikel 1942 aufgenommenes Interview mit einer Sprecherin des Texas Deutschen aufgezeichnet war. Diese Sprecherin erinnerte sich während des Gesprächs 2006 an die alte Schallplatte, auf der sie als neunjähriges Mädchen im Gespräch mit Eikel zu hören ist.[6] Was mit den anderen Aufnahmen von Eikel passiert ist, ist leider unbekannt, was exemplarisch zeigt, dass das Problem der fehlenden systematischen institutionellen Archivierung entstandener Sprachinseltonaufnahmen sicherlich nicht zu unterschätzen ist.

Die archivierten analogen Tonaufnahmen sowie ihre auf Papier verfassten Metadaten und andere Begleitmaterialien bilden zwar eine empirisch wertvolle Datenbasis für die Sprachinselforschung, sie sind jedoch, genau wie nur auf

6 Im Sommer 2017 fand Lars Hinrichs aus der Anglistik der University of Texas at Austin im Institut alte Tonbandbestände, die eigentlich „nur“ historische Aufnahmen des Texas Englischen sein sollten. Wie sich herausstellte, waren darunter aber auch ca. zehn Stunden historische Aufnahmen des Texasdeutschen aus den 1930er und 1940er Jahren. Leider konnte bisher nicht festgestellt werden, wer diese Aufnahmen angefertigt hatte.

Papier festgehaltene Sprachdaten wie der Deutsche Sprachatlas, aus der Benutzer-, Archivierungs- und Zugangsperspektive problematisch.[7] Erstens haben analoge Tonträger (Wachs, Platten, Magnetbänder) nur eine begrenzte Lebensdauer, je nach Qualität der Tonträger durchschnittlich 30–50 Jahre, selbst wenn diese professionell behandelt und archiviert werden.[8] Ein damit verbundenes Problem ist die Abhängigkeit von funktionierenden technischen Abspielgeräten: technische Standards ändern sich und im Laufe der technischen Entwicklung stellen Firmen die Produktion von Abspielgeräten und die Produktion von Ersatzteilen ein. Diese Entwicklung war bereits während der 1990er Jahre absehbar, wie Schüller bemerkt:

> In view of the limited life expectancy of carriers, and limited availability of replay equipment, it had become clear that the classical paradigm of museums and archives, namely to preserve the original objects or documents placed in their care, would ultimately be in vain. Long-term preservation has to concentrate on the content by extracting the signals from the original carriers, by digitising them, and by migrating these digitized contents losslessly from one IT preservation platform to the next. (Schüller 2012: 864)

Zweitens bieten analoge Sprachinseltonaufnahmen und ihre Begleitmaterialien nur eine begrenzte Möglichkeit, mit ihnen wissenschaftlich intensiv zu arbeiten. So muss man oftmals zu der Institution, an der die Sprachinseltonaufnahmen archiviert sind, reisen und dort ggf. längere Zeit verbringen. Darüber hinaus sind analoge Tonträger und deren Abspielgeräte eher mühsam zu bedienen und das wiederholte Abspielen von sensiblen Tonträgern wie Kassettenbändern kann zur starken Abnutzung und im schlimmsten Fall zu deren Unbrauchbarkeit führen. Außerdem können mehrere Wissenschaftler nicht mit denselben Tonaufnahmen gleichzeitig arbeiten, so dass gemeinschaftliche Forschungsvorhaben eher schwer zu realisieren sind.

---

7 So stellt Herrgen z. B. auch bezüglich der Zugänglichkeit und Verwendungsmöglichkeit des Deutschen Sprachatlas fest, dass eine „umfassende, auch anschauliche Version des Sprachatlasses nie publiziert werden konnte, weil die aufwendige Farbreproduktion nicht realisierbar erschien. Hier wie an mehreren anderen Stellen wirkten sich – bei einem Pionierprojekt wie dem Sprachatlas möglicherweise unvermeidbare – unglückliche methodologische Entscheidungen aus, die wiederholt, in verschiedenen Projektphasen, ohne kritischen Blick auf später sich stellende Auswertungs- und Publikationsprobleme getroffen wurden." (Herrgen 2001: 1527).

8 Ähnliche Probleme gibt es auch beim Deutschen Sprachatlas: Zum einen sind die handgezeichneten historischen Karten in ihrem Bestand gefährdet, denn die Farben (oft wurden auf ein und demselben Kartenblatt bis zu 22 Farben verwendet) beginnen zu verblassen. Zum Zweiten ist es absolut wünschenswert, die sehr anschaulichen Karten einer interessierten Öffentlichkeit zur Verfügung zu stellen. Und zum Dritten sind die wissenschaftlichen Auswertungsmöglichkeiten des Atlasses noch nicht annähernd ausgeschöpft (Herrgen & Lenz 2003).

Drittens verursachen analoge Sprachinselkorpora, wie alle Laut- und Bildarchive, erhebliche Kosten und können somit von möglichen Einsparungen oder politisch motivierten Aktionen betroffen sein. So stellt z. B. Schüller Folgendes fest:

> This often leads to dramatic situations of audiovisual collections within the realm of universities in the Western world: Their democratic and autonomous organisation is an endangering factor for the adequate or further support of audiovisual research collections, as researchers, specifically under the prevailing neo-liberal climate of our times, have an increasing tendency to optimise their short-term success at the expense of long-term strategies in the interest of the scientific community at large, including further generations to come. Additionally, such a policy implicitly takes into account that the results of present publications cannot be evaluated by researchers in the future. This would not be in line with basic scientific principles. There are several cases of audiovisual collections under the umbrella of universities whose existence is threatened. (Schüller 2008: 8)

Wenn auch die Probleme der Erstellung und langfristigen Sicherung digitaler Sprachkorpora nicht trivial sind, so lassen sich doch für diese einige Vorteile ins Feld führen. Durch den stetig sinkenden Preis von Speicherplatz lassen sich auch große Mengen von Sprachdaten verhältnismäßig günstig speichern. Ferner kann eine Vielzahl von Forschern an verschiedensten Orten parallel mit den gleichen Daten arbeiten, ohne dass vorher aufwändig Datenträger physisch vervielfältigt worden wären. Der nächste Abschnitt gibt einen Überblick über digitale Sprachinselkorpora, sowohl solche, die online über das Internet zugänglich sind, als auch solche, die „nur" vor Ort an der jeweiligen Institution genutzt werden können.

3 Digitale Sprachinselkorpora

3.1 Digitale Sprachinselkorpora am IDS

Mit dem *Archiv für gesprochenes Deutsch* verfügt das IDS Mannheim über eine bedeutende Sammlung von Sprachkorpora. In seinen Ursprüngen kann dieses zum einen in der Tradition des von Eberhard Zwirner gegründeten Deutschen Spracharchivs in die Zeit vor dem zweiten Weltkrieg zurückverfolgt werden. Dieses wurde 1971 vom IDS übernommen. Schon vor diesem Zeitpunkt hatte das IDS jedoch selbst auf die Erstellung gesprochener Sprachkorpora hingearbeitet, und auch in der Folgezeit wurden weitere Korpora erhoben (Stift & Schmidt 2014). Zum Zeitpunkt des Aufstiegs digitaler Tonaufnahmen zum Ende des 20. Jahrhunderts verfügte das IDS damit bereits über eine umfassende Sammlung von Tonaufnahmen. Im Laufe der 1990er Jahre wurde das *Deutsche*

Spracharchiv (so der damalige Name) am IDS schrittweise digitalisiert, um den Bestand der analogen Tonaufnahmen (auch von deutschen Sprachinseln) für die Zukunft zu erhalten. Ein Hauptziel war auch, durch die Verschmelzung von Ton- und Computertechnik Tonaufnahmen sehr viel schneller zugänglich zu machen, was eine einfachere Bearbeitung mit verschiedenen Instrumenten der Computertechnik ermöglicht.[9] Die Erstellung der *Datenbank Gesprochenes Deutsch* (DGD) wurde ab 1997 durch die Volkswagenstiftung finanziell gefördert und erlaubte letztendlich die komplette Überführung der digitalisierten Tonaufnahmen des Deutschen Spracharchivs in eine virtuelle, über das Internet zugängliche Version (Fiehler & Wagener 2005). Die DGD besteht aus unterschiedlichen Korpora, die sich wiederum aus unterschiedlichen Materialtypen zusammensetzen: Tonaufnahmen in unterschiedlichen Formaten (WAVE, WMA, MP3), Transkripte, die überwiegend als Fließtexte vorliegen, und Metadaten, die allgemeine Informationen über die Korpora und je nach Korpus ausführliche Informationen über die Sprecher der Interaktion und zur Situation der Aufnahme enthalten (Wagener 2005).

Mit der Digitalisierung und Archivierung analoger Tonaufnahmen seit den 1990er Jahren sowie dem Bereitstellen der Daten online folgt das IDS einem allgemeinen internationalen Trend zur digitalen Sprachdokumentation und archivierung. In Europa wurde z. B. im Jahr 2000 durch Finanzierung der Volkswagenstiftung das Projekt *Dokumentation bedrohter Sprachen* (DoBeS)[10] am Max-Planck-Institut (MPI) für Psycholinguistik in Nijmegen gegründet (Haig et al. 2012). Dieses Projekt setzt sich u. a. zum Ziel, vom Aussterben bedrohte Sprachen zu dokumentieren und die Ton- und Bildaufnahmen sowie die dazugehörigen Begleitmaterialien zu archivieren, um so Daten zu bedrohten Sprachen zu retten, und, soweit möglich, über das Internet für Forschungszwecke zur Verfügung zu stellen. So entstand im Rahmen von DoBeS in den letzten 17 Jahren nicht nur eine beträchtliche Anzahl von Sprachdokumentationsprojekten, sondern es wurde auch Software zur Aufnahme, Verarbeitung und Archivierung von Sprachaufnahmen entwickelt, die kostenfrei zur Verfügung gestellt wurde. Die so gesammelten wissenschaftlichen und technischen Erfahrungen wurden vor einigen Jahren am MPI Nijmegen in einer eigenen Abteilung (*The Language Archive* [TLA]) gebündelt, um so eine langfristige Unterstützung von Sprachdokumentations- und Archivierungsaktivitäten zu

9 Die Digitalisierung hat nicht nur die Archivierung und die Nutzung vorhandener Aufnahmen maßgeblich beeinflusst, sondern sie hat auch die linguistische Datenerhebung und -verarbeitung stark verändert. Siehe z. B. Bird & Simons (2003), Boas (2006) und Margetts & Margetts (2012).

10 http://dobes.mpi.nl (letzter Zugriff: 20. 6. 2017).

garantieren. Das am MPI angesiedelte *The Language Archive* ist genauso wie das IDS Mitglied im CLARIN-Verbund,[11] einer europäischen Forschungsinfrastruktur für Sprachressourcen und Sprachtechnologie.[12]

Gegenwärtig sind in der DGD mehrere Korpora aus Gebieten zugänglich, die als Sprachinseln des Deutschen gelten können. Umfangreich ist das Korpus zum Australiendeutsch, dessen Aufnahmen in den Jahren 1966 bis 1973 unter Leitung von Michael Clyne an der Monash University in Melbourne entstanden.[13] Ein weiterer umfangreicher Bestand liegt von Sprechern der ersten und zweiten Auswanderergeneration aus Israel vor, die im Wesentlichen in den 1990er und 2000er Jahren aufgenommen wurden. Ebenfalls existiert ein Korpus der ehemaligen deutschen Ostgebiete inklusive der Sprachinseln in Ost- und Südosteuropa, die zwischen 1962 und 1965 aufgenommen wurden, wobei dies überwiegend in Deutschland geschah. Im hauptsächlich zwischen 1955 und 1961 erhobenen Zwirner-Korpus, das den Versuch einer umfassenden Dokumentation der deutschen Dialekte darstellt, sind weitere Aufnahmen von Flüchtlingen, Vertriebenen und Übersiedlern aus diesen Gebieten gesammelt. Zuletzt findet sich im *Korpus Binnen- und Auslandsdeutsche Mundarten: Varia* eine Vielzahl von Aufnahmen, die in den 1960er und 1970er Jahren gemacht wurden. Neben den Aufnahmen aus dem deutschen Kerngebiet sind hier Sprecher aus Australien, Kanada, Mexiko und den USA dokumentiert.

Hier scheint es angebracht, einige Zahlen aufzuführen, um zum einen den Umfang dieser Korpora zu unterstreichen, gleichzeitig aber auch auf die Lücken hinzuweisen, die noch bestehen. Das Korpus *Australiendeutsch* kann mit 220 Aufnahmen mit einer Länge von mehr als insgesamt 64 Stunden als durchaus umfangreich gelten. Diese stammen aus 38 verschiedenen Orten in den Bundesstaaten South Australia und Victoria, wobei die Zahl der Aufnahmen pro Ort zwischen einer einzigen und 29 schwanken. Für 168 dieser 220 Aufnahmen ist in der DGD auch ein Transkript vorhanden. Das Korpus *Binnen- und Auslandsdeutsche Mundarten: Varia* ist dagegen deutlich stärker gestreut. Insgesamt enthält es zur Zeit 183 Aufnahmen mit mehr als 69 Stunden Länge. Von diesen stammen 3 aus Australien (sämtlich South Australia), 50 aus Deutschland, 26 aus Kanada (British Columbia 24, Ontario 2), 4 aus

11 CLARIN – *Common Language Resources and Infrastructure*, http://www.clarin.eu (letzter Zugriff: 20. 6. 2017).

12 Ähnliche Konsortien, die sich vorwiegend in Nordamerika mit dem Aufbau einer technischen-linguistischen Infrastruktur beschäftigen sind E-MELD – *Electronic Metastructure for Endangered Languages Data* (http://emeld.org [letzter Zugriff: 20. 6. 2017]) und OLAC – *Open Languages Archives Community* (http://www.language-archives.org/ [letzter Zugriff: 20. 6. 2017]).

13 Hier ist erwähnenswert, dass das IDS nicht sämtliche an der Monash University entstandenen Aufnahmen erhalten hat.

Mexiko (sämtlich Chihuahua), 12 aus Österreich, 12 aus der Schweiz sowie weitere 76 aus den USA (Illinois 9, Wisconsin 67). Insgesamt sind dies also 109 Aufnahmen aus deutschen Sprachinseln, Transkripte sind hier jedoch nicht vorhanden. Von diesen hat die überwiegende Mehrheit der Sprecher primär Englisch als Kontaktsprache, innerhalb dieser Gruppe sind jedoch mehr als 80 % aus nur zwei Regionen. Bezieht man das Korpus *Australiendeutsch* mit ein, wächst dieses Übergewicht der in einen englischen Kontext eingebetteten Sprachinseln zusätzlich. Für sprachwissenschaftliche Arbeit, die auf genau dieses Kontaktszenario abzielt, bietet die DGD damit eine wichtige Grundlage für empirische Arbeit. Sie ist jedoch räumlich eingeschränkt und wird auf sich allein gestellt einer vergleichenden Sprachinselforschung, wie sie später in diesem Beitrag diskutiert wird, noch nicht gerecht.

Neben diesen Aufnahmen besitzt das IDS weitere Daten, die jedoch bislang nicht öffentlich zugänglich sind. Dies sind etwa die Daten des Siebenbürgisch-Sächsischen Schallarchivs, welche jedoch auch über das Projekt Audioatlas Siebenbürgisch-Sächsischer Dialekte (ASD) an der Ludwig-Maximilians-Universität München verfügbar sind (siehe Abschnitt 3.2). Vorhanden aber nicht zugänglich sind dagegen Aufnahmen aus Panambi im brasilianischen Bundesstaat Rio Grande do Sul (vgl. Bärnert-Fürst 1994). Sowohl an diesen Daten aus Brasilien als auch an den von in die DGD integrierten Daten aus Australien lassen sich wichtige rechtliche Aspekte illustrieren, da diese die Nutzbarkeit der Daten grundlegend beeinflussen (mehr Details dazu in Gorisch, Schmidt & Stift i. V.). Für das IDS ist das Bundesdatenschutzgesetz (BDGS) verbindlich. Unter § 40 (3) heißt es:

> Die wissenschaftliche Forschung betreibenden Stellen dürfen personenbezogene Daten nur veröffentlichen, wenn
> 1. der Betroffene eingewilligt hat oder
> 2. dies für die Darstellung von Forschungsergebnissen über Ereignisse der Zeitgeschichte unerlässlich ist.[14]

Ob die Veröffentlichung von Aufnahmen aus deutschen Sprachinseln im Sinne von Satz 2 für die Darstellung von Forschungsergebnissen über Ereignisse der Zeitgeschichte unerlässlich ist, ist eine Frage, die sich nicht ohne weiteres bejahen lässt. Eindeutiger ist dagegen die Einwilligung der Sprecher. Im Falle der Daten aus Australien wurde damals das Einverständnis der Sprecher mündlich eingeholt und dies in der Regel aufgenommen. Neben diesem Regelfall gibt es jedoch auch Ausnahmen, in denen dieses Einverständnis nicht auf den Aufnahmen erscheint. Für die Veröffentlichung dieser Aufnahmen besteht damit

14 http://www.gesetze-im-internet.de/bdsg_1990/__40.html (letzter Zugriff: 19. 5. 2017).

keine hundertprozentige rechtliche Sicherheit, so dass im IDS davon abgesehen wird. In verwandter (doch anderer) Weise ist Datenschutz auch für die Daten aus Panambi ein Grund, der der Veröffentlichung bislang im Wege stand. In das aufgenommene Deutsch der Sprecher mischt sich auch oft Portugiesisch, welches in gleicher Weise anonymisiert werden muss wie die in Deutsch aufgenommenen Teile des Gesprächs. Mit entsprechenden Kapazitäten ist das IDS jedoch derzeit nicht ausgestattet, so dass andere Sprachinseldaten in der Priorität vor diese getreten sind.

Dieser kurze Einblick in den Stand der Archivierung und Digitalisierung von Sprachinselaufnahmen zeigt, dass das IDS eine zentrale Rolle in der Verortung und Verfügbarmachung entsprechender Sprachinselkorpora innehat. Er zeigt aber auch gleichzeitig, dass die Archivierung und Bereitstellung von Sprachinselaufnahmen im Internet auch in Zukunft weiterer Arbeit bedarf. Dass seit der Umfrage von Wagener & Bausch (1997) die Aufnahmen der Varietäten an bereits dokumentierten Orten fortgeführt oder an weiteren Orten neu begonnen wurden, erhärtet letzteren Befund. Gerade bei den Dokumentationsprojekten der letzten Jahre sind das IDS und seine Mitarbeiter jedoch sehr involviert. So leisten sie wichtige Unterstützung bei der Planung, Konzeption und Realisierung von digitalen Korpora, die von Forschungsprojekten auf der Basis aktueller linguistischer Feldforschungen zusammengestellt und später der DGD zugeführt werden. Besonders sei hier die zentrale Rolle von Thomas Schmidt genannt, der in den letzten Jahren in seiner Rolle als Leiter des Programmbereichs „Mündliche Korpora“ am IDS nicht nur den Ausbau und die technische Modernisierung der DGD vorangetrieben hat, sondern auch gleichzeitig mehrere Forschungsprojekte wie Unserdeutsch (Peter Maitz, Universität Augsburg), und Namibiadeutsch (Horst Simon, FU Berlin, und Heike Wiese, Universität Potsdam) bei der Planung und Zusammenstellung ihrer digitalen Korpora technisch beraten und unterstützt hat.

3.2 Digitale Sprachinselkorpora außerhalb des IDS

Auch an anderen Institutionen gibt es digitale Sprachinselkorpora, wobei sich grundsätzlich drei Kategorien unterscheiden lassen. In die erste Kategorie fallen digitale Sprachinselkorpora, die auf bereits vorhandenen analogen Aufnahmen beruhen, welche digitalisiert worden sind und primär vor Ort zugänglich sind.

Ein Beispiel ist das *North American German Dialect Archive*, das im Max Kade Institute for German-American Studies an der University of Wisconsin-Madison beheimatet ist. Dort sind tausende Stunden von Aufnahmen deutscher Einwandererdialekte aus den USA vorhanden, die seit den 1940ern auf-

genommen wurden.[15] Die dort vorhandenen Daten stammen teilweise von an der University of Wisconsin ansässigen Forschern wie etwa Lester W. J. („Smoky") Seifert oder Jürgen Eichhoff, teilweise jedoch auch von Forschern an anderen Institutionen, so etwa die von Glenn Gilbert in den 1960er Jahren an der University of Texas erhobenen Aufnahmen des Texasdeutschen. Diese sind zwar mittlerweile digitalisiert worden, bislang ist jedoch nur ein Bruchteil der Bestände online in kurzen Ausschnitten zugänglich, und auch Archivlisten sind nicht online, so dass der Zugang nur umständlich möglich ist.[16] Eine Überarbeitung des Portals ist gegenwärtig in Planung, im Zuge derer auch detailliertere Informationen zu den Archivalien verfügbar gemacht werden sollen. Zusätzlich ist vorgesehen, zumindest Teile der Bestände Forschern auch aus der Ferne zugänglich zu machen.

Die zweite Kategorie umfasst online verfügbare digitale Sprachinselkorpora, die aus einer spezifischen Sorte linguistischer Daten aus einem bestimmten Gebiet bestehen. Ein Beispiel ist der *Linguistic Atlas of Kansas German*,[17] der die Sprachvariation unterschiedlicher deutscher Dialekte in Kansas mit Hilfe von Wenker-Sätzen erfasst. Nutzer der Online-Version können über eine Landkarte von Kansas (siehe den Ausschnitt aus der Webseite mit der interaktiven anklickbaren Karte von Kansas in Abb. 7.1) auf Tonaufnahmen von Wenker-Sätzen aus elf Landkreisen in Kansas zugreifen, insgesamt handelt es sich um Aufnahmen mit 21 Sprechern. Die Aufnahmen können auf der Webseite des *Linguistic Atlas of Kansas German* mit Hilfe eines eingebetteten Audio-Players abgespielt werden.[18]

Ein weiteres Beispiel ist die digitalisierte Version von Gesprächen mit ca. 20 Dialektsprechern wolgadeutscher Herkunft in North Dakota.[19] Diese Gespräche wurden in den 1970er Jahren von Allen Spiker aufgenommen, in den letzten paar Jahren digitalisiert und dann an der North Dakota State University online gestellt. Insgesamt handelt es sich um ca. 12 Stunden digitaler Tonaufnahmen, die man entweder mit einem eingebetteten Audio-Player hören oder als Datei im MP3-Format herunterladen kann.

---

15 https://mki.wisc.edu/content/north-american-german-dialect-archive (letzter Zugriff: 26. 5. 2017).

16 http://csumc.wisc.edu/AmericanLanguages/german/germ_us.htm (letzter Zugriff: 27. 5. 2017).

17 http://www2.ku.edu/~germanic/LAKGD/Atlas_Intro.shtml (letzter Zugriff: 20. 6. 2017).

18 Nicht alle vorhandenen Aufnahmen des Kansasdeutschen sind online gestellt. Es gibt insgesamt 136 Aufnahmen (zwischen 15–45 Minuten lang), die z. T. noch auf Kassetten sind (Chris Johnson, pers. Mitt.).

19 https://library.ndsu.edu/ndsuarchives/allen-spiker-german-russian-dialect-tapes (letzter Zugriff: 20. 6. 2017).

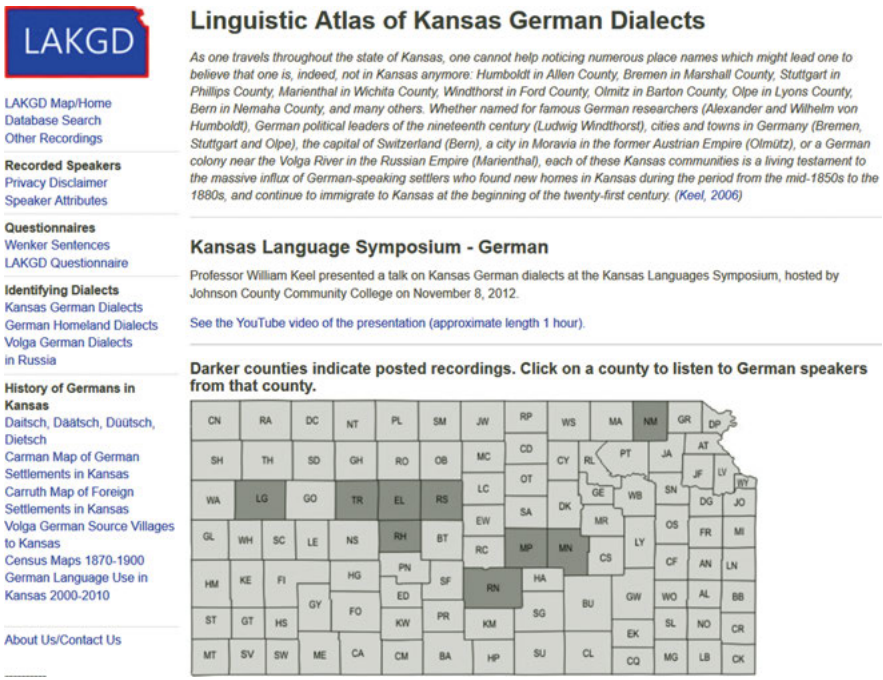


Abb. 7.1: Linguistic Atlas of Kansas German Dialects. Karteninterface zur Abfrage von Tondaten.

Als drittes Beispiel lässt sich das Korpus russlanddeutscher Sprecher in Sibirien an der Universität Göteborg nennen.[20] Die bisher online verfügbaren Tonaufnahmen stammen von Gesprächen mit vier Sprecherinnen, wobei jedes Gespräch zwischen 60–90 Minuten dauert.[21] Zu den Tonaufnahmen gibt es außerdem Transkripte, die knapp 4.000 Sätze umfassen (siehe auch Andersen 2016).[22]

20 https://spraakbanken.gu.se/korp/?mode=siberian_german#?stats_reduce=word&cqp=%5B%5D (letzter Zugriff: 20. 6. 2017). Zum Russlanddeutschen siehe auch Behrend (2003).

21 An der Universität Krasnoyarsk, Russland, gibt es außerdem ein regionales Forschungszentrum zur Erforschung und Dokumentation der Sprache und Kultur der Deutschen in Sibirien. Dort gibt es auch einige Tonaufnahmen online (http://deu.kspu.ru/ru/about-us.html [letzter Zugriff: 23. 6. 2017]).

22 Momentan wird auch an der Universität Erfurt von Csaba Földes das Projekt „Ungarndeutsches Zweisprachigkeits- und Sprachkontaktkorpus“ aufgebaut (https://www.ungarndeutsch.de/information [letzter Zugriff: 20. 6. 2017]), welches auch bald Tonaufnahmen und Transkripte über das Internet anbieten wird (Földes 2016).

Als letztes Beispiel sei nochmals der an der Ludwig-Maximilians-Universität München angesiedelte Audioatlas Siebenbürgisch-Sächsischer Dialekte genannt, der auf digitalisierten Tonaufnahmen aus den 1960er und 1970er Jahren basiert (Krefeld, Lücke & Mages 2016).[23] Die ca. 360 Stunden umfassenden Tonaufnahmen bestehen aus Wenker-Sätzen, Erzählungen, Märchen und Liedern, zu vielen der Aufnahmen gibt es auch Transkripte. Der Audioatlas kann mit Hilfeunterschiedlicher Abfragemasken durchsucht werden (Datenbankabfrage, Karten, Vergleich von Wenker-Sätzen etc.).

Die dritte Kategorie digitaler Sprachinselkorpora umfasst solche Korpora, die das direkte Resultat von aktuellen Forschungsprojekten sind. Solche Sprachinselprojekte und -korpora haben sich zum Ziel gesetzt, aktuell erhobene Sprachinselaufnahmen im Rahmen der Forschungstätigkeit zeitnah zu bearbeiten, zu transkribieren und zu archivieren. Ein Beispiel ist das im Jahre 2001 an der University of Texas at Austin gegründete *Texas German Dialect Project* (TGDP) (Boas 2002; Boas et al. 2010), das sich zum Ziel gesetzt hat, die letzten der noch ca. 6.000 verbliebenen Sprecher des Texasdeutschen aufzunehmen und die Aufnahmen zu archivieren.[24] Die vom TGDP erhobenen Daten gliedern sich in drei Teile. Zum einen werden in einem Interview systematisch biographische Informationen über Person und Leben der Sprecher sowie ihren Gebrauch der deutschen Sprache und ihre Spracheinstellungen erhoben (siehe Boas 2005, Boas & Fingerhuth 2017). Die so erhobenen Sprecherdaten bilden die Grundlage für die im *Texas German Dialect Archive* archivierten Tonaufnahmen, d. h. jede Tonaufnahme kann durch eine einzigartige Identifikationsnummer nur einem bestimmten Satz an Metadaten zugeordnet werden (siehe Boas 2006 und Abb. 7.2). Die von Eikel (1954) und Gilbert (1972) benutzten Übersetzungslisten werden vom TGDP wieder verwendet, um so einen kontrollierten Datensatz zu erstellen und um auch gleichzeitig feststellen zu können, wie sich das Texasdeutsche in den letzten 50–60 Jahren verändert hat. Als dritten Ansatz gibt es 30–60 Minuten lange soziolinguistische Interviews mit den Sprechern des Texasdeutschen, die sich, je nach Lage und Interesse um unterschiedliche Themen drehen (Vorfahren, Wetter, Politik, persönliche Interessen etc.). Außerdem wird teilweise ein Gespräch in Abwesenheit der Forscher nur unter Sprechern von Texasdeutsch aufgenommen, etwa beim Verrichten von Alltagsarbeiten oder beim Kartenspiel. In Summe kann die Erhebung sämtlicher Daten eines einzigen Sprechers ohne weiteres drei Stunden in Anspruch nehmen. Dies bietet grundsätzlich eine breite Datengrundlage für sprachwissenschaftliche Forschung. In der Praxis erweist es sich jedoch teilweise als

23 http://www.asd.gwi.uni-muenchen.de/ (letzter Zugriff: 20. 6. 2017).
24 http://www.tgdp.org (letzter Zugriff: 20. 6. 2017).

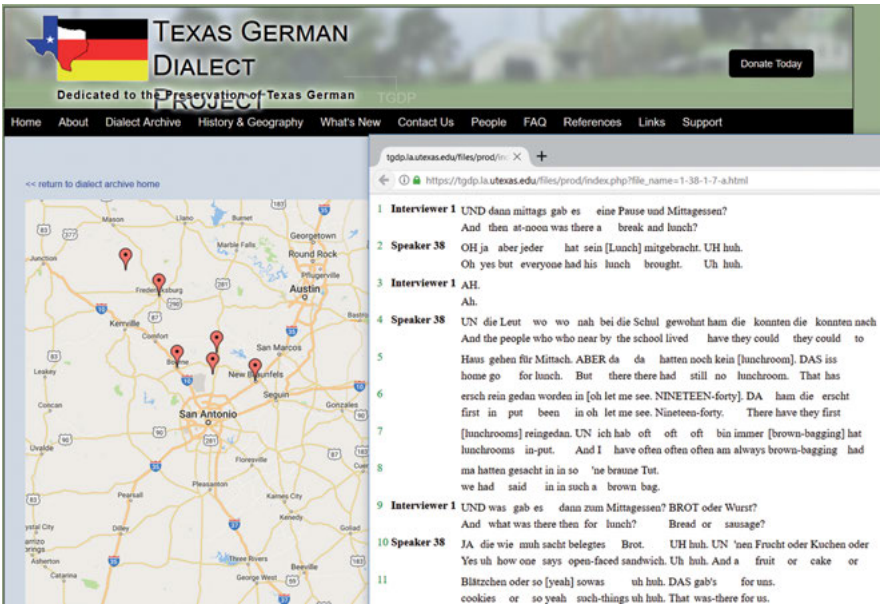


Abb. 7.2: Das Texas German Dialect Archive. Ausschnitt aus der Webseite (screen shot), mit Anzeige der transkribierten Interviews.

Problem, da z. B. das Interesse der Sprecher an der Teilnahme schwindet, und dadurch nur ein Teil der Daten erhoben werden kann. Von 2001–2018 hat das TGDP mehr als 660 Sprecher des Texas-Deutschen aufgenommen. Der Plan ist bis 2030 die Grenze von tausend Sprechern zu überschreiten, falls es bis dahin noch genügend Sprecher geben sollte (siehe Boas 2009).

Die vom TGDP digital aufgenommenen Eikel- und Gilbert-Sprachdaten werden nach den Interviews segmentiert und im *Texas German Dialect Archive* (TGDA) archiviert. Die soziolinguistischen Interviews werden mit ELAN (Wittenburg et al. 2006) transkribiert,[25] ins Englische übersetzt und ebenfalls im TGDA archiviert (für Details siehe Boas 2003/2006). Nutzer des Archivs können auf alle drei Sorten von Daten zugreifen, entweder über eine Datenbankabfrage, über ein geographisches Suchinterface mit Karten, auf denen die Erhebungsorte verzeichnet sind, oder über die Wort- und Satzlisten der Gilbert- und Eikel-Fragebögen. Die im TGDA archivierten Daten (Tonaufnahmen, Transkripte, Metadaten) sind nach Anmeldung und Einverständniserklärung mit den Nutzungsbedingungen online abrufbar entweder als MP3 im eingebetteten

25 https://tla.mpi.nl/tools/tla-tools/elan/ (letzter Zugriff: 20. 6. 2017).

Audio Player zusammen mit den Transkripten im HTML-Format oder zum Herunterladen als WAVE- (Ton) und EAF- (Text) Format zur Analyse und weiteren Bearbeitung mit ELAN (für Details, siehe Boas et al. 2010). Von den insgesamt mehr als 1.000 Stunden Tonaufnahmen sind bisher nur ca. 100 Stunden transkribiert und übersetzt.

Die Konzeption und technische Realisierung des Arbeitsablaufs des TGDP sowie die daraus resultierende Archivierung und Bereitstellung von Sprachaufnahmen und begleitenden Materialien im Internet wurde in den vergangenen Jahren von anderen Forschungsprojekten aufgegriffen und als Grundlage für deren eigene Aufnahme- und Archivierungsaktivitäten genutzt. So haben das *Indiana German Dialect Project* (Karen Roesch, Indiana University Purdue University Indianapolis), das *Wisconsin Low German Dialect Project* (Ryan Dux, Bucknell University), und das *Barossa-Deutsch Projekt* (Claudia Riehl, Ludwig-Maximilians-Universität München) (Riehl 2012) die vom TGDP verwendeten Fragebögen benutzt, um selbst vergleichbare Daten zu erheben. Die von diesen Forschergruppen erhobenen Daten sind nicht nur für sich alleine genommen interessant, sondern sie erlauben es auch zu vergleichen, was passiert, wenn unterschiedliche deutsche Auswandererdialekte in Kontakt mit unterschiedlichen Varietäten des Englischen stehen. Wie der folgende Abschnitt zeigt, werden momentan die technischen Voraussetzungen dafür geschaffen, dass Daten in digitalen Sprachinselkorpora systematisch miteinander verglichen werden können.

4 Aufgaben für die Zukunft: Lokalisierung, Digitalisierung und eine digitale vergleichende Sprachinselforschung

Die Sprachinselforschung hat sich im 20. Jahrhundert weitgehend dezentral entwickelt – ein Umstand, der sich vielleicht aus der Verteilung ihres Untersuchungsgegenstands und auch der Forschenden über sämtliche Kontinente erklären lässt. In dieser Diaspora ist daher eine Vielzahl von Korpora entstanden, zu deren Umfang und Status sich verlässliche Angaben nur teilweise machen lassen. Das oben besprochene Korpus Australiendeutsch in der DGD zeigt einen Fall, in dem historische Sprachinseltonaufnahmen digitalisiert und der Öffentlichkeit über das Internet zugänglich gemacht wurden. Die Digitalisierung und Aufnahme der Australiendeutschdaten in die DGD ist ein Glücksfall für die Forschung, da diese Daten sonst wahrscheinlich nicht langfristig gesichert gewesen wären, geschweige denn einem breiteren Forschungspubli-

kum zur Verfügung gestanden hätten. Doch leider ist dieses Vorgehen eher die Ausnahme.

In vielen Fällen scheinen die Sprachinseldaten (Tonaufnahmen, Transkripte, Begleitmaterialien etc.) weiterhin ausschließlich am Ort der ursprünglichen Erhebung oder bei der Person, die sie erhoben hat, vorzuliegen. Der Umfang dieser Daten dürfte beachtlich sein, und sie sind, wenn überhaupt, nur vor Ort abzufragen, wobei es gerade bei älteren Magnetbändern nicht klar ist, wie lang diese noch haltbar sind. Auf Grundlage der von Wagener und Bausch durchgeführten Befragung lassen sich als Beispiele hierfür etwa unter Leitung von Evelyn S. Firchow in Minnesota geführte Interviews nennen, wo zum Zeitpunkt der Erhebung Anfang der 1990er Jahre bereits ca. 150 Interviews vorlagen, jedoch auch weitere Aufnahmen geplant waren (Wagener & Bausch 1997: 223, vgl. auch Firchow 1991). Im rumänischen *Arhiva de Folclor a Academiei Române* (sowie als Kopie an der siebenbürgisch-säschsischen Forschungsstelle in Gundelsheim – seit 2003 als Siebenbürgen-Institut ein An-Institut der Universität Heidelberg – dort jedoch nicht zugänglich) liegen nach Wagener & Bausch (1997: 208) 85 Magnetophonbänder mit Aufnahmen aus Siebenbürgen. Es ließen sich weitere Beispiele anführen, doch sollten allein diese beiden genügen, um zu illustrieren, dass eine beträchtliche Menge von Sprachdaten existiert, jedoch derzeit nicht oder nur schwer zugänglich ist. Ferner ist unsicher, ob und wie eine Archivierung und Sicherung der auf analogen Datenträgern erhobenen Daten in den unterschiedlichen Institutionen durchgeführt wird, geschweige denn ob Pläne vorhanden sind, diese einer breiteren Öffentlichkeit über das Internet zur Verfügung zu stellen.

Hier ist zu erwähnen, dass die von Wagener und Bausch durchgeführte Erhebung, die wohl als letzte und einzige ihrer Art gelten kann, einige in diesem Artikel bereits erwähnte Bestände nicht erfasst. Die Dokumentation der existierenden Korpora ist also lückenhaft. Ferner gibt es Aufnahmen, die derzeit als verschollen gelten müssen, wie etwa Eikels Aufnahmen des Texasdeutschen aus den 1940er Jahren. Eine entsprechende Übersicht existierender Sprachinseltonaufnahmen, die von der Dokumentation durch Wagener und Bausch ausgeht, findet sich im Anhang.

Für die Gegenwart stellen sich also die Aufgaben, die existierenden Aufnahmen zu erfassen, zu lokalisieren, zu digitalisieren, zu archivieren, sie für die Nutzung aufzubereiten und sie letztlich der Fachöffentlichkeit über das Internet zur Verfügung zu stellen. Dadurch ließen sich nicht nur die Früchte vergangener Arbeit sichern. Vielmehr bietet sich dadurch die Möglichkeit, der Sprachinselforschung die Daten und Infrastruktur zur Verfügung zu stellen, die die Erschließung eines Forschungsparadigmas ermöglicht, das bereits von Rosenberg (2003, 2005) angedacht wurde, das aber bislang noch nicht fest

etabliert ist: einer vergleichenden Sprachinselforschung. Boas (2016) skizziert, wie eine vergleichende Sprachinseldatenbank aussehen könnte. Konkret wird vorgeschlagen,

> [...] auf der Basis des TGDA eine erweiterte Sprachinseldatenbank aufzubauen, in der Daten von anderen Sprachinseln parallel korpuslinguistisch verarbeitet und archiviert werden. Der Umfang, die Qualität und die Art von bereits existierenden Sprachinseldaten sind natürlich recht unterschiedlich. [...] Aber alle bereits existierenden Daten könnten prinzipiell mit denselben im Rahmen des TGDP erprobten sowie evtl. noch zu erforschenden zusätzlichen Methoden korpuslinguistisch verarbeitet werden, um so in separaten Sprachinselarchiven archiviert zu werden. Diese Sprachinselarchive könnten dann miteinander vernetzt werden, um eine vergleichende Onlinesprachinseldatenbank zu implementieren. Ziel einer solchen vergleichenden Sprachinseldatenbank ist nicht nur die Archivierung existierender Sprachinseldaten für künftige Generationen, sondern auch die Bereitstellung der Sprachinseldaten für die vergleichende Sprachinselforschung. (Boas 2016: 39)

Die im Jahr 2016 durchgeführte Erneuerung der technischen Infrastruktur des TGDA an der University of Texas at Austin hat dazu geführt, dass die zugrunde liegende Datenbank nun auch Daten aus anderen Sprachinseln aufnehmen kann. Bei den Sprachinseldokumentationsprojekten, die sich an das TGDP anlehnen und vergleichbare Fragebögen und Einwilligungserklärungen benutzen, ist die Aufnahme von neuen Sprachinseldaten relativ einfach. Die ersten Schritte, Sprachinseldaten aus Wisconsin, Indiana und Australien in das vergleichende Sprachinselarchiv aufzunehmen, waren erfolgreich. Diese werden im Laufe des Jahres 2019 auch für die Öffentlichkeit freigeschaltet.

Etwas anders sieht die Lage bei älteren Sprachinselaufnahmen aus, bei denen es keine Einwilligungserklärungen gibt. Die Probleme bei der Zugänglichmachung von Daten in Deutschland wurden bereits im Kontext der über das IDS verfügbaren Korpora diskutiert. Aufgrund der unterschiedlichen Gesetzeslage gestaltet sich die Praxis in den USA anders. Forschung mit menschlichen Teilnehmern muss grundsätzlich von einem sogenannten *Institutional Review Board* (IRB) genehmigt werden.[26] Diese Genehmigung hängt neben grundsätzlichen ethischen Kriterien nicht zuletzt vom Einverständnis der Teilnehmer ab. Diese Genehmigungspflicht hat sich jedoch erst seit den 1970ern entwickelt. Abseits der Forschung sind jedoch persönliche Daten als solche in den USA im Gegensatz zu etwa Deutschland nicht in gleicher Weise geschützt. Anwenden lassen sich jedoch die Regularien des sogenannten *Fair Use* aus

26 Details zu den Begriffen „Forschung“ und „menschlicher Teilnehmer“ finden sich etwa unter https://research.utexas.edu/ors/human-subjects/what-is-human-subjects-research/ (letzter Zugriff: 18. 6. 2017).

dem Bereich des Urheberrechts. Nach diesen lassen sich jegliche Materialien für Zwecke wie Forschung oder Bildung verwenden, wenn Umfang und Art der Nutzung dies rechtfertigen.

Die Kriterien für eine solche Nutzung sind nicht rigide definiert, sondern verlangen eine Abschätzung von Seiten des Nutzers, und die Verwendung alter Aufnahmen für Forschungszwecke lässt sich vor diesem Hintergrund rechtfertigen. Eine Verpflichtung zur Wahrung der Persönlichkeitsrechte der Sprecher, wie sie bei der Bewilligung neuer Forschung durch das IRB gefordert würde, bleibt dabei jedoch weiterhin bestehen. Auch alte Sprachaufnahmen können deshalb nicht ohne weiteres veröffentlicht werden, sondern müssen soweit wie möglich anonymisiert werden. Neben der Digitalisierung analoger Sprachinselaufnahmen sowie der Anfertigung von Transkripten stellt die Anonymisierung daher einen recht zeitaufwendigen Arbeitsschritt dar. Im ersten Halbjahr 2017 wurde die erste Testphase zur Erschließung und Einbindung bereits existierender historischer Sprachinselaufnahmen aus Pennsylvania und Brasilien erfolgreich abgeschlossen. Diese Daten werden auch im Verlauf des Jahres 2019 der Öffentlichkeit im vergleichenden Sprachinselarchiv der University of Texas at Austin zur Verfügung gestellt.

Die Zusammenführung unterschiedlicher Sprachinselkorpora unter einem Dach hat den Vorteil, dass man viele linguistische Phänomene besser und intensiver untersuchen kann,

> [...] da eine vergleichende Sprachinseldatenbank es erlaubt, nach Daten zu bestimmten sprachlichen Phänomenen zu suchen und diese dann systematisch zwischen Sprachinseln zu vergleichen. So könnte der Nutzer einer vergleichenden Sprachinseldatenbank z. B. nach bestimmten Präpositionen suchen, um zu sehen, welche Kasus diese in unterschiedlichen Sprachinselvarietäten markieren. Ein vergleichendes Sprachinselarchiv hätte nicht nur den Vorteil, dass bereits existierende Daten gesichert und archiviert würden und so der Nachwelt erhalten blieben, sondern es würde endlich auch eine wirklich systematische vergleichende Sprachinselforschung ermöglichen. Forscher, die ihre Sprachinseldaten in einem solchen Archiv deponieren würden, erhielten außerdem die Möglichkeit, die Früchte ihrer langjährigen Arbeit zu teilen. Entsprechende Richtlinien stellen außerdem sicher, dass die Forscher, die ihre Daten anderen Kollegen zur Verfügung gestellt haben, durch entsprechende Zitate ihrer Daten Anerkennung erhalten. (Boas 2016: 39–40)

Mit diesem Schritt ließe sich die Sprachinselforschung in ihrer Breite sinnbildlich aus dem analogen ins digitale Zeitalter, aus dem 20. Jahrhundert in die Gegenwart überführen und die Grundlage für zukünftige Arbeit schaffen. Während die bisherige Forschung mit ihrem Fokus auf einzelne Varietäten eine unerlässliche Grundlage gelegt hat, bietet der vergleichende Ansatz neue Fragen und Erkenntnisse, die auf Muster jenseits der einzelnen Varietäten oder Kontaktsprachen hinzielen. Weiter könnte dazu beigetragen werden,

das Potenzial der Sprachinselforschung als Schnittstelle zwischen der Sprachwissenschaft des Deutschen und Nachbardisziplinen zu verwirklichen. Sprachinselforschung ist fast notwendig Sprachkontaktforschung, die Perspektive der Kontaktsprache ist jedoch bislang weitgehend unbeachtet geblieben.

5 Zusammenfassung und Ausblick

Das Gesagte hinterlässt ein durchwachsenes Bild vom gegenwärtigen Stand der Sprachinselkorpora des Deutschen, erlaubt aber gleichzeitig einen positiven Ausblick. Angelegt in ihrem zergliederten Forschungsgegenstand ist die Sprachinselforschung weitgehend unkoordiniert entstanden. Die erhobenen Sprachkorpora sind diesem Umstand entsprechend ebenfalls zerstreut. Um das Potenzial der in einem Zeitraum von mehr als einem Jahrhundert gesammelten Daten zu nutzen, stellen sich in der Gegenwart die Aufgaben, die existierenden Aufnahmen zu lokalisieren, zu digitalisieren, und innerhalb der gesetzlich gesteckten Rahmen zugänglich zu machen.[27]

Es gibt jedoch bereits Ansätze zur Lösung dieser Aufgabe. Dem IDS kommt dabei eine zentrale Rolle zu, und die Übernahme der von Michael Clyne in Australien gemachten Aufnahmen sowie deren Veröffentlichung in der DGD können als ein Beispiel dafür angeführt werden, wie ein analog erhobenes Korpus von beachtlichem Umfang in ein digitales und über das Internet weltweit zugängliches umgewandelt wurde. Die Bedeutung des IDS wächst zusätzlich durch seine Betreuung bei der Planung von neuen Erhebungen, wie etwa im Fall der gegenwärtigen Arbeit am Namibiadeutschen oder am Unserdeutsch in Papua Neu Guinea und Australien.

Außerhalb des IDS gibt es ferner mit dem vergleichenden Sprachinselarchiv an der University of Texas at Austin Bemühungen, eine Infrastruktur aufzubauen, die gezielt auf die Bereitstellung von Sprachinselkorpora hin ausgerichtet ist. Dieses Sprachinselarchiv besitzt das Potenzial, die Sprachinselforschung dabei zu unterstützen, über ihren bisherigen Forschungsansatz hinauszuwachsen, der im Wesentlichen auf einzelne Varietäten beschränkt ist. Auf Grundlage zentralisiert zugänglicher Korpora könnte so eine systematisch

27 Ohne vom Gegenstand der Korpora des gesprochenen Deutschen abweichen zu wollen, scheint es doch angemessen, an dieser Stelle auf Lücken in der Erforschung des geschriebenen Deutsch der Sprachinseln hinzuweisen. Dieses ist in der Forschung bisher nur vereinzelt genutzt worden, etwa in Form von Auswandererbriefen (etwa Elspaß 2005), aber auch in der Sprachinselpresse (Földes 2015). Das weitere Erschließen solcher schriftsprachlicher Ressourcen, gegenwärtig wie auch historisch, stellt eine weitere Aufgabe für die Sprachinselforschung dar.

vergleichende Sprachinselforschung entstehen, die auf Grundlage historischer und neu erhobener Korpora mit neuen Fragen Forschung betreibt. Sie könnten Brücken zu anderen Disziplinen schlagen und zur Verankerung der Sprachinselforschung jenseits ihrer traditionellen Heimat in der Germanistischen Linguistik beitragen.

Literatur

Andersen, Christiane (2016): Nachfeld im Kontakt. Eine Korpusuntersuchung am Russlanddeutschen in Sibirien. *Göteborger Arbeitspapiere zur Sprachwissenschaft* 6, 1–15.

Atwood, E. Bagby (1962): *The regional vocabulary of Texas*. Austin, TX: University of Texas Press.

Bärnert-Fürst, Ute (1994): Conversation and displacement processes of the German language in the speech community of Panambi, Rio Grande Do Sul, Brazil. In Nina Berend & Klaus J. Mattheier (Hrsg.), *Sprachinselforschung. Eine Gedenkschrift für Hugo Jedig*, 273–287. Frankfurt am Main: Peter Lang.

Berend, Nina (2003): Zur Vergleichbarkeit von Sprachkontakten: Erfahrungen aus wolgadeutschen Sprachinseln in den USA und Russland. In William D. Keel & Klaus J. Mattheier (Hrsg.), *German language varieties worldwide: Internal and external perspectives (Deutsche Sprachinseln weltweit: Interne und externe Perspektiven)*, 239–269. Frankfurt am Main: Peter Lang.

Bird, Steven & Gary Simons (2003): Seven dimensions of portability for language documentation and description. *Language* 79 (4), 557–582.

Boas, Hans C. (2002): The Texas German Dialect Archive as a tool for analyzing sound change. In Peter Austin, Helen A. Dry & Peter Wittenburg (Hrsg.), *Proceedings of the International Workshop on Resources and Tools in Field Linguistics held in conjunction with the Third International Conference on Language Resources and Evaluation*. Las Palmas, Spain.

Boas, Hans C. (2003): Tracing dialect death: The Texas German Dialect Project. In Julie Larson & Mary Paster (Hrsg.), *Proceedings of the 28th Meeting of the Berkeley Linguistics Society*, 387–398. Berkeley, CA: Berkeley Linguistics Society.

Boas Hans C. (2005): A dialect in search of its place: The use of Texas German in the public domain. In Craig Cravens & David Zersen (Hrsg.), *Transcontinental encounters: Central Europe meets the American heartland*, 78–102. Austin, TX: Concordia University Press.

Boas, Hans C. (2006): From the field to the web: Implementing best-practice recommendations in documentary linguistics. *Language Resources and Evaluation* 40 (2), 153–174.

Boas, Hans C. (2009): *The life and death of Texas German*. Durham: Duke University Press.

Boas, Hans C. (2016): Variation im Texasdeutschen: Implikationen für eine vergleichende Sprachinselforschung. In Alexandra Lenz (Hrsg.), *German Abroad. Perspektiven der Variationslinguistik, Sprachkontakt- und Mehrsprachigkeitsforschung*, 11–44. Wien: Vienna University Press.

Boas, Hans C. & Matthias Fingerhuth (2017): „I am proud of my language but I speak it less and less!" – Der Einfluss von Spracheinstellungen und Sprachgebrauch auf den Spracherhalt von Heritage-Sprechern des Texasdeutschen. *Linguistische Berichte* 249, 95–121.

Boas, Hans, C., Marc Pierce, Karen Roesch, Guido Halder & Hunter Weilbacher (2010): The Texas German Dialect Archive: A multimedia resource for research, teaching, and outreach. *Journal of Germanic Linguistics* 22 (3), 277–296.

Bremer, Otto (1895): *Beiträge zur Geographie der deutschen Mundarten in Form einer Kritik von Wenkers Sprachatlas des Deutschen Reichs*. Leipzig. Breitkopf & Härtl.

Brenner, Koloman, Maria Erb & Karl Manherz (2008): *Ungarndeutscher Sprachatlas*. Band 1. Budapest: ELTE Germanistisches Institut.

Eikel, Fred (1949): The use of cases in New Braunfels German. *American Speech* 24 (4), 278–281.

Eikel, Fred (1954): *The New Braunfels German dialect*. Manuskript. Johns Hopkins University.

Eikel, Fred (1966): New Braunfels German: Part I. *American Speech* 41 (1), 5–16.

Eller-Wildfeuer, Nicole (im Druck): *Sprecherbiographien und Mehrsprachigkeit. Deutschbasierte Minderheitensprachen in Osteuropa und Übersee*. Tübingen: Stauffenburg.

Elspaß, Stephan (2005): *Sprachgeschichte von unten. Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert*. Tübingen: Niemeyer.

Fiehler, Reinhard & Peter Wagener (2005): Die Datenbank Gesprochenes Deutsch (DGD) – Sammlung, Dokumentation und Untersuchung gesprochener Sprache als Aufgaben der Sprachwissenschaft. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 6, 136–147.

Firchow, Evelyn Scherabon (1991): Deutsche Sprachinseln im amerikanischen Bundesstaat Minnesota. In Yoshinori Shichiji (Hrsg.), *Begegnung mit dem „Fremden". Grenzen, Traditionen, Vergleiche; Akten des VIII. Internationalen Germanisten-Kongresses, Tokyo 1990. Band 3: Sprachgeschichte: Sprachkontakte im germanischen Sprachraum*, 252–263. München: Iudicium.

Földes, Csaba (2015): Literalität im Schnittfeld von zwei Sprachen und Kulturen. Beobachtungen anhand der Phraseologie in der Sprache der Lokalpresse. In Regula Schmidlin, Heike Behrens & Hans Bickel (Hrsg.), *Sprachgebrauch und Sprachbewusstsein. Implikationen für die Sprachtheorie*, 239–260. Berlin, Boston: de Gruyter.

Földes, Csaba (2016): Ungarndeutsche Sprachvariation und Mehrsprachigkeit. Ein Korpusprojekt auf der Basis empirischer Feldforschung und Online-Sprachdokumentation. *Sprachtheorie und Germanistische Linguistik* 26 (2), 167–190.

Gilbert, Glenn G. (1963): *The German dialect spoken in Kendall and Gillespie Counties, Texas*. PhD Dissertation, Harvard University.

Gilbert, Glenn G. (1972): *The Linguistic Atlas of Texas German*. Austin, TX: The University of Texas Press.

Goebl, Hans & Guillaume Schiltz (2006): Neuere Entwicklungen in der europäischen Dialektologie (1950–2000). In Sylvain Auroux, E. F. K. Koerner, Hans-Josef Niederehe & Kees Versteegh (Hrsg.), *Geschichte der Sprachwissenschaften. Ein internationales Handbuch zur Entwicklung der Sprachforschung von den Anfängen bis zur Gegenwart*, 3. Teilband (Handbücher zur Sprach- und Kommunikationswissenschaft 18.3), 2352–2362. Berlin, Boston: Walter de Gruyter.

Gorisch, Jan, Schmidt, Thomas & Ulf Michael Stift (i.V.): Data of German speech minorities in the Archive for Spoken German: An overview. Manuskript. IDS Mannheim.

Haig, Geoffrey, Nicole Nau, Stefan Schnell & Claudia Wegener (2012): Introduction: Documenting endangered language before, during, and after the DoBes programme. In Geoffrey Haig, Nicole Nau, Stefan Schnell & Claudia Wegener (Hrsg.), *Documenting*

endangered languages: Achievements and perspectives, 1–14. Berlin, Boston: de Gruyter.

Herrgen, Joachim (2001): Die Rolle des Wenker-Atlasses in der Geschichte der Dialektologie. In Sylvain Auroux, E. F. K. Koerner, Hans-Josef Niederehe & Kees Versteegh (Hrsg.), *Geschichte der Sprachwissenschaften. Ein internationales Handbuch zur Entwicklung der Sprachforschung von den Anfängen bis zur Gegenwart*, 2. Teilband (Handbücher zur Sprach und Kommunikationswissenschaft 18.2), 1513–1535. Berlin, Boston: de Gruyter.

Herrgen, Joachim & Alexandra Lenz (2003): Digitale Dialektologie. Online Publikation des Wenker Atlasses im Internet. *Marburger UniJournal* 14, 43–48.

John, Eckhard & Natalia D. Swetosarowa (Hrsg.) (2005): *Traditionelle Lieder der Russlanddeutschen. Die Volksliedsammlung von Viktor M. Schirmunski. Ein Quellenhandbuch*. Münster: Waxmann.

Klein, Karl Kurt & Ludwig Erich Schmitt (Hrsg.) (1961/1965): *Siebenbürgisch-Deutscher Sprachatlas*. 2 Bände. Marburg: Elwert.

Knetschke, Edeltraud & Margret Sperlbaum (1983): *Das Deutsche Spracharchiv im Institut für Deutsche Sprache* (Mitteilungen des Instituts für Deutsche Sprache 6). 2. Aufl. Mannheim: Institut für Deutsche Sprache.

Knoop, Ulrich, Wolfgang Putschke & Herbert-Ernst Wiegand (1982): Die Marburger Schule: Entstehung und frühe Entwicklung der Dialektgeographie. In Werner Besch, Ulrich Knoop, Wolfgang Putschke & Herbert-Ernst Wiegand (Hrsg.), *Dialektologie: Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. 1. Teilband (Handbücher zur Sprach- und Kommunikationswissenschaft 1.1), 38–92. Berlin, New York: Walter de Gruyter.

Krefeld, Thomas, Lücke, Stephan & Emma Magnes (Hrsg.) (2016): *Zwischen traditioneller Dialektologie und digitaler Geolinguistik: Der Audioatlas siebenbürgisch-sächsischer Dialekte (ASD)* (Korpus im Text 2). Münster: Monsenstein & Vannerdat.

Kuhn, Walter (1934): *Deutsche Sprachinselforschung. Geschichte, Aufgaben, Verfahren*. Plauen: Wolff.

Lessiak, Primus & Anton Pfalz (1918): Sprachproben aus den Sieben Gemeinden (Sette Comuni Vicentini), Italien. *Sitzungsberichte/Akademie der Wissenschaften in Wien, Philosophisch-Historische Klasse* 187 (1), 59–74.

Lüdeling, Anke & Merja Kytö (Hrsg.) (2008/2009): *Corpus linguistics. An international handbook*, 2 Bde. (Handbücher zur Sprach- und Kommunikationswissenschaft 29.1–2). Berlin, Boston: Mouton de Gruyter.

Margetts, Anna & Andrew Margetts. 2012. Audio and video recording techniques for linguistic research. In Nick Thieberger (Hrsg.), *The Oxford handbook of linguistic fieldwork*, 13–53. Oxford: Oxford University Press.

Moelleken, Wolfgang Wilfried (1988): A new linguistic atlas of Pennsylvania German. *Monatshefte* 80 (1), 105–114.

Reed, Carroll & Lester W. Seifert (1954): *A linguistic atlas of Pennsylvania German*. Marburg: Becker.

Riehl, Claudia (2012): Deutsch als Reliktvarietät: Der Fall des Barossa-Deutschen (Australien). In Claudia Riehl & Elisabeth Knipf-Komlosi (Hrsg.), *Kontaktvarietäten des Deutschen in historischer und gegenwärtiger Sicht*, 37–40. Wien: edition praesens.

Rosenberg, Peter (2003): Comparative speech island research: Some results from studies in Russia and Brazil. In Willian D. Keel & Klaus J. Mattheier (Hrsg.), *German language varieties worldwide: Internal and external perspectives*, 199–238. Frankfurt am Main: Peter Lang.

Rosenberg, Peter (2005): Dialect convergence in the German language islands (*Sprachinseln*). In Peter Auer, Frans Hinskens & Paul Kerswill (Hrsg.), *Dialect change: Convergence and divergence in European languages*, 221–235. Cambridge: Cambridge University Press.

Schirmunski, Viktor M. (1962): *Deutsche Mundartkunde. Vergleichende Laut- und Formenlehre der deutschen Mundarten*. Aus dem Russischen übersetzt und wissenschaftlich bearbeitet von Wolfgang Fleischer. Berlin: Akademie Verlag.

Schmeller, Johann Andreas (1821): *Die Mundarten Bayerns grammatisch dargestellt*. München: Karl Thienemann.

Schmeller, Johann A. (1838): Ueber die sogenannten Cimbern der VII und XIII Communen auf den Venedischen Alpen und ihre Sprache. *Abhandlungen der Bayerischen Akademie der Wissenschaften, Philosophisch-Philologische und Historische Klasse*, 559–708. http://www.mdz-nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bvb:12-bsb10523345-1 (letzter Zugriff: 23. 6. 2017).

Schüller, Dietrich (2008): *Audiovisual research collections and their preservation*. Amsterdam: European Comission on Preservation and Access.

Schweizer, Bruno (1939): *Zimbrische Sprachreste. Teil 1: Texte aus Giazza (Dreizehn Gemeinden ob Verona)*. Nach dem Volksmunde aufgenommen und mit deutscher Übersetzung herausgegeben. Halle (Saale): Niemeyer.

Stift, Ulf-Michael & Thomas Schmidt (2014): Mündliche Korpora am IDS: Vom Deutschen Spracharchiv zur Datenbank für Gesprochenes Deutsch. In Melanie Steine & Franz Josef Berens (Hrsg.), *Ansichten und Einsichten. 50 Jahre Institut für Deutsche Sprache*, 360–375. Mannheim: Institut für Deutsche Sprache. http://nbn-resolving.de/urn:nbn:de:bsz:mh39-24779 (letzter Zugriff 23. 6. 2017).

Wagener, Peter (1988): *Untersuchungen zur Methodologie und Methodik der Dialektologie*. Marburg: Elwert.

Wagener, Peter (2002): Gesprochenes Deutsch online. Zur Modernisierung des Deutschen Spracharchivs. *Zeitschrift für Dialektologie und Linguistik* 3, 314–335.

Wagener, Peter (2005): Die Datenbank Gesprochenes Deutsch. Archivierung, Dokumentation und Erschließung des Deutschen Spracharchivs (DSAv). *IDS Sprachreport* 3/2005, 23–26.

Wagener, Peter & Karl-Heinz Bausch (Hrsg.) (1997): *Tonaufnahmen des gesprochenen Deutsch. Dokumentation der Bestände von sprachwissenschaftlichen Forschungsprojekten und Archiven*. Tübingen: Niemeyer.

Wildfeuer, Alfred (2017a): Sprachinseln, Sprachsiedlungen, Sprachminderheiten. Zur Bezeichnungsadäquatheit dieser und weiterer Termini. In Alexandra Lenz, Ludwig Breuer, Tim Kallenborn, Peter Ernst, Manfred Glauninger & Franz Patocka (Hrsg.), *Bayerisch-österreichische Varietäten zu Beginn des 21. Jahrhunderts – Dynamik, Struktur, Funktion* (Zeitschrift für Dialektologie und Linguistik – Beihefte 167), 373–388. Stuttgart: Steiner.

Wildfeuer, Alfred (2017b): *Sprachenkontakt, Mehrsprachigkeit und Sprachverlust. Deutschböhmisch-bairische Minderheitensprachen in den USA und in Neuseeland* (Linguistik – Impulse & Tendenzen 73). Berlin, Boston: de Gruyter Mouton.

Wittenburg, Peter, Hennie Brugman, Albert Russel, Alex Klassmann & Han Sloetjes (2006). ELAN: A professional framework for multimodality research. In Nicoletta Calzolari et al. (Hrsg.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, *Genoa, Italy*, 1556–1559. Paris: European Language Resources Association (ELRA).

Zwirner, Eberhard (1983): Fünfzig Jahre Deutsches Spracharchiv. *Zeitschrift für Dialektologie und Linguistik* 50 (1), 35–43.

Anhang: Sprachinselkorpora

Region/Land	Zeitraum	Hauptverantwortlich	Umfang	Wo vorhanden?[28]	Weitere Informationen
Colonia Tovar, Venezuela	1986/1987	Klaus Lückert	68 Aufnahmen	Privat?	Wagener & Bausch (1997: 4–6)
Israel	1989–1994, 2000–2011	Anne Betten, Kristine Hecker	178 Aufnahmen	über DGD zugänglich (Korpora IS, ISW, ISZ)	Wagener & Bausch (1997: 51–53)
Siebenbürgen, Rumänien	1966–1975	Ruth Kisch, Heinrich Mantsch	190 Aufnahmen	Ludwig-Maximilians-Universität München (Zugriff möglich), IDS (kein Zugriff)	Wagener & Bausch (1997: 69), http://www.asd.gwi.uni-muenchen.de/
Slowakei	1977–1980	R. Melzer	39 Tonbänder	Karpatendeutsches Heimatmuseum, Karlsruhe	Wagener & Bausch (1997: 85)
Ehemalige Sowjetunion, USA, Kanada, Mexiko, Paraguay	1959–1991	Ulrich Tolksdorf	Ungewiss, Teil einer Sammlung mit 544 Aufnahmen	Privatarchiv Tolksdorf, Kiel?	Wagener & Bausch (1997: 91–92)
Nord- und Südamerika	1980–?	Ulrich Tolksdorf	80 Aufnahmen	Universität Kiel? Evtl. Deutscher Sprachatlas Marburg	Wagener & Bausch (1997: 92–96)
Panambi, Brasilien	1985	Ute Bärnert-Fürst	23 Aufnahmen	Universität Mannheim, IDS	Wagener & Bausch (1997: 144–145)

28 Sofern die Autoren ermitteln konnten, dass es einen Wechsel in der die Aufnahmen beherbergenden Institution gegeben hat, ist dies hier aktualisiert. Dabei wird grundsätzlich davon ausgegangen, dass die neue Institution die Bestände weiterhin verwahrt. Wenn unklar ist, ob der bei Wagener & Bausch (1997) angegebene Verbleibeort aufgrund von Institutionswechseln oder ähnlichem weiterhin zutrifft, ist dies mit einem Fragezeichen markiert.

Region/Land	Zeitraum	Hauptverantwortlich	Umfang	Wo vorhanden?	Weitere Informationen
Ungarn, Slowakei	1957–1992	Alfred Camman	Ungewiss, Teil einer Sammlung von ca. 1.000 Aufnahmen	Verein der Freunde des Archivs für Heimatforschung in Rotenburg (Wümme) e. V.	Wagener & Bausch (1997: 179–180)
Südosteuropa	1954–1992	Hans Gehl (Überspielung)	212 Aufnahmen	Institut für donauschwäbische Landeskunde, Tübingen	Wagener & Bausch (1997: 192–194
Diverse USA	1964–?	Wolfgang W. Moelleken	ca. 430 Aufnahmen	University at Albany?, IDS (teilweise), Max-Kade-Institute	Wagener & Bausch (1997: 202–205)
Wisconsin, USA	1968–1969	Jürgen Eichhoff	64 Aufnahmen	Max-Kade-Institute, IDS (teilweise)	Wagener & Bausch (1997: 222–223)
Minnesota, USA	1986–?	Eveylin S. Firchow	ca. 150 Aufnahmen	University of Minnesota?	Wagener & Bausch (1997: 222–223)
Italien, Tschechien, Slowakei, Ungarn u. a.	1958–?	Maria Hornung	ca. 500 Aufnahmen	Verein der Freunde der von Österreich aus im MA besiedelten Sprachinseln/Verein der Sprachinselfreunde	Wagener & Bausch (1997: 247–248)
Brasilien, Neuseeland, Rumänien, Ukraine, USA (Kansas und Minnesota)	2005–2013	Nicole Eller-Wildfeuer, Alfred Wildfeuer	unter anderem ca. 50 Stunden Tonaufnahmen	Privat	Eller-Wildfeuer (im Druck) Wildfeuer 2017b