

Deutsche Sprache

DS 3.26

Zeitschrift für
Theorie • Praxis • Dokumentation

54. Jahrgang
3. Quartal 2026

ISSN 0340-9341

www.DSdigital.de

IDS

LEIBNIZ-INSTITUT FÜR
DEUTSCHE SPRACHE

ESV
ERICH
SCHMIDT
VERLAG

Hans C. Boas / Margo Blevins / Thomas Schmidt / Ellen J. Schoedler / Joren Somers

25 Jahre Texas German Dialect Project: Methodologie, digitale Ressourcen und Analysen zu diachronen und vergleichenden Aspekten einer Kontaktvarietät des Deutschen

Abstracts

Das Texas German Dialect Project (TGDP) wurde 2001 gegründet und ist an der University of Texas in Austin angesiedelt. Seit 25 Jahren dokumentiert es Texasdeutsch, eine vom Aussterben bedrohte Kontaktvarietät, die von den Nachfahren deutschsprachiger Einwanderer gesprochen wird, die sich ab den 1840er Jahren in Zentraltexas niedergelassen haben. Das Projekt hat bislang über 850 SprecherInnen aufgenommen und stellt die Tonaufnahmen mit Transkripten und englischen Übersetzungen im frei zugänglichen TGDA-Online bereit. Der Beitrag beschreibt die Datenerhebung und -verarbeitung, die technische Entwicklung des Projekts sowie die jüngste Integration des Korpus in die ZuMult-Plattform, die erweiterte Korpusabfragen ermöglicht.

Founded in 2001 and housed at the University of Texas at Austin, the Texas German Dialect Project (TGDP) has spent 25 years documenting Texas German, a critically endangered contact variety spoken by the descendants of German-speaking immigrants who settled in central Texas from the 1840s onwards. To date, the project has recorded over 850 speakers, creating a freely accessible online archive (TGDA-Online) of audio recordings, transcripts, and English translations. This article describes the project's data collection methodology, its technical evolution, and the recent integration of the corpus into the ZuMult platform, which enables advanced corpus linguistic queries.

Keywords: Texasdeutsch, Sprachkontakt, Sprachtod, Sprachdokumentation, Korpuslinguistik; Texas German, Language Contact, Language Death, Language Documentation, Corpus Linguistics

1. Einleitung¹

Der als „Texasdeutsch“ seit den 1970er Jahren bekannte Auswandererndialekt (siehe Gilbert 1972) hebt sich durch verschiedene Merkmale von anderen Varietäten außerhalb des geschlossenen deutschen Sprachraums ab.² Erstens ist das Texasdeutsche im Unterschied zu anderen Varietäten ein relativ junger Auswandererndialekt, dessen Ursprünge sich auf die 1840er Jahre zurückführen lassen. Zweitens ist das heute noch in der letzten Generation gesprochene Texasdeutsche im Gegensatz zu anderen deutschen Kontaktvarietäten ein Mischdialekt, der sich aus mindestens fünf unterschiedlichen regionalen Dialekten des Deutschen zusammensetzt.³ Die Entwicklung hin zu einem einheitlichen neuen Dialekt im Sinne von Trudgills (2004) Modell der „New Dialect Formation“ wurde u. a. durch die Folgen eines 1918 erlassenen Gesetzes gegen Deutsch als Schulsprache unterbrochen. Dies führte dazu, dass ab den 1920er Jahren immer weniger Eltern mit ihren Kindern Texasdeutsch sprachen, so dass es gegen Ende der 1940er Jahre zum kompletten Erliegen

¹ Für hilfreiche Kommentare bedanken wir uns bei Steffen Höder und Ekatarina Levina sowie bei zwei anonymen GutachterInnen.

² Der einheitliche Begriff „Texas German“ wird erstmals von Gilbert (1972) im Rahmen seines *Linguistic Atlas of Texas German* verwendet. Davor gibt es andere Publikationen, die ortsspezifische Angaben benutzen, wie z. B. „New Braunfels German“ (Clardy 1954; Eikel 1949, 1954), „German in Kendall and Gillespie Counties“ (Gilbert 1963) und „The German in Fredericksburg“ (Gilbert 1965).

³ Diese sind u. a. Hessisch, Ostfälisch, Moselfränkisch, Thüringisch und Alemannisch.

der Weitergabe des Texasdeutschen an jüngere Generationen kam. Deshalb wird das Texasdeutsche, wie von Wilson (1986) vorhergesagt, bis Mitte der 2030er Jahre ausgestorben sein.⁴ Drittens ist der Grad an sprachlicher Variation zwischen den SprecherInnen des Texasdeutschen so hoch (siehe Gilbert 1972; Boas 2009a), dass die Frage gestellt worden ist, ob es überhaupt eine einzige Varietät gibt, die als Texasdeutsch bezeichnet werden kann, oder ob man vielleicht eher von mehreren tausend SprecherInnen deutscher Dialekte (mit starkem Einfluss des Englischen auf allen systemlinguistischen Ebenen) ausgehen sollte, die in der 5. und 6. Generation nach der Einwanderung noch immer unterschiedlich gemischte Varietäten sprechen.

All dies macht das Texasdeutsche zu einem interessanten Forschungsgegenstand, unter anderem für die Bereiche Kontaktlinguistik, Dialektologie, Variationslinguistik und Dokumentation bedrohter Sprachen und Dialekte. Um das vom Aussterben bedrohte Texasdeutsche vor dessen Verschwinden zu dokumentieren und zu erforschen, wurde im Jahr 2001 an der University of Texas at Austin das Texas German Dialect Project (TGDP) gegründet. Das Hauptziel des TGDP seit 2001 ist es, so viele wie möglich der damals noch ca. 10.000 lebenden SprecherInnen des Texasdeutschen zu interviewen und die Tonaufnahmen der Interviews im Texas German Dialect Archive Online (TGDA-Online) zu archivieren und online bereit zu stellen, damit sie für Lehr- und Forschungszwecke online genutzt werden können.

Dieser Beitrag gibt einen Überblick darüber, wie das TGDP über ein Vierteljahrhundert hinweg Daten erhoben, verarbeitet und archiviert hat und wie diese Daten u. a. genutzt worden sind, um neue Einsichten über die Struktur und Entwicklung des Texasdeutschen sowie die Dialektologie deutsch-amerikanischer Dialekte und Sprachvariation und Sprachwandel im weiteren Sinne zu gewinnen. Dabei wird besonders auf den Ablauf der Datenerhebung und -verarbeitung sowie der dabei angewandten Methoden eingegangen und es wird diskutiert, wie sich technische und organisatorische Aspekte des TGDP in den letzten 25 Jahren entwickelt haben.

Der Beitrag ist wie folgt strukturiert. Abschnitt 2 skizziert die Hintergründe der deutschen Einwanderung nach Texas und zeigt, wie sich die deutschsprachige Bevölkerungsgruppe in Texas seit den 1840er Jahren entwickelt hat. Dabei werden auch einige relevante strukturelle Merkmale des Texasdeutschen besprochen. Abschnitt 3 stellt den Organisationsablauf und die Entwicklung des Texas German Dialect Projects vor, welches seit 2001 Daten von mehr als 850 SprecherInnen des Texasdeutschen erhoben hat. In diesem Zusammenhang erläutern wir auch, welche Arten von Daten das TGDP erhebt und wie diese Daten weiterbearbeitet werden, bevor sie in unterschiedlichen Formaten mit Transkripten und englischen Übersetzungen archiviert und online bereitgestellt werden. Vor diesem Hintergrund zeigen wir in Abschnitt 4, wie Nutzer im frei verfügbaren Texas German Dialect Archive Online (TGDA-Online) auf die im Archiv gespeicherten Daten unterschiedlich zugreifen können. Wir erläutern dann in Abschnitt 5 die Normalisierung der texasdeutschen Transkripte, ihre Anreicherung mit zusätzlichen Informationen und die Bereitstellung der Daten in der ZuMult-Korpusplattform. Die Nutzungsmöglichkeiten der ZuMult-

⁴ Wilson (1986) sagte bereits Mitte der 1980er Jahre das Aussterben des Texasdeutschen voraus: „Fifty years from now (in the year 2036) there will be no speakers of Texas German left, and monolingual Texans will find it hard to believe that German was a living language in Texas, spoken by hundreds of thousands for two hundred years. Every sample of Texas German that we can collect now, while Texas German is still alive, will be precious in the future.“ (ebd., S. 222).

Plattform für Forschung, Lehre und Outreach werden in Abschnitt 6 besprochen. Abschnitt 7 zeigt, wie andere texasdeutsche Datenquellen in das Archiv aufgenommen und verarbeitet werden und stellt ein Projekt vor, das mit Hilfe künstlicher Intelligenz versucht, ältere Sprachstufen des Texasdeutschen aus dem 19. Jahrhundert zu rekonstruieren. Abschnitt 8 fasst den Beitrag zusammen.

2. Texasdeutsch: Ein Überblick

Campbell (2003) zufolge zogen die ersten deutschsprachigen Einwanderer schon Ende der 1820er Jahre nach Texas. In den 1830er Jahren ließen sich Friedrich Ernst und seine Familie in Texas (Austin County) nieder. Mehrere Familien folgten ihrem Beispiel und begannen mit der Besiedlung des Gebiets, das später zum östlichen Teil des „German Belt“ wurde. Die Migration kam aber erst in den 1840er Jahren voll in Gang, und zwar auf Initiative des 1842 gegründeten „Mainzer Adelsvereins“ (Biesele 1930). Diese Gesellschaft erwarb Land in Texas, das sie deutschen Einwanderern zur Besiedlung anbot, was bis Mitte der 1840er Jahre zur Ansiedlung von ca. 7.000 Deutschen in Texas führte (Boas 2009a). Die meisten Neuankömmlinge wanderten laut Boas (ebd.) aus ökonomischen Gründen aus, aber manche flohen auch aus politischen Gründen aus ihrer Heimat nach Texas (Nicolini 2004). Den Brennpunkt der deutschen Migration bildete ein Gebiet in der Mitte von Texas, welches als der „German Belt“ (siehe Abb. 1) bekannt ist und u. a. Ortschaften wie Weimar, New Ulm, Brenham, La Grange, Boerne, New Braunfels, Comfort und Fredericksburg umfasst. Insbesondere die letzten beiden zählen als traditionelle Bastionen des Texasdeutschen (vgl. Boas/Pierce 2025, S. 159).

In den darauffolgenden Jahren stieg die Zahl der DeutschsprecherInnen in Texas rasch an: Um 1860 gab es etwa 20.000 deutschsprachige TexanerInnen und Ende des 19. Jahrhunderts, d. h. um 1890, war ihre Zahl schon auf 40.000 angestiegen (Pierce/Boas/Roesch 2015, S. 118). Zu Beginn des Ersten Weltkriegs betrug die Anzahl der deutschsprachigen TexanerInnen laut Eichhoff (1986) etwa 100.000. Der Höhepunkt wurde um 1940 erreicht, als es um die 160.000 SprecherInnen des Deutschen in Texas gab (Kloss 1977). Im Jahr 2026 ist das Texasdeutsche akut vom Aussterben bedroht: Boas (2018a) schätzte die damalige Sprecherzahl auf nicht mehr als 3.000. Seitdem hat sich diese Zahl u. a. auch wegen der Covid-19 Pandemie weiter drastisch verringert. Aller Wahrscheinlichkeit wird das Texasdeutsche, wie von Wilson (1986) vorausgesagt, bis zum Jahr 2036 völlig ausgestorben sein.

Die Blütezeit des Texasdeutschen dauerte von den 1840er Jahren bis zum Anfang des 20. Jahrhunderts. Während dieser Zeit war die Varietät stark in der Gesellschaft verankert, mit deutschsprachigen Schulen, Gottesdiensten und Zeitungen, in denen überwiegend Standarddeutsch benutzt wurde (Boas 2009a, S. 47–54).⁵ Der Beginn des Ersten Weltkrieges stellte aber einen drastischen Wendepunkt für das Fortbestehen der deutschen Sprache dar.

⁵ Es bestand in Texas bis zum Ende des Ersten Weltkriegs eine Diglossie mit Standarddeutsch als Prestigevarietät und Texasdeutsch als „Nicht-Prestigevarietät“, die hauptsächlich in der Familie und unter Freunden und Nachbarn sowie in den unterschiedlichen Tanz-, Gesangs-, Sport- und Schützenvereinen benutzt wurde. Über den Einfluss des Standarddeutschen auf das Texasdeutsche gibt es unterschiedliche Analysen. Salmons/Lucht (2006) vertreten die Ansicht, dass das Standarddeutsche bis zum Ende des Ersten Weltkriegs einen relativ großen Einfluss auf die Verwendung und Struktur des Texasdeutschen hatte, während Boas (2009a), Boas/Fuchs (2018) und Boas/Levina (2024) argumentieren, dass dieser Einfluss des Standarddeutschen, wenn überhaupt, nur sehr gering gewesen sein kann.

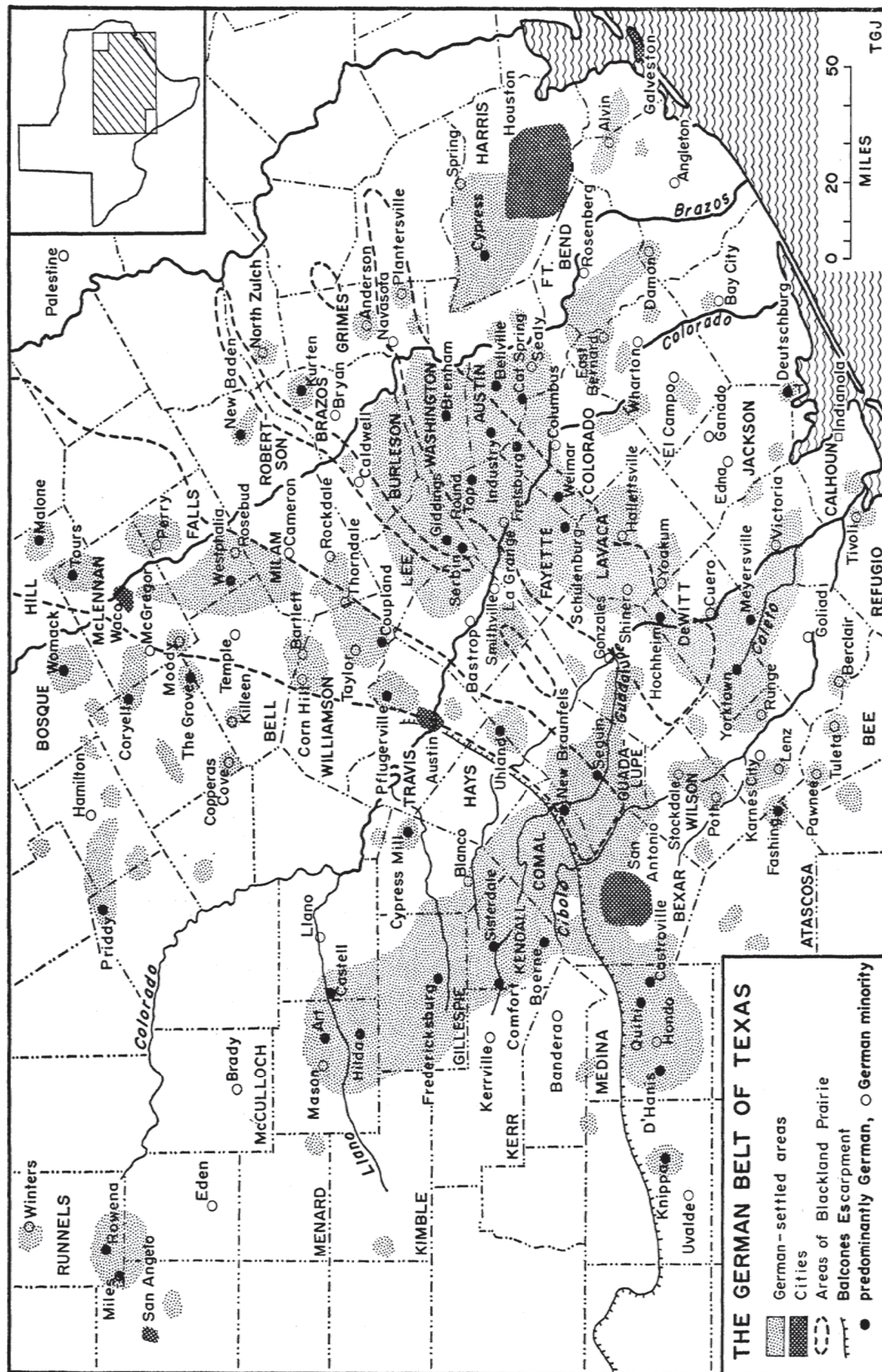


Abb. 1: Der „German Belt“ in Zentral-Texas (Jordan 1969, S. 200)

So wurde 1918 ein Gesetz verabschiedet, das die Verwendung des Deutschen in Schulen verbot, und auch Kirchen und Zeitungen wechselten in zunehmendem Maße zum Englischen; auch deutschsprachige Vereine trafen sich nicht länger (Boas 2009a, S. 55–61). Boas (2018a, S. 175 f.) fasst zusammen, dass der Erste Weltkrieg eine regelrechte „Deutschfeindlichkeit“ auslöste. Infolgedessen verlor das Deutsche in Texas (sowie in den gesamten Vereinigten Staaten) drastisch an Prestige und das Englische übernahm die Rolle der gesellschaftlich führenden Prestigesprache (Boas 2018a, S. 176).

Nach dem Ersten Weltkrieg verschlechterte sich die Lage des Texasdeutschen weiter, wofür Boas (ebd., S. 176 f.) zufolge fünf Gründe zu nennen sind. Erstens kam in den Nachkriegsjahren die deutsche Einwanderung nach Texas zum Stillstand. Zweitens sorgten groß angelegte Infrastrukturprojekte dafür, dass die deutschsprachigen Gebiete in Texas besser mit den englischsprachigen verbunden wurden, was den zuvor stabilen Zustand des Texasdeutschen untergrub. Drittens führten immer häufiger vorkommende Mischehen zwischen Deutsch- und EnglischsprecherInnen dazu, dass das Englische in vielen Haushalten de facto zur Familiensprache wurde. Viertens schuf die Nachkriegswirtschaft viele Beschäftigungsmöglichkeiten in englischsprachigen Städten, wodurch Texasdeutsche dazu gezwungen waren, bei der Arbeit Englisch zu sprechen. Fünftens sei angemerkt, dass der Zweite Weltkrieg einen weiteren Schlag für das Prestige des Texasdeutschen darstellte (Boas 2009a, S. 70).

Sprachlich ist das Texasdeutsche von einer weitreichenden Variation geprägt, was unter anderem damit zu tun hat, dass die ursprünglichen deutschsprachigen Einwanderer aus verschiedenen Dialektgebieten mit unterschiedlichen niederdeutschen, mitteldeutschen oder oberdeutschen Regionaldialekten nach Texas kamen, z. B. aus den Gebieten der heutigen deutschen Bundesländer Hessen, Rheinland-Pfalz, Thüringen, Mecklenburg-Vorpommern und Niedersachsen (Boas 2009a, S. 36).⁶ Außerdem gab es nie eine einheitliche, flächendeckende deutschsprachige Gemeinschaft in Texas, nicht nur aufgrund der Größe des Siedlungsgebiets, sondern auch, weil die Mehrheit der Einwanderer in geographisch isolierten Gemeinden lebte. Dennoch lassen sich bestimmte strukturelle Regularitäten ausmachen, die im Folgenden skizziert werden.

Im lexikalischen Bereich fällt vor allem die Anzahl englischer Lehnwörter auf, deren Zahl sich im Laufe der Jahrzehnte ständig gesteigert hat. Gilbert (1965, S. 110) geht zum Beispiel für das Texasdeutsche der 1960er Jahre davon aus, dass nur 5% des gesamten Wortschatzes dem Englischen entstammen. Boas/Pierce (2011) stimmen dieser Schlussfolgerung zu: Aufbauend auf in New Braunfels gesammelten Daten für den Zeitraum 2001–2006 zeigen sie, dass das Lexikon im Vergleich zu den von Gilbert erhobenen Daten relativ stabil geblieben ist. Typische Entlehnungsbereiche umfassen laut Boas/Pierce (2011) die Fauna und Flora (z. B. *der Pecanbaum, die Creek*), die Landwirtschaft (z. B. *die Fence, der Pasture*), die Politik (z. B. *der County Commissioner, naturalizen*) und das Schulwesen (z. B. *die Teacher, die Grade*).

Das texasdeutsche Lautsystem lässt sich am besten anhand von Gilbert (1972), Boas (2009a), Warmuth (2023) und Boas/Pierce (2025, S. 162–170) skizzieren. Eines der auffälligsten phonologischen Merkmale des Texasdeutschen ist demnach der Mangel an vor-

⁶ Eine Ausnahme bildet das Texas-Elsässische, welches als einheitlicher Dialekt von den 1850er Jahren bis in die 2010er Jahre westlich von San Antonio gesprochen wurde, aber seit ca. fünf Jahren ausgestorben ist (siehe Roesch 2012).

deren gerundeten Vokalen /y:/, /ø:/, /ʏ/ und /œ/ zugunsten ihrer ungerundeten Gegenstücke. Wörter wie *Tür*, *schön*, *fünf* und *zwölf* klingen also oft wie ,T[i:]r‘, ,sch[e:]n‘, ,f[ɪ]nf‘ und ,zw[ɛ]lf‘. Zweitens hat sich in englischen Lehnwörtern das Lehnphonem /æ/ durchgesetzt, z.B. in *Tractor* [ˈtræktər] ‚Traktor‘ oder *Candy* [ˈkændi] ‚Süßigkeit‘ (im Gegensatz zu [ɛ]). Drittens kommt auch der Reibelaut /pf/ im Texasdeutschen nur selten vor: wortinitial wird er vom Reibelaut [f] abgelöst, wortmedial und -final bleibt er entweder unverändert oder erscheint als Verschlusslaut [p]. Schließlich wird in Wörtern wie *Donnerstag* [ˈdɒnəs, ta:k] oder *Haarbürste* [ˈha:ɐ̯, byʁstə] das [s] oft als ein an ein vorausgehendes r assimiliertes [ʃ] realisiert und zwar wie folgt: [ˈdɒnɐʃ, ta:k] und [ˈha:ɐ̯, byʁʃtə] (siehe Gilbert 1972 und Lindemann 2019).

Die Morphosyntax des Texasdeutschen unterscheidet sich Boas (2009a, S. 238) zufolge in einigen Bereichen von den Ursprungsdiakten und ist nur begrenzt vom Kontakt mit dem Englischen beeinflusst worden. Was die Wortstellung angeht, folgt das Texasdeutsche oft dem standardsprachlichen Muster, was bedeutet, dass Hauptsätze meistens nach der geläufigen Verbzweitstellung und Nebensätze meist nach der geläufigen Verbendstellung aufgebaut werden. Auch die Genuszuweisung geschieht noch immer weitgehend gemäß dem dreigliedrigen System (mit einigen Abweichungen), das auch im Hochdeutschen geläufig ist. Andererseits lassen sich auch morphosyntaktische Vereinfachungen beobachten (siehe Boas 2016). So hat sich im Texasdeutschen auch der sogenannte Präteritumschwund durchgesetzt, der bereits in einigen Ursprungsdiakten vorhanden war. Dieses Phänomen besteht darin, dass der Bereich des Präteritums (z.B. *ich kam*) von dem des Perfekts (z.B. *ich bin gekommen*) übernommen wird. Schließlich ist auch das texasdeutsche Kasussystem zu nennen, das nur noch in stark reduzierter Form vorhanden ist (siehe Boas 2009b, 2011).

In den letzten 70 Jahren gab es mehrere größere Untersuchungen zum Texasdeutschen: Eikel (1954), Gilbert (1972), Boas (2009a) und Warmuth (2023). Eikel (1954) ist eine der ältesten Studien zur Phonologie, Morphologie und Syntax des Texasdeutschen in Neu Braunfels und basiert auf Daten von 24 Teilnehmern aus drei Generationen.⁷ Gilbert (1963) erweitert die Analyse der Struktur und Entwicklung des Texasdeutschen auf den westlichen Teil des German-Belts (Gillespie und Kendall Counties) und Gilbert (1972) präsentiert eine variationslinguistische Studie des Texasdeutschen: Sein bahnbrechendes Werk, der *Linguistic Atlas of Texas German*, ein Dialektatlas im Stil der Marburger Schule, enthält 148 Karten, die sich über ein aus 31 Landschaften bestehendes Gebiet in Zentral-texas erstrecken (ebd., S. 1). Boas (2009a) ist die bislang umfangreichste Monographie zur Struktur und Entwicklung des Texasdeutschen in Comal County, wo sich auch New Braunfels befindet. Dieses Werk ist die erste großangelegte Untersuchung des Texasdeutschen, die auf der Datensammlung im Rahmen des 2001 gegründeten TGDp basiert. Warmuth (2023) untersucht die Entstehung von einzigartigen Diphthongen im Texasdeutschen. Diese Studien bilden wichtige Bezugspunkte, die zusammengekommen eine Längsschnittuntersuchung diachroner Veränderungen im Texasdeutschen ermöglichen und damit auch unser Verständnis von bestimmten sprachlichen Phänomenen in den Bereichen Sprachwandel, Sprachtod, Sprachkontakt und Dialektvermischung erweitern.

⁷ Zum Status der von Eikel (1954) erhobenen Daten siehe Pierce/Boas/Gilbert (2019, 2021).

3. Das Texas German Dialect Project

Zum Zeitpunkt der Gründung des Texas German Dialect Project (TGDP) durch Hans C. Boas an der University of Texas at Austin (UT-Austin) im Jahr 2001 betrug die geschätzte Zahl der SprecherInnen des Texasdeutschen nur noch 8.000–10.000, wovon die Mehrheit damals schon 60 Jahre und älter war. Das primäre Ziel des TGDP war und ist deshalb, so viele verbliebene SprecherInnen des vom Aussterben bedrohten Dialekts wie möglich zu finden und mit ihnen Interviews zu führen, um Tonaufnahmen, Transkripte und englische Übersetzungen in einem frei zugänglichen Digitalarchiv zu sichern und sie damit auch für Lehr- und Forschungszwecke zur Verfügung zu stellen.⁸ Zwischen 2001 und 2002 führte Boas nicht nur die ersten Interviews des TGDPs, sondern beantragte auch Drittmittel, um die Finanzierung weiterer Interviews sowie deren Bearbeitung und Archivierung sicherzustellen. Er beschäftigte sich auch intensiv mit allen technischen Aspekten des Aufbaus eines digitalen Archivs des Texasdeutschen.

Im Jahr 2001 war es noch relativ einfach, SprecherInnen des Texasdeutschen zu finden. Die ersten Interviews fanden im Umkreis von Austin statt, in den Gemeinden New Braunfels, Fredericksburg, Schulenburg und Freyburg. Nachdem die ersten paar Dutzend Interviews, die von Boas alleine von 2001–2003 durchgeführt worden waren, segmentiert, transkribiert, übersetzt und archiviert worden waren, konnten sie in Linguistik- und Deutschkursen an der UT-Austin im Unterricht eingesetzt werden. Dadurch wuchs das Interesse der Studierenden an linguistischer Feldforschung, und ab 2004 fingen auch Studierende und DoktorandInnen an, bei Interviews mitzuhelfen.

Ab diesem Zeitpunkt wurden dann auch regelmäßig Interviews mit Gruppen organisiert, wobei bis zu zwölf SprecherInnen über einen Tag von mehreren InterviewerInnen befragt und aufgenommen wurden. Diese Aufnahmen wurden in öffentlichen Bibliotheken, Archiven oder Altenheimen durchgeführt. Interviews dauern üblicherweise zwei bis drei Stunden, oft mit Pausen zwischendurch. Vor dem Interview müssen die Gewährspersonen ein Formular ausfüllen, mit dem sie ihre Zustimmung zum Interview und dessen weiterer Bearbeitung, Archivierung und Auswertung für Lehr- und Forschungszwecke geben.

Am Anfang des Projekts wurden die Interviews mit Sony MiniDisc Rekordern aufgezeichnet, welche damals zwar einfach in der Bedienung und vergleichsweise günstig waren, aber deren Aufnahmequalität suboptimal war. Ab 2004 kamen dann auch digitale DAT-Rekorder zum Einsatz, die eine viel höhere Aufnahmequalität lieferten, aber auch deutlich teurer als die ursprünglichen MiniDisc Rekorder waren. Seit 2008 verwendet das TGDP fast ausschließlich sog. Solid-State Flash-Rekorder von Roland und Zoom, die eine

⁸ Es gibt auch frühere Aufnahmen mit SprecherInnen des Texasdeutschen, welche die Basis für wissenschaftliche Publikationen gewesen sind. In den 1940er Jahren hat Fred Eikel einige seiner Gewährspersonen in Neu Braunfels für sein Manuskript (Eikel 1954) aufgenommen, aber diese Aufnahmen sind leider bis auf eine Schellack-Schallplatte verschollen. In den 1960er Jahren nahm Glenn Gilbert mehrere seiner Gewährspersonen für seine Dissertation (Gilbert 1963) und seinen *Linguistic Atlas of Texas German* (1972) auf. In den 1990er Jahren nahm Susan Guion (1996) 16 texasdeutsche SprecherInnen in Gillespie County für ihren Artikel auf. Das TGDP besitzt Kopien der Tonaufnahmen von Gilbert und Guion (diese sind aber noch nicht vollumfänglich online zugänglich). Seit Dezember 2025 sind die freien Gespräche von Gilberts Sammlung (mit Transkriptionen) online verfügbar, siehe Abschnitt 5. Auf die von Guion durchgeführten narrativen Interviews kann derzeit noch nicht über TGDA-Online und ZuMult zugegriffen werden.

hohe Aufnahmequalität im WAV-Format erlauben (Boas et al. 2010).⁹ Von April 2020 bis August 2021 wurden die Interviews wegen der Covid-19-Pandemie nur per Videokonferenz (Zoom) oder Telefon geführt. Nach August 2021 haben die Mitglieder wieder angefangen, Interviews in Präsenz zu organisieren. Seitdem ist es bedauerlicherweise immer schwieriger geworden, SprecherInnen des Texasdeutschen zu finden und zu überzeugen, an einem Interview teilzunehmen. Einerseits liegt das daran, dass die Zahl der verbliebenen SprecherInnen jedes Jahr sinkt. Andererseits ist es auch schwierig, interessierte SprecherInnen zu finden, mit denen das Projekt bis jetzt noch kein Interview geführt hat.

Sowohl Einzelinterviews als auch Interviews, bei denen mehrere SprecherInnen an einem Tag teilnehmen, werden normalerweise einzeln durchgeführt, d. h. eine Gewährsperson wird von einem Mitglied des TGDP befragt. Die InterviewerInnen sprechen während eines Interviews meistens Hochdeutsch, welches die SprecherInnen des Texasdeutschen fast immer verstehen. Die InterviewerInnen sind manchmal MuttersprachlerInnen, haben aber auch oft Deutsch als Zweit- oder Fremdsprache. Wenn die Gewährspersonen zu viele Probleme mit dem Hochdeutschen haben oder Unklarheiten entstehen, wechseln die InterviewerInnen ins Englische, um die Fragen oder Anweisungen zu wiederholen. Es ist daher von entscheidender Bedeutung, dass alle Mitglieder des TGDPs, die Interviews durchführen, sowohl Hochdeutsch als auch Englisch beherrschen.

3.1 Datenerhebung

Der Fokus der Datenerhebung lag während der Anfangsphase des Projekts auf narrativen Interviews. Seit 2005 werden drei unterschiedliche Arten von Daten anhand dreier unterschiedlicher Fragebögen aufgenommen: 1) biografische bzw. demografische Daten; 2) narrative Interviews, in denen unterschiedliche Themen besprochen werden (z. B. Wetter, Herkunft der Vorfahren, Feste feiern, Zubereitung von Essen usw.); und 3) Übersetzungen von Wörtern, Phrasen und Sätzen vom Englischen ins Texasdeutsche (siehe Tab. 1 unten).

Das TGDP benutzt einen elfseitigen englischsprachigen Fragebogen, um biografische bzw. demografische Informationen über die fragliche Gewährsperson zu sammeln, wie z. B. Alter, Herkunft, Religion und Bildungsgrad sowie unterschiedliche Aspekte des Sprachgebrauchs und Sprach- und Sprechereinstellungen (siehe Abb. 2). Während des Interviews wird der Fragebogen von dem Interviewer oder der Interviewerin schriftlich ausgefüllt und das auf Englisch geführte Gespräch nebenbei aufgenommen.

Dieser Teil des Interviews wird auf Englisch geführt, um festzustellen, ob Probleme beim Sprechen des Texasdeutschen wie Wortfindungsprobleme, Stottern, häufige Reparatur von Satzstrukturen usw. auch beim Sprechen des Englischen auftreten. Solche Auffälligkeiten oder Schwierigkeiten sind vielleicht auch in ihrer Produktion des Texasdeutschen zu erwarten. Um ein umfassenderes Bild der Sprachkompetenzen zu bekommen, wird auch erhoben, wie die Gewährsperson Englisch spricht, d. h. ob er oder sie Englisch mit einem deutschen Akzent spricht oder nicht. Der biografische Teil eines Interviews wird auch deshalb auf Englisch geführt, weil die Gewährspersonen komplexere Fragen zu soziologischen, lokal-historischen oder linguistischen Themen besser auf Englisch beantworten können.

⁹ Bei einigen Interviews werden auch, je nach Lage, Videoaufnahmen aufgezeichnet. Diese werden jedoch ausschließlich projektintern verwendet und sind der Öffentlichkeit nicht zugänglich.

Biographical Information Speaker No. 24

Questionnaire – Texas German Dialect Project

In case a question does not apply to you, please skip it. All answers are kept strictly confidential. In case you feel like you would like to add information, please use sections labeled "additional comments." Thank you very much for your help. We greatly appreciate it!

Name [redacted]

Year of Birth 1915

Place of Birth New Braunfels

Current place of residence ditto

Previous places of residence US Airforce
Corpus Christi / MO, FL,
UK, other countries in Europe

Religious affiliation Other

Which town or region did your ancestors come from in Europe?
Subject of the Duke of
Namau, a farmer. Other
ancestors from
1945/6
Sclesburg x Mucklenburg

1) Which language(s) did you learn first at home?
both German
and English; German
predominant

2) When you started school, which language(s) could you speak?
Engl., German
7 yrs old
segregation, i.e. no
Mexicans until
1930s

3) How old were you when you learned your second language?
does not remember when, started
right away

4) Where and from whom did you learn your second language?
parents, neighbors, friends,
grandparents

5) When you were a kid how often did you speak (a) German, (b) English, with _____?

PLEASE CIRCLE YOUR ANSWER(S). Please make sure to include information on your use of German AND English for each point. For example, if you circle "always" for German, then you are likely to circle "never" or "sometimes" for English.

	(German) Always	Often	regularly	sometimes	never
Parents:	☒	☐	☐	☐	☐
Grandparents:	☒	☐	☐	☐	☐
Teacher:	☒	☐	☐	☐	☐
Friends:	☒	☐	☐	☐	☐
Siblings:	☒	☐	☐	☐	☐
Neighbors:	☒	☐	☐	☐	☐
Church:	☒	☐	☐	☐	☐
School:	☒	☐	☐	☐	☐
Home:	☒	☐	☐	☐	☐
Shops:	☒	☐	☐	☐	☐
Large Family Gatherings:	☒	☐	☐	☐	☐
Other (boy scouts, etc.):	☒	☐	☐	☐	☐

Still when you were a kid-how often did you speak (a) German, (b) English, at _____?

older folks more likely to sp. Gm. dep. on circumstances

Gm. discouraged during WWI and WWII

Abb. 2: Die ersten beiden Seiten des biografischen bzw. demografischen Fragebogens

Die narrativen Interviews bestehen aus längeren Gesprächen, wobei einer Gewährsperson sowohl thematisch gegliederte Fragen über Lebensbereiche wie z. B. ihre Kindheit, Familie, Vorfahren, Ausbildung, Schulerfahrungen, Sprachgebrauch, Essen und Trinken als auch Wetter, Lokalgeschichte und -gebräuche gestellt werden.¹⁰ Diese Gespräche dauern in der Regel zwischen 30 und 60 Minuten und sollen in diesem Kontext so natürlich wie möglich sein. Die Fragen sollen auf Texasdeutsch beantwortet werden, denn das Ziel ist, die Gewährspersonen in ein Gespräch zu verwickeln, bei dem sie sich so natürlich wie möglich auf Texasdeutsch ausdrücken und nicht nur kurze Antworten auf einzelne Fragen geben. Solche freien Gespräche zeigen, wie gut die SprecherInnen das Texasdeutsche noch beherrschen. Es kommt auch oft vor, dass Gewährspersonen am Anfang des Gesprächs Schwierigkeiten haben, Sätze zu formulieren oder sich an bestimmte Wörter zu erinnern, aber diese Schwierigkeiten verschwinden bei den meisten SprecherInnen im Laufe des Interviews.¹¹

Schließlich beruht der Übersetzungsteil des Interviews auf den Listen von Eikel (1954), Gilbert (1972) und Guion (1996), siehe die Beispiele in Tabelle 1.¹² Die Daten, die während dieses Interviewabschnitts erhoben werden, ermöglichen einen systematischen Ver-

¹⁰ Zum Konzept des narrativen Interviews siehe Wolfram (2011) und Becker (2017). Obwohl die InterviewerInnen eine Liste mit möglichen Fragen erhalten, haben sie auch die Freiheit, eigene Fragen zu formulieren. Das bedeutet, dass es keine Standardfragen gibt, die jeder Gewährsperson gestellt werden – jedes narrative Interview ist einzigartig. Viele der oben genannten Themen werden jedoch häufig angesprochen.

¹¹ Dieses Phänomen wurde noch nicht eingehend untersucht, aber von mehreren InterviewerInnen beobachtet.

¹² Während des Übersetzungsinterviews bittet der Interviewer oder die Interviewerin die Gewährsperson, eine Liste englischer Wörter und Sätze ins Texasdeutsche zu übersetzen. Der Interviewer oder die Inter-

gleich zwischen SprecherInnen, um zu untersuchen, wie unterschiedlich dieselben Wörter, Phrasen oder Sätze im Texasdeutschen realisiert werden (siehe z.B. Boas 2009a; Lindemann 2019; Warmuth 2023). Da die Listen auf früheren Studien von Eikel, Gilbert und Guion basieren, ist es auch möglich, Veränderungen im Lexikon, der Phonologie, der Syntax und der Morphologie seit den 1940er bzw. 1960er und 1990er Jahren festzustellen. Dieser Teil des Interviews dauert insgesamt zwischen 15 und 75 Minuten.¹³

Eikel (1954)	Gilbert (1972)	Guion (1996)
Good morning. That serves him right. If I were a farmer I would raise more cotton. Do you like to eat sausage? The country roads are not as bad as they used to be. The cat is sitting in front of the door.	one apple two goats a horse He's running now. He told us that. Is that a squirrel on the pecan tree? This chicken has long feathers. His coughing is getting worse and worse.	My brother's friend is walking around the ranch to check up on the sheep. When I was young, everyone in town spoke German. Where were you going without the car?

Tab. 1: Beispielauszüge aus den Übersetzungslisten von Eikel (1954), Gilbert (1972) und Guion (1996)

3.2 Datenverarbeitung

Nach den Interviews werden die Tonaufnahmen katalogisiert. Jede Tonaufnahme bekommt eine vierstellige Zahlenkombination, die als Dateiname benutzt wird und eine eindeutige Identifikation bezüglich der wichtigsten Parameter (InterviewerIn, Gewährsperson) ermöglicht.¹⁴

Die vollständigen Tonaufnahmen werden nach der Katalogisierung mit der Audiosoftware *Audacity*¹⁵ in kleinere Abschnitte aufgeteilt – eine Aufgabe, die häufig von Studierenden der UT-Austin übernommen wird. Während dieser Bearbeitung werden persönliche Informationen maskiert, d. h. Namen von Personen bzw. Informationen, die Rückschlüsse auf

viewerin liest ein Wort oder einen Satz auf Englisch vor und fragt die Gewährsperson, wie sie dies in einem Gespräch auf Texasdeutsch ausdrücken würde. Es kommt selten vor, dass eine Gewährsperson alle drei Übersetzungsaufgaben vollständig beantwortet. Die Gilbert-Aufgabe wird oft zuerst erledigt, und wenn dann noch Zeit und Energie für weitere Übersetzungen vorhanden sind, kann der Interviewer oder die Interviewerin auch (einen Teil der) Guion- und/oder Eikel-Übersetzungsfragen stellen.

¹³ Die Gilbert- und Eikel-Übersetzungslisten sind auf TGDA-Online sichtbar. Die Übersetzungslisten sind auch unter <https://doi.org/10.18738/T8/PVWFB0> verfügbar.

¹⁴ Ein Beispiel: An dem Dateinamen 1-543-2-0-a.wav erkennt man, dass es sich um eine Tondatei (-a = „Audio“) handelt. Die Datei enthält ein Interview, das von InterviewerIn Nr. 1 mit SprecherIn Nr. 543 geführt wurde, und jede Art des Interviews erhält ihre eigene Nummer (hier: Interview Nr. 2). Die „0“ zeigt auch, dass die Datei die Originalaufnahme enthält, das sog. „Masterfile“ für Interview Nr. 2. Die vollständigen Tonaufnahmen („Masterfiles“) und digitalisierten Kopien des biografischen Fragebogens (als PDF) werden in einer projektinternen Cloud gespeichert und sind nicht öffentlich zugänglich.

¹⁵ <https://www.audacityteam.org/> (Stand: 11.4.2026).

die Identität der Gewährsperson ermöglichen könnten, werden mit einem Ton überspielt und auch im Transkript z. B. durch „[NAME]“ ersetzt. Diese Abschnitte werden danach in einer Datenbank auf dem Server des *Liberal Arts Instructional Technology Services* (LAITS) der UT-Austin gespeichert. Jede Datei, die Teil des narrativen Interviews ist, erhält auch einen deskriptiven Titel, der den Inhalt kurz zusammenfasst, z. B. „Speaking German in school“ oder „Ancestors coming from Germany to Texas“ (Boas et al. 2010; Boas/Fingerhuth 2017).

Auf dem LAITS-Server sind nicht nur die Teilausschnitte gesichert, sondern auch die Daten aus den biografischen Fragebögen. Solche Metadaten ermöglichen soziolinguistische Analysen der Tonaufnahmen, z. B. um festzustellen, ob bestimmte soziolinguistische Variablen wie Alter, Geschlecht, Bildungsgrad oder positive Spracheinstellung gegenüber dem Texasdeutschen einen Einfluss auf bestimmte sprachliche Merkmale haben. Als Beispiel hat Rybarski (2006) mit Hilfe der Metadaten der Gewährspersonen herausgefunden, dass unter den Texasdeutschen die Frauen dazu neigen, Akkusativmarkierungen in Nominalphrasen zu verändern und bei gleichem Kontext deutlich weniger Dativmarkierungen als Männer zu verwenden, was einen geschlechtsspezifischen Unterschied in der Kasusmarkierung aufzeigt.¹⁶

Nachdem die einzelnen Abschnitte einer Tonaufnahme gesichert worden sind, folgen die Transkription der Tonaufnahmen der narrativen Interviews und deren Übersetzung ins Englische.¹⁷ Mitglieder des TGDG laden zuerst eine bestimmte Tondatei auf ihren Computer herunter und bearbeiten sie dann mit ELAN, einem frei verfügbaren vom *Max-Planck Institut in Nijmegen* im Rahmen des Forschungsprogramms *Dokumentation bedrohter Sprachen (DobeS)* entwickelten Transkriptionsprogramm (siehe Sloetjes/Wittenburg 2008).¹⁸ Die Tondateien werden transkribiert und ins Englische übersetzt.

3.3 Datenarchivierung

Die Transkriptionen werden in einer literarischen Umschrift (Zwirner/Bethge 1958) verfasst, denn die Transkripte sollen nicht nur für Linguisten, sondern auch für andere Akademiker bzw. Laien zugänglich sein. Die Transkripte werden Wort für Wort ins Englische übersetzt, da nicht alle Nutzer des Dialektarchivs die Tonaufnahmen oder sogar Deutsch selbst verstehen können. Die Transkriptionen und deren Übersetzungen werden dann auf dem LAITS-Server abgelegt. Danach werden die Tondateien mit den zugehörigen Transkriptionsdateien über TGDA-Online veröffentlicht.

Bis Dezember 2025 hat das TGDG über 850 Texasdeutsch-SprecherInnen interviewt. Zum Zeitpunkt der Aufnahmen war die jüngste Person 52 Jahre alt und die älteste 101 Jahre alt, mit einem Durchschnittsalter von 78 Jahren. Die Geburtsjahre der Texasdeutsch-SprecherInnen reichen von 1903 bis 1963. Was das Geschlecht angeht, sind 35% der aufgenommenen Texasdeutschen weiblich und 65% männlich.¹⁹

¹⁶ Die vollständigen biografischen Fragebögen sind derzeit nicht öffentlich zugänglich.

¹⁷ Die Interviewsegmente, die Teil der Übersetzungsinterviews sind, werden auf den Server hochgeladen und nicht transkribiert.

¹⁸ <https://archive.mpi.nl/tla/elan>.

¹⁹ Alle hier angegebenen Daten und Prozentsätze basieren auf den Daten von TGDA-Online (Version 2.1, Dezember 2025) und nicht auf denen der größeren TGDA (siehe Kap. 3.4). Die Daten von TGDA-Online sind relativ repräsentativ für die größere TGDA.

Von 2001 bis 2025 haben über 120 Studierende bzw. DoktorandInnen Interviews geführt. Mehr als 50 Studierende haben Tonaufnahmen transkribiert und übersetzt. Einige Studierende waren nur für ein Semester Mitglied des TGDP, während andere mehrere Jahre daran arbeiteten. So entstanden an der UT-Austin auch einige Bachelorarbeiten (z. B. Rybarski 2006), Masterarbeiten (z. B. Bathe 2005; Whitworth 2005; Thompson 2005; Blevins 2014; Jones 2022) und Dissertationen (Roesch 2009; Blevins 2022; Warmuth 2023) über das Texasdeutsche (siehe auch Abschn. 6 zu weiteren Publikationen, die auf den TGDA-Online Daten beruhen).

In den letzten Jahren haben Mitglieder des TGDP nicht nur Interview-Daten erhoben, sondern das TGDP hat auch historische Briefe, Tagebücher, Protokolle von Vereinssitzungen und Fotos von Gewährspersonen gespendet bekommen. Solche Materialien wurden (wenn nötig) digitalisiert und in der projektinternen Cloud des TGDP gesichert, um so zu ermöglichen, dass die Materialien für künftige Forschungszwecke genutzt werden können. Es ist geplant, sie in Zukunft auch einem breiteren Publikum zur Verfügung zu stellen.

3.4 TGDP-Ressourcen

Das TGDP enthält eine Vielzahl von Ressourcen. Abbildung 3 veranschaulicht die verschiedenen Ressourcen, die in diesem Beitrag beschrieben werden, und wie sie miteinander in Beziehung stehen. Die verschiedenen Sammlungen und Zugangsweisen werden im Folgenden kurz skizziert und in späteren Abschnitten näher beschrieben.

Das **Texas German Dialect Project (TGDP)** umfasst das gesamte Forschungsprojekt mit all seinen Bestandteilen. Dazu gehören auch Datensammlungen, die nicht aus Texas stammen (z. B. die Huffines-Sammlung zu den Pennsylvaniadeutschen und die Dux-Sammlung zu den Wisconsindeutschen), sowie Öffentlichkeitsarbeit wie Vorträge bei lokalen Geschichtsvereinen. Das **Texas German Dialect Archive (TGDA)** ist die vollständige Sammlung aller texasdeutschen Materialien, die das TGDP zusammengetragen hat. Dazu gehören schriftliche Daten (Briefe, Tagebücher, Zeitungen usw.) sowie Audio- und Videomaterialien (sowohl die vom TGDP erstellten als auch die dem TGDP gespendeten Materialien).

Das **Texas German Dialect Archive Online (TGDA-Online)** (tgdp.org/dialect-archive/) bezieht sich auf den öffentlich zugänglichen Teil des TGDA, der sowohl vom TGDP selbst zusammengetragen wurde, als auch mehrere kleinere Sammlungen, die dem TGDP gespendet wurden.²⁰ Das TGDA-Online enthält drei Aufnahmetypen: 1) narrative Interviews (transkribiert), 2) Übersetzungsaufgaben nach der Liste von Gilbert (1972) und 3) Übersetzungsaufgaben nach der Liste von Eikel (1954). Weitere Informationen zur TGDA-Online sind in Abschnitt 4 zu finden. Das **TGDP-Korpus** enthält derzeit alle öffentlich zugänglichen, transkribierten narrativen Interviews innerhalb des TGDA-Online. Das **GTGXG-Korpus** („(Glenn) **G**ilbert **T**exas **G**erman“) ist eine Sammlung von Aufnahmen, die Glenn Gilbert in den 1960er Jahren angefertigt hat und die kürzlich an das TGDP gespendet wurde, zusammen mit den dazugehörigen Transkriptionen. Ähnlich wie die vom TGDP gesammelten Daten (siehe Abschn. 3.1) enthält das GTXG-Korpus narrative Interviews und Übersetzungsaufgaben. Sowohl das TGDP-Korpus als auch der narrative

²⁰ Einige wenige Audioaufnahmen aus den Jahren 1989, 1992 und 1994 wurden dem TGDP gespendet und liegen somit außerhalb des allgemeinen Zeitrahmens des Projekts. Informationen zu den Aufnahmejahren können in der ZuMult-Plattform gefiltert werden (siehe Abschn. 5.2).

Teil des GTXG-Korpus sind online über **ZuMult** verfügbar (<https://tgdp-zumult.la.utexas.edu/index.jsp>, siehe Abschn. 5.2).

TGDP - Texas German Dialect Project

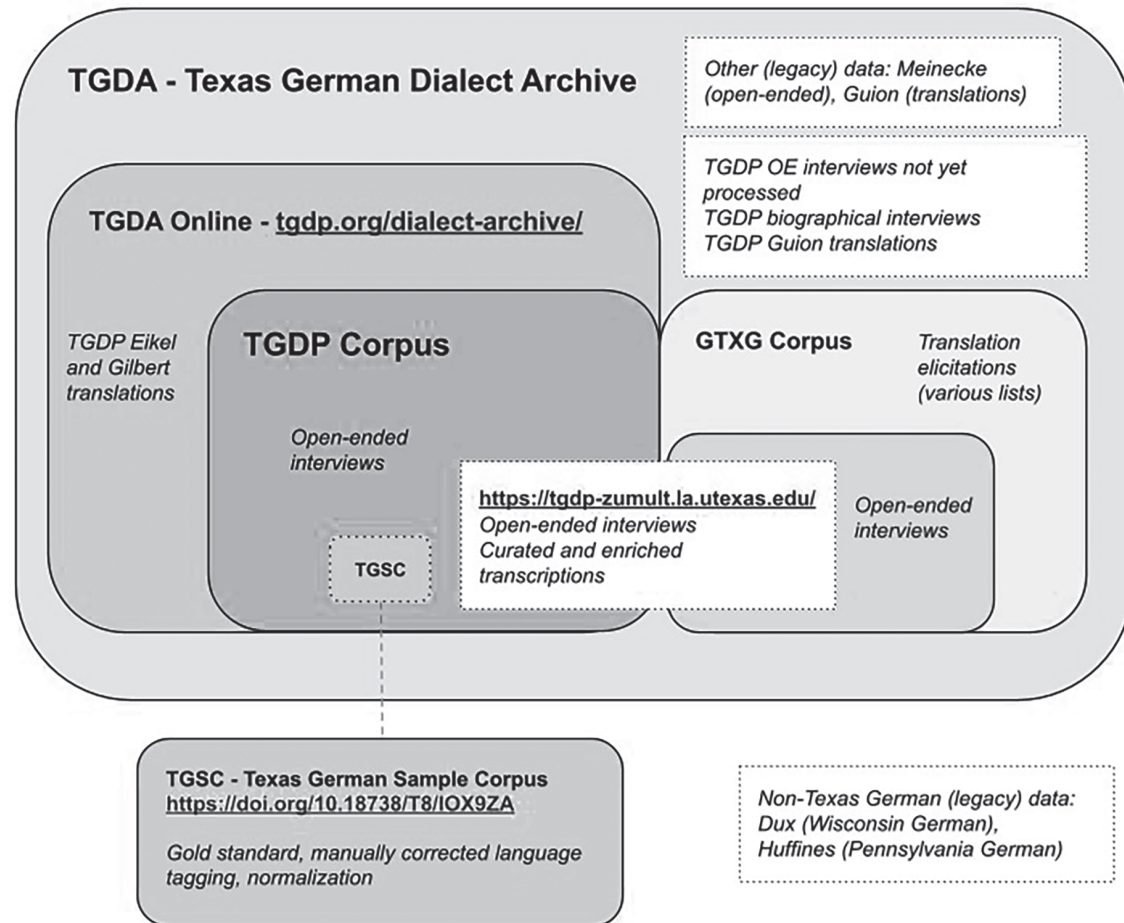


Abb. 3: Ressourcen, die vom Texas German Dialect Project erstellt wurden

Tabelle 2 zeigt, wie viele Stunden an Tonaufnahmen in verschiedenen Teilen des TGDA verfügbar sind. Die ersten drei Abschnitte, gekennzeichnet mit (TGDP), repräsentieren Audiodateien, die vom TGDP selbst erhoben wurden und mehrere kleinere Sammlungen, die dem TGDP gespendet wurden.²¹ Die letzten beiden Abschnitte, gekennzeichnet mit (GTXG), enthalten Audioaufnahmen, die Teil der historischen texasdeutschen Sammlung von Glenn Gilbert sind. Die ZuMult-Plattform umfasst derzeit nur narrative Interviews (transkribiert), während das TGDA-Online sowohl diese Interviews als auch die (nicht-transkribierten) Aufnahmen der Gilbert- und Eikel-Übersetzungsaufgaben umfasst. Die ZuMult-Plattform enthält jedoch auch einen Teil der Audioaufnahmen und Transkripte der historischen Aufnahmen von Glenn Gilbert aus den 1960er Jahren, die er dem Projekt zur Verfügung gestellt hat (GTXG), während TGDA-Online diese Daten nicht beinhaltet.²²

²¹ Siehe Fußnote 17.

²² Einige SprecherInnen des Texasdeutschen wurden im Laufe der Jahre mehrmals interviewt, was erklären kann, warum es mehr offene Interviews gibt als Gewährspersonen in der TGDA. Die Zeitangaben in

25 Jahre Texas German Dialect Project

Teil des TGDPs			Aufnahmetyp TGDA-Online ZuMult		Umfang aller Aufnahmen		Online verfügbar? ²³	
TGDA	TGDA-Online	TGDP Corpus	narrative Interviews (TGDP) ²⁴		~930 Interviews (~310 Stunden)		529 Interviews 6.058 Interviewabschnitte (196 Stunden, 4 Min., 45 Sek.)	
			Übersetzungen (TGDP)	Eikel	~165 Interviews (~40 Stunden)	~1.415 Interviews (~320 Stunden)	~150 Interviews (<i>nicht</i> transkribiert)	Nicht verfügbar
	Gilbert			~825 Interviews (~210 Stunden)	~750 Interviews (<i>nicht</i> transkribiert)			
	Sonstige Übersetzungen (TGDP)		Guion	~390 Interviews (~65 Stunden)	Nicht verfügbar			
			Sonstige	~35 Interviews (~5 Stunden)				
			biografische Fragebogen (normalerweise auf Englisch) (TGDP)		~500 Interviews (~85 Stunden)		Nicht verfügbar	

Tabelle 1 basieren auf folgenden Annahmen: offenes Interview = ~20 Min.; Interviews mit Eikel- bzw. Gilbertsätzen = ~15 Min. (wobei sich einige Interviews überschneiden, da mehrere Personen gleichzeitig interviewt wurden); Guion- und andere Interviews = ~10 Min.; biografische Interviews = ~10 Min. Dies sind recht grobe Schätzungen, da beispielsweise offene Interviews zwischen 10 Min. und über 2 Stunden dauerten, Übersetzungsinterviews zwischen 10 Min. und 1 Stunde und biografische Interviews zwischen 5 Min. und 1 Stunde. Aufzeichnungen der englischsprachigen biografischen Interviews, die Guionsätze-Interviews und die „anderen“ Übersetzungen sind derzeit nicht online verfügbar.

²³ Aus verschiedenen Gründen sind derzeit nicht alle Audioaufnahmen des TGDA-Online verfügbar. Etwa 230 Interviews sind noch nicht online, da sie beispielsweise noch nicht segmentiert wurden. Sobald ein narratives Interview segmentiert worden ist, muss jedes einzelne Segment zunächst transkribiert werden, und personenbezogene Daten, wie z. B. Namen von Personen, müssen unkenntlich gemacht werden, bevor das Segment veröffentlicht werden kann. Darüber hinaus gibt es derzeit mehr als 1.200 Interviewsegmente von narrativen Interviews, die noch nicht transkribiert worden sind. Wir sind aktuell dabei, sie zu transkribieren, und werden jedes Jahr weitere Transkripte veröffentlichen, bis alle Teile der narrativen Interviews transkribiert sind.

²⁴ Obwohl Abbildung 3 die Altdaten als außerhalb von TGDP-Online dargestellt zeigt, sind die Altdaten (narrative Interviews) von Guion, Roesch und Anderson in TGDP-Online enthalten.

Teil des TGDPs			Aufnahmetyp TGDA-Online ZuMult		Umfang aller Aufnahmen		Online verfügbar?	
TGDA	GTXG Corpus		freie Gespräche (GTXG)		10 Stunden, 53 Min.		Nicht verfügbar	49 Inter- views 395 Inter- viewab- schnitte (10 Stunden, 27 Min., 56 Sek.)
			Übersetzun- gen (GTXG)	GA	7 Stunden, 4 Min.	35 Stunden, 59 Min.	Nicht verfügbar	In Vor- bereitung
				GL	25 Stunden, 3 Min.			
				WS	2 Stunden, 25 Min.			
				Sonstige (GM, GS, GE)	1 Stunde, 27 Min.			

Tab. 2: Ein Überblick über die Bestände des Texas German Dialect Projects und deren Verfügbarkeit (Stand: Dezember 2025)

In den folgenden Abschnitten werden TGDA-Online und ZuMult näher erläutert und es wird beschrieben, wie das TGDP die Daten für die Online-Veröffentlichung vorbereitet hat.

4. Das Texas German Dialect Archive Online

Das Texas German Dialect Archive Online (TGDA-Online), welches über die Homepage des TGDP unter www.tgdp.org/dialect-archive abrufbar ist, wurde im Oktober 2002 freigeschaltet.²⁵ Seitdem kann man dort, nach einer einmaligen kostenlosen Registrierung mit Angabe von Namen und Verwendungszweck, die Tonaufnahmen der Interviews abrufen. NutzerInnen des Archivs müssen sich aus Datenschutzgründen bei der Registrierung damit einverstanden erklären, die im Archiv enthaltenen Daten ausschließlich für Lehr- und Forschungszwecke zu verwenden.

²⁵ TGDA-Online war damit – neben z. B. der Datenbank Gesprochenes Deutsch (DGD1, Fiehler/Wagener 2005, jetzt DGD2, Schmidt 2017, <https://dgd.ids-mannheim.de/>), Archiven des Projekts „Dokumentation bedrohter Sprachen“ (DoBeS, Brugman et al. 2002, <https://dobes.mpi.nl/>) am MPI Nijmegen und dem Endangered Languages Archive (ELAR, Austin 2006, Seyfeddinipur et al. 2019, <https://www.elararchive.org/>) – Teil einer ersten Reihe von Projekten, die audiovisuelle Sprachdokumentation in den frühen 2000er Jahren online zugänglich gemacht haben.

Über das TGDA-Online lassen sich drei unterschiedliche Arten von Daten abrufen: 1) narrative Interviews, 2) Übersetzungsdaten, die auf Gilbert (1972) beruhen und 3) Übersetzungsdaten, die auf Eikel (1954) beruhen.²⁶

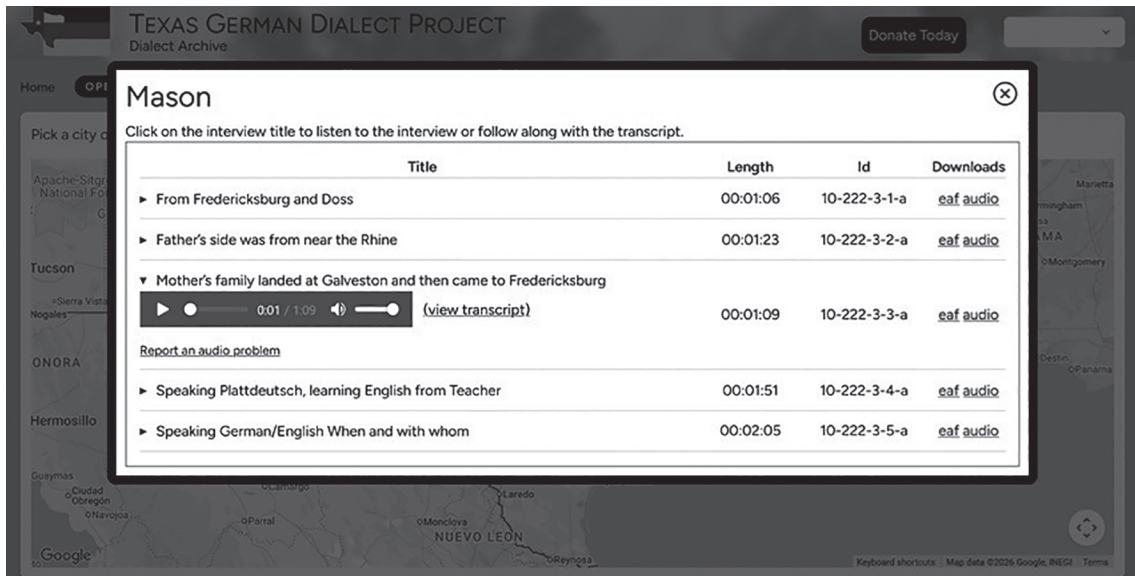


Abb. 4: Zugriff auf die narrativen Interviews im TGDA-Online

Die Tonaufnahmen der narrativen Interviews mit den dazugehörigen Transkriptionen und englischen Übersetzungen kann man abrufen, indem man im TGDA-Online auf den Link „Open-ended Interviews“ klickt. Dieser führt zu einer neuen Webseite mit einer Landkarte von Zentraltexas, auf der alle Orte markiert sind, für die im TGDA-Online narrative Interviews archiviert sind. Wenn man auf einen der Orte klickt, öffnet sich ein neues Fenster mit einer Liste von verfügbaren Interviewsegmenten, wie z. B. in Abbildung 4, wo eine Liste von Interviewsegmenten für den Ort Mason, ca. 50 km nordwestlich von Austin gelegen, dargestellt wird. Die erste Spalte auf der linken Seite in Abbildung 4 listet die Titel der jeweiligen Interviewsegmente auf, wie z. B. „Speaking German with a friend as a kid“ oder „An uncle who refused to speak English to kids“. Die zweite Spalte gibt an, wie lang jedes der Interviewsegmente in Minuten und Sekunden ist, und die dritte Spalte enthält die Identifikationsnummern der Interviewsegmente (vgl. Boas et al. 2010). Die vierte Spalte ganz rechts zeigt an, für welche Interviewsegmente es eine entsprechende Transkriptionsdatei (in EAF-Format) und Audiodatei (in .wav Format) gibt, die heruntergeladen und mit ELAN geöffnet werden können.

Ein Klick auf den Namen eines Interviewsegments, z. B. „An uncle who refused to speak English to kids“ in Abbildung 5, startet einen integrierten Audioplayer, der die Tonaufnahme im Browser abspielt. Wenn man das dazugehörige Transkript lesen möchte, klickt man auf „view transcript“, was ein neues Fenster öffnet, in dem man die gesamte Tonaufnahme abspielen und parallel dazu die Transkription sowie die dazugehörige englische Wort-für-Wort-Übersetzung zeitaligniert mitlesen kann. Jede Zeile des Transkripts ist

²⁶ Die Übersetzungsdaten, die auf Guion (1996) beruhen, sind noch nicht im TGDA-Online veröffentlicht worden.

nummeriert, so dass man bei wissenschaftlichen Arbeiten, die sich auf bestimmte Aufnahmen beziehen, auch immer die jeweils relevante Zeilennummer angeben kann.

The screenshot shows the Texas German Dialect Project (TGDA) website. At the top, there is a header with the project name and a 'Donate Today' button. Below the header is a media player interface with a play button, a progress bar showing 0:00 / 1:23, and a volume icon. The main content area displays a transcript with numbered entries (1-6) and corresponding audio player controls. The transcript includes German text and its English translation.

TEXAS GERMAN DIALECT PROJECT
Dialect Archive

Donate Today

0:00 / 1:23

Father's side was from near the Rhine

1 TRANSCRIPTION_Interviewer_0010
WISSEN Sie wo in Europa Ihre - uh - Ihre Vorfahren herkommen?
Know you where in Europe your - uh - your forefathers from-come?

2
AUS aus welchen Teil von Europe Europa Ihre Ihre Vor - uh - ... - uh - Vorfahren hergekommen sind?
From from which part of Europe Europe your your fore - uh - ... - uh - forefathers from-came had?

3 TRANSCRIPTION_Speaker_0222
ICH weiß dass von mein Papa sein [side] die waren ... bei [die] ... Rhein [river].
I know that of my Papa his side they were ... by the ... Rhine river.

4
UND meine Mama ... ich weiß nicht.
And my mama ... I know not.

5
DIE waren ... in Deutschland und die hamm - uh - Wein gemacht.
They were ... in Germany and they had - uh - wine made.

6 TRANSCRIPTION_Interviewer_0010
HMM.
Hmm.

Abb. 5: Transkript mit Übersetzungen im TGDA-Online

Die zweite Kategorie von Daten, die das TGDA-Online zur Verfügung stellt, sind die vom TGDP erhobenen Übersetzungsdaten, die auf Gilbert (1972) beruhen. Ein Klick auf „Gilbert Interviews“ auf der Archivseite öffnet eine neue Webseite mit einer aus Gilberts *Linguistic Atlas of Texas German* (1972) digitalisierten Landkarte von Zentraltexas, auf der alle Erhebungsorte, an denen das TGDP vor 2021 Aufnahmen gemacht hat, markiert sind, für die das TGDP-Tonaufnahmen der Gilbert-Daten aufgezeichnet hat (siehe Abb. 6). Die einzelnen Tonaufnahmen können mit einem Klick auf eins der 148 englischen Wörter oder eine(n) der Phrasen oder Sätze in der Liste abgerufen werden. Wenn man z. B., wie in Abbildung 6 dargestellt, auf das Wort „Thursday“ klickt, dann öffnet sich über der digitalisierten Landkarte ein neues kleines Fenster mit allen für den jeweiligen Ort verfügbaren relevanten Tonaufnahmen. Ein Klick auf die Nummer einer Gewährsperson spielt dann die Tondatei mit der texasdeutschen Übersetzung ab. So kann man sich

The screenshot shows the Texas German Dialect Project (TGDA) website, specifically the 'Gilbert' section. The header is the same as in Abb. 5. Below the header, there is a navigation bar with 'Home', 'Open-Ended', 'GILBERT', and 'Eikel'. The main content area is divided into two parts. On the left, there is a list of phrases with their corresponding numbers. On the right, there is a map of Texas with various locations marked. The map is titled 'MEXICO' and 'NATION' and shows the state of Texas with its major cities and neighboring states.

TEXAS GERMAN DIALECT PROJECT
Dialect Archive

Donate Today

Home Open-Ended **GILBERT** Eikel

Select a sound phrase and then click a red city on the map.

Phrase

- 1 A hairbrush (1)
- 2 He's running now (1)
- 3 Two goats
- 4 One apple
- 5 Iclies
- 6 The cooking pot
- 7 My head
- 8 The door
- 9 The table is broken
- 10 The animal died out in the pasture (1)
- 11 The animal died out in the pasture (2)
- 12 It is too dry this year
- 13 Water faucet
- 14 Thursday
- 15 He took the most sausage (1)
- 16 A hairbrush (2)
- 17 The door
- 18 Two daughters
- 19 Sweet potatoes (ipomoea batatas) (1)
- 20 Twin cooking pots

MEXICO

Fort Worth Dallas

TEXAS

COAHUILA

NUEVO LEON

LOUISIANA

MISSISSIPPI

Google

Keyboard shortcuts Map data ©2025 Google, INEGI Terms

Abb. 6: Zugriff auf die Gilbert-Daten im TGDA

z. B. für den Ort New Braunfels (zwischen Austin und San Antonio gelegen) von insgesamt 104 Gewährspersonen die Übersetzungen von „Thursday“ anhören, um die Variation bei der Aussprache des Wortes „Donnerstag“ zu untersuchen (siehe Lindemann 2019). Diese Ergebnisse können dann mit den Tonaufnahmen aus anderen Orten in Zentraltexas verglichen werden (vgl. Boas 2006).

Die dritte Kategorie von Daten umfasst die vom TGDP in New Braunfels erhobenen Übersetzungsdaten, welche auf Eikel (1954) beruhen. Auf der Hauptseite des TGDA-Online klickt man auf „Eikel Interviews“ und gelangt so auf eine neue Webseite, von der die texasdeutschen Übersetzungen aller 190 von Eikel auf Englisch verfassten Sätze abgerufen werden können. Ähnlich wie bei dem Abrufen der Gilbert-Daten kann man sich dann für jeden der 190 Sätze eine Liste aller verfügbaren texasdeutschen Übersetzungen von unterschiedlichen Gewährspersonen anzeigen und diese per Mausklick abspielen lassen. Ein Vergleich der unterschiedlichen Übersetzungen der Gewährspersonen erlaubt auch hier Untersuchungen zur sprachlichen Variation im Texasdeutschen von New Braunfels (siehe z. B. Boas 2009a, 2016).

Seit Oktober 2002 werden die texasdeutschen Daten interessierten Forschern, Studierenden und Laien über das TGDA-Online zur Verfügung gestellt, welches über die Projekt-Homepage unter <http://tgdp.org/dialect-archive/> zugänglich ist. Die archivierten texasdeutschen Daten sind bis April 2026 von über 5.200 NutzerInnen aus über 20 Ländern für unterschiedliche Zwecke wie Forschung, Lehre usw. abgerufen worden.

5. Weiterentwicklung der Daten und Zugänge

Im Laufe der 25-jährigen Entwicklung des TGDP und der von den Mitarbeitenden online bereitgestellten texasdeutschen Daten haben sich eine Reihe von Herausforderungen entwickelt. Erstens gab es bis 2024 keine öffentlich zugängliche Möglichkeit, die Transkripte der soziolinguistischen Interviews systematisch mit einer Volltextrecherche zu durchsuchen. Dieses Problem lässt sich u. a. auf das zweite Problem zurückführen, dass sich auf Grund der relativ großen Zahl von über 50 Studierenden, die seit 2001 texasdeutsche Daten transkribiert haben, mehrere Unregelmäßigkeiten bei den Transkriptionen eingeschlichen hatten. Drittens machten es diese Unregelmäßigkeiten fast unmöglich, eine einheitliche Analyse der transkribierten Daten zu entwickeln, die als Basis für eine systematische Volltextrecherche benutzt werden konnten. Dieser Abschnitt beschreibt die einzelnen Schritte, die seit 2022 zur Lösung dieser Probleme in die Wege geleitet worden sind.

5.1 Weiterentwicklung der Daten

5.1.1 Bereinigung und Überführung nach ISO/TEI

ELAN wurde bereits zu Beginn des Projekts 2001 als Lösung für die Erstellung alignierter Transkripte der Audioaufnahmen ausgewählt, sodass alle Transkripte im EAF-Format dieses Tools verfügbar sind. Allerdings haben Änderungen an diesem Dateiformat im Laufe der Jahre und Anpassungen der Transkriptionspraktiken zu einer gewissen Inkonsistenz in den Daten geführt, und die Tatsache, dass ELAN in seinem Datenmodell viele Freiheitsgrade (also verschiedene Möglichkeiten, identische Strukturen auszudrücken) zulässt, schränkt auch die Zuverlässigkeit (*reliability*) der Original-Transkripte für die automatische Verarbeitung ein. Zur Vorbereitung aller weiteren Schritte wurden daher zunächst Vereinheitlichungen an den knapp 6.000 vorhandenen Transkripten narrativer Interview-

abschnitte durchgeführt. Die Annotationsstrukturen und -benennungen folgen nun einem gemeinsamen, klar definierten Schema, und dieses Schema ist zukünftig auch als Vorlage verfügbar, um sicherzustellen, dass neue Transkripte strukturell mit den bereinigten Daten übereinstimmen.

Die bereinigten Daten können automatisch weiterverarbeitet werden, wobei die Wahrscheinlichkeit von Fehlern aufgrund inkonsistenter Benennungen oder Strukturen nun deutlich reduziert ist. In einem ersten Schritt wurden die ELAN-Dateien in den ISO-Standard für Transkriptionen gesprochener Sprache (ISO 2016; Schmidt 2011; Hedeland/Schmidt 2022) konvertiert, der auf den Richtlinien der Text Encoding Initiative (TEI) basiert. In diesem Prozess wurden die Äußerungen der SprecherInnen und InterviewerInnen tokenisiert und englische Übersetzungen sowie der ursprüngliche Transkriptionstext als Annotationen den tokenisierten Äußerungen hinzugefügt. Die gesamte nachfolgende Verarbeitung wird anhand dieser Fassung der Daten durchgeführt.

5.1.2 Hinzufügen und Automatisieren von Token-Annotationen

Frühere Versionen der Daten enthielten nur die Transkriptions- und Übersetzungsebenen als Quelle für systematische Abfragen. Durch Hinzufügen von Tokenisierung, Sprachkennzeichnungen, orthografischer Normalisierung, Lemmatisierung und Part-of-Speech-Tagging als zusätzliche Annotationsebenen eröffnen sich verbesserte Möglichkeiten für korpuslinguistische Analysen. Im Folgenden erklären wir, wie diese Ebenen hinzugefügt wurden.

5.1.2.1 Sprachkennzeichnung, orthografische Normalisierung und das Texas German Sample Corpus

Wie oben erläutert, enthalten die Originaltranskripte zwei Annotationsebenen pro SprecherIn: eine literarische Transkription nach modifizierter deutscher Rechtschreibung und eine Wort-für-Wort-Übersetzung ins Englische. Dies stellt zwar eine ausreichende Grundlage dar, um die Materialien allgemein verfügbar und zugänglich zu machen, doch die hohe orthografische Variabilität erschwert die Durchführung präziser linguistischer Analysen in zweierlei Hinsicht. Einerseits kann es schwierig sein, z. B. ein bestimmtes Lexem zu finden, wenn nur literarische Transkripte durchsuchbar sind, da es dafür notwendig wäre, alle möglichen Schreibweisen eines bestimmten Lexems vorherzusagen. Beispielsweise gibt es mindestens zehn verschiedene orthografische Realisierungen des Lexems *gehabt*: *gehab*, *ghab*, *gehap*, *ghat*, *gahab*, *gehat*, *gehad*, *gehabt*, *gahabt* und *ehabt*.²⁷ Andererseits führt die Verwendung (halb)automatischer Verarbeitungswerkzeuge zum Hinzufügen von Informationen wie Lemma und Wortart oft zu relativ vielen Fehlern, da diese Tools in der Regel mit (schriftlichen) standardnahen Daten wie Zeitungstexten trainiert werden und daher keine Einträge für nicht-standardisierte Schreibweisen in ihrem Lexikon enthalten.²⁸ Um die Transkriptionen sowohl manuell (durch Menschen) als auch

²⁷ Ein weiteres Beispiel: Wenn man nicht wüsste, dass im Texasdeutschen „wir“ oft als „mir“ ausgesprochen wird, und einfach nach <wir> suchen würde, um alle Vorkommen der ersten Person Plural zu finden, würde man versehentlich Hunderte von Tokens übersehen.

²⁸ Als Westpfahl/Schmidt (2013, S. 140) einen automatischen Part-of-Speech-Tagger mit literarischen Transkriptionen (gemäß den cGAT-Richtlinien) testeten, kam es zu einer hohen Fehlerquote, da es für die transkribierten gesprochenen Formen keine lexikalischen Einträge im Tagger gab.

automatisch (durch Computer)²⁹ leichter durchsuchbar und verarbeitbar zu machen, entwickelte Blevins (2022) daher ein orthografisches Normalisierungssystem für gesprochene Kontaktsprachdaten, wobei sie Texasdeutsch als Fallstudie verwendete.³⁰

Zur Evaluierung der vorgeschlagenen Richtlinien zur Sprachkennzeichnung und Normalisierung wurde das Texas German Sample Corpus (TGSC) erstellt. Das TGSC basiert auf einer Zufallsstichprobe der ersten 600 SprecherInnen des Texasdeutschen, die vom TGDGP befragt wurden, mit dem Ziel, die Proportionalität in Bezug auf Geburtsort (County) und Geschlecht zu wahren.³¹ Das resultierende Korpus umfasst 13 Stunden und 35 Minuten an Audioaufnahmen, 75.604 Token (basierend auf der „Norm“-Ebene) und 162 SprecherInnen des Texasdeutschen aus 28 Geburts-Counties. Gemäß den in Blevins (2022) dargelegten Richtlinien wurden 40% der transkribierten Token im TGSC in der Normalisierungsebene geändert, und die Rechtschreibvarianten wurden um etwa 25% reduziert. Das TGSC kann im XML-basierten EXB-Format von EXMARaLDA (Schmidt/Wörner 2014) von der *Texas Data Repository* heruntergeladen werden.³²

5.1.2.2 Automatisierung von Sprachkennzeichnung und Normalisierung

Das manuelle Hinzufügen einer Sprachen- und Normalisierungsebene zu den transkribierten Daten ist sehr zeitaufwendig; angesichts der begrenzten Ressourcen des Projekts und seiner Priorität auf der Transkription weiterer Aufnahmen ist eine Ausweitung dieser manuellen Arbeit auf das gesamte Korpus nicht durchführbar. Das TGDGP hat daher nach Möglichkeiten gesucht, den Prozess der Sprachkennzeichnung und -normalisierung zu automatisieren, wobei das TGSC als Ausgangspunkt („Goldstandard“) sowohl für die Entwicklung der automatisierten Methoden als auch für deren Evaluation herangezogen wurde. Die Automatisierung der Sprachkennzeichnung und Normalisierung bedeutet, einen Prozess zu definieren, der transkribierte Formen als Eingabe nimmt und ihnen automatisch eine Sprachkennzeichnung und eine normalisierte Form zuweist. Die Fehlerquote kann berechnet werden, indem 1) der Prozess auf einen Teil des Goldstandards angewendet wird, 2) die automatisch zugewiesenen Formen mit den manuell zugewiesenen verglichen werden und 3) alle Abweichungen als Fehler gezählt werden.

In einem ersten Schritt wurde die Aufgabe vereinfacht, indem gegenüber Blevins (2022) a) die Sprachkennzeichnungen auf vier Kategorien reduziert wurden („deu“ für Deutsch, „eng“ für Englisch, „amb“ für mehrdeutig und „xxx“ für nicht-lexikalische Formen wie Häsimationsmarker) und b) die Struktur der Normalisierungen vereinfacht wurde, indem nur 1:1-Beziehungen zwischen transkribierter und normalisierter Form zugelassen wurden (z. B. *hamma* → *haben wir*, während solche Kontraktionen zuvor 1:n-Beziehungen waren). Anschließend wurden die vorhandenen TGDGP-Transkripte in einer mehrstufigen Pipeline tokenisiert und jedes Token mit einer Sprachkennzeichnung versehen und orthografisch normalisiert (für eine detaillierte Beschreibung des Sprachkennzeichnungs- und Normalisierungsprozesses, siehe Boas et al. im Dr.). Ein Vergleich mit dem manuell anno-

²⁹ Siehe Blevins (2022, S. 25 f.).

³⁰ Dieses System basiert auf zehn bestehenden, ähnlichen Systemen (Blevins 2022, S. 41–62). Für eine Zusammenfassung des gesamten Sprachkennzeichnungs- und orthografischen Normalisierungssystems von Blevins siehe Blevins (2022, S. 688–695).

³¹ Siehe Blevins (2022, S. 664–668, 670–681) für weitere Informationen zur Zielgruppe und zum Stichprobenverfahren.

³² Siehe <https://doi.org/10.18738/T8/IOX9ZA>.

tierten TGSC (dem „Goldstandard“) zeigt, dass diese einfache Heuristik zu einer Fehlerquote von 6% bei der Sprachkennzeichnung und einer Fehlerquote von 7% bei der Normalisierung führt.

5.1.2.3 POS-Tagging und Lemmatisierung

Lemmatisierung und Part-of-Speech-Tagging erweitern die Möglichkeiten systematischer Abfragen im Korpus zusätzlich. Für die Lemmatisierung und das POS-Tagging wurde TreeTagger verwendet (Schmid 1994). Als Ausgangspunkt für diese Aufgabe wurde die Normalisierungsebene gewählt. Während sie für schriftliche Sprachkorpora zum Standard gehören und erfolgreich auf einsprachige, standardnahe gesprochene Sprachen übertragen wurden (z.B. Schmidt 2016a, b), stellen die Lemmatisierung und das POS-Tagging von Kontaktvarietätsdaten wie denen des Texasdeutschen eine besondere Herausforderung dar.

Für die deutschen Token wurde die Parameterdatei für das Stuttgart-Tübingen-Tag-Set (STTS 2.0, Westpfahl/Schmidt 2016) verwendet. In einem Nachbearbeitungsschritt wurden alle auf der Sprachebene als „eng“ markierten Token mit dem POS-Tag „FM“ (für: „Fremdsprachliches Material“) versehen. In einem zweiten Durchlauf wurde die mit dem PENN-Tagset (Santorini 1990) trainierte englische Parameterdatei verwendet, jedoch nur die Lemmata in den endgültigen Daten beibehalten. Auf diese Weise folgt die POS-Ebene konsequent den Richtlinien des STTS 2.0-Tagsets, das für deutschsprachiges Material entwickelt wurde. Alles, was nicht deutsch ist (oder „amb“ – mehrdeutig), wird als FM markiert, aber die sprachensitive Lemmatisierung ermöglicht weiterhin das Abrufen von Wörtern über ihre Grundformen. Für die im Dezember 2025 veröffentlichte zweite Version der Korpusplattform wurden zusätzlich alle (d.h. deutsche STTS- und englische PENN-) POS-Tags auf das Universal Dependencies (UD) POS-Tagset³³ abgebildet. Damit steht nun auch eine sprachübergreifend nutzbare POS-Annotation für Abfragen zur Verfügung.³⁴ Abbildung 9 veranschaulicht den Annotationsprozess, einschließlich der Tokenisierung, Normalisierung, und Lemmatisierung, und des Sprachzuweisungsprozesses.

original	SIE ham ham gegessen und die die gerollt.								
tok	sie	ham	ham	gegessen	un	die	die	gerollt	.
lang	deu	deu	eng	deu	deu	deu	eng	amb	
norm	sie	haben	ham	gegessen	und	die	die	gerollt	.
POS (STTS 2.0)	PPER	VAFIN	FM	VVPP	KON	ART	FM	FM	
POS (UD)	PRON	VERB	NOUN	VERB	CCONJ	DET	NOUN	VERB	
lemma	sie	haben	ham	essen	und	die	die	roll	
eng	They have ham eaten and the die rolled.								

Abb. 7: Neue Annotationsebenen seit Dezember 2025

³³ <https://universaldependencies.org/u/pos/> (Stand: 18.6.2026).

³⁴ Bei dem Beispiel in Abbildung 7 handelt es sich um ein fiktives Beispiel, das erstellt wurde, um mehrere Herausforderungen für die POS-Aufgabe in einem einzigen Beispiel zu kombinieren: deutsch/englische Homographen (*ham* ‚haben (Deutsch)‘ vs. *ham* ‚Schinken (Englisch)‘, *die* ‚die (Deutsch)‘ vs. *die* ‚Würfel (Englisch)‘) und gemischte Wörter (*gerollt* ‚gewürfelt‘).

Die Auswertung der Ergebnisse der automatischen STTS 2.0 POS-Tagging-Analyse anhand des Goldstandards des TGSC ergibt eine Fehlerquote von 14,6%. Dieser Wert liegt deutlich über den für das FOLK-Korpus angegebenen 5%, ist aber für Korpusabfragen in einem qualitativen Ansatz, d. h. bei denen es eher darum geht, gute Beispiele zu finden als um größere, vollständige Sätze von echten Positiven, immer noch akzeptabel. Die Verwendung der Normalisierungs- und Sprachkennzeichnungsebenen als Grundlage für die Lemmatisierung und das POS-Tagging bedeutet, dass Fehler aus der automatischen Normalisierung möglicherweise auf die beiden anderen Ebenen übertragen werden. Wird das POS-Tagging auf einer manuellen (d. h. weitestgehend fehlerfreien) Normalisierung ausgeführt, reduziert sich die Fehlerquote auf 11,6%.

5.2 Grundlage für vergleichende Sprachinselforschung: Die TGDP-Daten in der ZuMult-Plattform

Im November 2024 wurden als Ergebnis einer Zusammenarbeit zwischen dem TGDP, LAITS und linguisticbits.de die so aufbereiteten und erweiterten Daten erstmals über die Korpusplattform ZuMult zugänglich gemacht, um, über die Möglichkeiten des TGDA-Online hinaus, zusätzliche Abfrage- und Darstellungsmethoden zur Verfügung zu stellen (<https://tgdp-zumult.la.utexas.edu/>). Mit dem zweiten Release im Dezember 2025 stehen auf dieser Plattform neben zusätzlichen narrativen Interviews aus dem TGDP-Projekt nun auch die von Glenn Gilbert in den 1960er-Jahren erhobenen Daten zum Texasdeutschen bereit, die die Analysemöglichkeiten um eine diachrone Dimension erweitern (Schmidt et al. 2026).

ZuMult ist eine Open-Source-Architektur für audiovisuelle Sprachkorpora, die zwischen 2018 und 2022 in einem gemeinsamen Projekt des Leibniz-Instituts für Deutsche Sprache (IDS), des Herder-Instituts der Universität Leipzig und des Hamburger Zentrums für Sprachkorpora (HZSK) entwickelt wurde. Die Architektur ist quelloffen verfügbar und so gestaltet, dass sie adäquat an unterschiedliche Formen der Datenhaltung im Backend und unterschiedliche Nutzungserfordernisse im Frontend angepasst werden kann (Fandrych et al. (Hg.) 2023). Über ZuMult sind mittlerweile als ein Teil der Korpora des Archivs für Gesprochenes Deutsch insbesondere auch Korpora weiterer extraterritorialer Varietäten wie Deutsch in Namibia (Zimmer et al. 2020), Deutsch in Australien (Clyne 1981), Unserdeutsch (Maitz/Volker 2017) und Mennonitenplautdietsch (Kaufmann/Gorisch/Schmidt 2023) (siehe <https://zumult.ids-mannheim.de/>) und auch die umfangreichen Daten der Études Sociolinguistiques à Orléans (Eshkol-Taravella et al. 2012; Schmidt 2025) verfügbar. Weitere Implementierungen entstehen aktuell an der Universität Duisburg-Essen (multimodale Korpora, Ferger/Krause/Pitsch 2023) und an der Universität Hamburg. Die Tatsache, dass verteilt vorliegende Daten auf diese Weise auf einer gemeinsamen technischen Grundlage bereitgestellt werden können, vereinfacht kontrastive Analysen zwischen extraterritorialen Varietäten in verschiedenen Konstellationen, siehe die Vorschläge von Boas (2025) zur vergleichenden Sprachinselforschung, ebenso wie einen Abgleich mit Referenzkorpora für das deutschsprachige Kerngebiet (etwa FOLK oder Deutsch Heute)³⁵.

³⁵ Siehe Boas (2025) zum Einsatz unterschiedlicher Korpora für die systematische vergleichende Sprachinselforschung.

Eine umfassendere Darstellung des ZuMult-Systems findet sich in Fandrych et al. (Hg.) (2023). Boas/Schmidt/Blevins (2026) illustrieren die Nutzung der Plattform für die TGD- Daten im Detail. Schmidt/Ferger (2025) fokussieren auf die technischen Grundlagen der Architektur. Schmidt (in Vorb.) beinhaltet eine kompakte Gesamtdarstellung von ZuMult. Mit Verweis auf diese Publikationen sei hier daher nur ein kurzer Einblick in die wesentlichen Funktionalitäten der TGD-Instanz der Plattform gegeben.

5.2.1 Browsing

Für den Einstieg in die Daten stellt die Plattform eine Übersichtsseite bereit, die es erlaubt, Interviews nach Aufnahmeort, Aufnahmejahr, InterviewerIn und Gewährspersonen zu filtern. Zu jedem Interview können dann erstens die Metadaten zur Aufnahme und den daran beteiligten SprecherInnen angezeigt werden. Zweitens werden die betreffenden thematischen Abschnitte aufgelistet, und die zugehörigen Audioaufnahmen und Transkripte können abgerufen werden (siehe Abb. 8).

The screenshot displays the ZuMult platform interface. At the top, there is a navigation bar with a map icon, the text "Browse the GTXG Corpus", a "Navigation" dropdown, and a "Help" link. Below this, the main content area is divided into two columns. The left column, titled "Interviews", contains a filter panel with dropdown menus for "Interview Location" (Any), "Recording Year" (1965 (9)), "Interviewer" (Any), and "Speaker" (Any), along with a "Filter" button. Below the filter panel is a list of interview entries, each with a speaker icon, ID, name, and date. The right column, titled "Interview GTXG_E_0009", shows the details for a specific interview. It includes the title "Austin (1964-05-15)" and a list of speakers: GTXG S 022, GTXG S 026, and GTXG S 001. Below this is a table of metadata for the selected speaker (GTXG S 022). The table has two columns: the attribute name and its value. The attributes include Gender (female), Birth Year (1906), Age at Interview (58), Approximate Age (58), Birth Location (City) (Near Lincoln), Birth Location (County) (Lee), Birth Location (State) (Texas), Birth Location (Country) (USA), Residence Location (City) (Austin), Residence Location (County) (Not specified), Residence Location (State) (Texas), Residence Location (Country) (USA), Other Locations (Not specified), Religious Affiliation (Not specified), Role (Informant), Language(s) (Not specified), Speaker ID in Gilbert 1972 (X0015), and Speaker ID in Gilbert 1963 (Not specified). Below the metadata table is a section titled "Audio & Transcripts" which contains three audio player cards. Each card has a title, a description, a speaker icon, a play button, a progress bar, and a volume icon. The first card is titled "The speaker's childhood and early schooling" with the ID GTXG_E_009_SE_OE_A_0001. The second card is titled "The speaker's experience growing up on a farm, discussing her parents" with the ID GTXG_E_009_SE_OE_A_0002. The third card is titled "The speaker discusses her father" with the ID GTXG_E_009_SE_OE_A_0003. Each card also has buttons for "Transcript" and "Annotations".

Attribute	Value
Gender	female
Birth Year	1906
Age at Interview	58
Approximate Age	58
Birth Location (City)	Near Lincoln
Birth Location (County)	Lee
Birth Location (State)	Texas
Birth Location (Country)	USA
Residence Location (City)	Austin
Residence Location (County)	Not specified
Residence Location (State)	Texas
Residence Location (Country)	USA
Other Locations	Not specified
Religious Affiliation	Not specified
Role	Informant
Language(s)	Not specified
Speaker ID in Gilbert 1972	X0015
Speaker ID in Gilbert 1963	Not specified

Abb. 8: Auswahl von Interviewsegmenten mit ZuMult

5.2.2 Query

Das für korpuslinguistische Herangehensweise wichtigste Element der Plattform ist die Query-Seite. Dort können Suchanfragen auf allen Annotationsebenen der Korpora ausgeführt werden. Eine einfache Suchanfrage besteht aus einfachen Zeichenketten, die mit den normalisierten Formen abgeglichen werden. Beispielsweise findet die Suchanfrage *bin gegangen* Belege für die literarisch transkribierten Formen *bin gegang*, *bin gegange* und *bin gegangen*. Wesentlich komplexere Suchanfragen können mit der CQP-Anfragesprache formuliert werden. Diese erlaubt es, die Suche auf die verschiedenen Annotationsebenen zuzuschneiden. So findet [word="gehen"] nur solche Formen, die literarisch als *gehen* transkribiert wurden, [norm="gehen"] findet Formen, die zu *gehen* normalisiert wurden (also etwa auch *gehn*), und [lemma="gehen"] findet alle Formen der Grundform *gehen* (also: *gehst*, *ging*, *gegangen* etc.). Entsprechende Anfragen können auch auf der Sprachebene (z. B. [lang="eng"]) und auf der Part-of-Speech-Ebene ausgeführt werden, z. B. [pos="VVFIN"] für alle finiten Vollverben oder [udpos="ADJ"] für eine sprachübergreifende Suche nach Adjektiven.

Anfragen können weiterhin zum einen für ein einzelnes Token kombiniert werden: [norm="der" & pos="PDS"] findet beispielsweise nur solche Instanzen der normalisierten Form *der*, die als Demonstrativpronomen (und nicht als Artikel) gekennzeichnet sind. Zum anderen können durch Aneinanderreihen solcher Ausdrücke auch Token-Abfolgen abgefragt werden, wobei einzelne Bestandteile auch als optional markiert werden können. Der Ausdruck [pos="ART"][pos="ADJA"][pos="NN"] findet zum Beispiel Artikel-Nomen-Abfolgen mit optional dem Nomen vorangestellten Adjektiv. Weiterhin können über reguläre Ausdrücke in den Attributwerten auch ganze Zeichenmuster (statt fixer Zeichenketten) abgefragt werden. So findet [norm="ver.+en" & pos="ADJ."] diejenigen normalisierten Tokens, die mit „ver-“ beginnen und mit „-en“ enden und als Adjektiv (ADJA oder ADJD) getaggt sind.

Schließlich lassen sich in eine CQP-Anfrage auch Annotationsebenen einbeziehen, die nicht auf einzelne Tokens bezogen sind. Die Anfrage <translation="*.always.*"/> containing [norm="immer"] findet etwa alle Äußerungen mit der normalisierten Form *immer*, deren englische Übersetzung das Wort *always* enthält. Ähnlich können Metadaten in eine Suchanfrage einbezogen werden: [norm="in"] within <Interview_Location="New Braunfels"/> beschränkt beispielsweise die Anfrage nach der normalisierten Form *in* auf Aufnahmen, die in New Braunfels gemacht wurden.

Das Ergebnis einer CQP-Suchanfrage wird als Keyword-in-Context-Konkordanz (KWIC) mit dem Treffer in der Mitte und einigen Wörtern des linken und rechten Kontexts angezeigt (siehe Abb. 9). Aus der KWIC heraus kann das zugehörige Audio abgespielt, zusätzlicher Transkriptkontext abgerufen oder der Treffer in der interaktiven Transkriptansicht (siehe 5.2.3) angezeigt werden. Zur weiteren Verarbeitung (z. B. Quantifizierung) kann die KWIC auf den eigenen Computer heruntergeladen und dort mit einer Tabellenkalkulation (wie MS Excel) weiterverarbeitet werden.


Query the **TGDP** and **GTXG** Corpus
Navigation ▾
Help

CQP query:
🔍
📄

☒ **TGDP**
☒ **GTXG**
- 15 + tokens in context
☒ Restrict query to informants

Query intro ▾

Total: 631 in 484 documents.

First Previous **1** 2 3 4 5 Next Last

1	1-59-1-18-a	Speaker_0059	... uhh mein Mann uh un die Eltern ham Haus da gehabt . UN mir ham von	die gerent	denn . UN uh uh das war nich mehr so . ABER was interessant war EIN Tag ...	▶	🔍
2	1-59-1-19-a	Speaker_0059	is uh die ham wo mir frieher gewohnt ham . UH unser Platz war grad an	den Comal	was er ka- gekauft hat nach mei Vater gestorben is wo mein wo meine Eltern ...	▶	🔍
3	1-59-1-21-a	Speaker_0059	... Paar Jahre zurick , da kam ' ne Frau von Jordanton Texas un uh die kam nach	die archives	da . UN uhh sie wollt hat uh Rezepte gehabt . UND sie war sehr interessiert in ...	▶	🔍
4	1-59-1-24-a	Speaker_0059	... hab ich gsagt mir missen da aber eins finden was nich so viel schmoke hat .	DIE schmoken	schrecklich da drieben . DAS war alles . MA konnt kaum sehen . UN laugh das is wahr . DAS ...	▶	🔍
5	1-27-1-2-a	Speaker_0027	älder geworden sinn , denn musst mir den Vater helfen ... JA mit ah auf die Farm	das corn	- corn einholen und ah ... aahh das Futter balen . MIR hatten uh eins Geschäft gehab . MEIN ...	▶	🔍

Abb. 9: CQP Abfrage mit ZuMult-Plattform

5.2.3 Interaktive Transkriptanzeige

Zum detaillierten Studium eines einzelnen Transkripts kann dieses in einer interaktiven Transkriptanzeige wie in Abbildung 10 abgerufen werden.

Angezeigt werden neben dem eigentlichen Transkripttext inklusive der englischen Übersetzungen eine Lemma-Frequenzliste, ein Density-Viewer, der einen kompakten Überblick über die interaktive Struktur des Interviews gibt, sowie ein Audio-Player. Diese Komponenten sind miteinander synchronisiert, so dass z. B. ein Doppelklick ins Transkript den Audio-Player an der entsprechenden Stelle startet oder ein Klick auf ein Lemma in der Frequenzliste die betreffenden Formen im Transkript hervorhebt.

Die Darstellung kann parametrisiert werden. Beispielsweise können statt der literarisch transkribierten die normalisierten Formen angezeigt werden, was es leichter macht, den Inhalt der Transkription nachzuvollziehen. Ausschnitte des Transkripts können markiert und die zugehörigen Daten zur Weiterbearbeitung auf den eigenen Rechner in unterschiedlichen Formaten heruntergeladen werden.

25 Jahre Texas German Dialect Project

Hail and rain storms
1-51-1-24-a

Navigation Parameters Interview Sections (27) Selection Help

pos=ART

00:00 / 1:56

00:00 00:10 00:20

Letter	Count
A	1
ab	1
aber	1
abschlagen	2
alle	2
also	1
ander	1
ans-	1
anstreichen	1
auch	7
auf	1
aus	1
auswaschen	1
B	
Berg	1
bleiben	1
C	
creeken	1
D	
d	13

0019 von das Haus. OH ja wir hamm oft Unwetter.
from the house. Oh yes we have often bad-weather.

0020 0001 Und wie groß ist der Hagel, so, oder?
And how big is the hail, like-this, or?

0021 0051 NO. WAS war das, vielleicht uh
No. What was that, maybe uh

0022 I gue wahrscheinlich wo der die Farbe abgeschlagen hat
I guess probably where it the color beat-away has

0023 WAR wahrscheinlich so gewesen.
Had probably like-this been.

0024 0507 Wir haben auch so groß gesehen.
We have also this big seen.

0025 0051 JA, das hammir schon gesehen.
Yes, that have-we already seen.

0026 0507 Hamm wir auch so groß gesehn.
Have we also this big seen.

0027 Ja.
Yes.

0028 0051 (???) isses hier nur wie so' ne Erbse.
(???) is-it here only like such a pea.

0029 ABER wir haben schon harter... schlimmer Hagel gehabt.
But we have already hard ... bad hail had.

0030 0507 Auch mal uh Zeit uh
Also one uh time uh

0031 Masse Wasser.
a-lot-of water.

Abb. 10: Interaktive Transkriptanzeige

6. Forschung, Lehre und Outreach

Das TGDA-Online wurde 2001 ursprünglich als Lernressource konzipiert (siehe Boas 2003), damit Studierende der Germanistik an der UT-Austin leichten Zugang zu authentischen gesprochenen Daten des Texasdeutschen bekommen können. Erst die eingehendere Beschäftigung mit der Struktur des Texasdeutschen im Rahmen mehrerer Lehrveranstaltungen legte das Forschungspotenzial der Daten offen. So entstanden auf der Grundlage der im TGDA-Online enthaltenen Daten erste Analysen des Lautinventars des Texasdeutschen (Boas 2002 und Boas et al. 2004) und Analysen der Bildung von Pluralformen von Substantiven (Bathe 2005) und der Distribution von Kasusmarkern (Rybarski 2006), gefolgt von ersten Untersuchungen zum Sprachgebrauch des Texasdeutschen (Boas 2005), zur Implementierungen von sog. „Best practice recommendations“ für die Dokumentation von bedrohten Sprachen und Dialekten (Boas 2006) und zur Distribution und Verwendungsweise von Modalpartikeln (Boas/Weilbacher 2007, 2009).³⁶

Das TGDA-Online hat mehr als 5.200 registrierte Nutzer (Stand April 2026) und wurde in den letzten 25 Jahren außer für den Einsatz in der Lehre an mehreren Hochschulen in Deutschland und den Vereinigten Staaten auch für weitere unterschiedliche wissenschaftliche Arbeiten genutzt. Neben den bereits oben genannten Untersuchungen erforscht z. B.

³⁶ Siehe auch Boas/Weilbacher (2009), Boas (2010), Weilbacher (2011) und Boas (2018b).

Boas (2009b) semantische Faktoren, die zum Kasusabbau im Texasdeutschen beigetragen haben. Die Monografie von Roesch (2012) untersucht, wie sich das Texaselsässische im Medina County (westlich von San Antonio) im letzten Jahrhundert entwickelt hat und warum. Eine Reihe weiterer Beiträge beschäftigt sich mit der Entlehnung von englischen Wörtern (Dux 2017, 2020) und Aspekt-Markern (Blevins 2018) ins Texasdeutsche. Die im TGDA-Online enthaltenen Transkripte und Tonaufnahmen sind nicht nur für Untersuchungen über strukturelle Merkmale des Texasdeutschen verwendet worden, sondern auch für soziolinguistische Untersuchungen: Boas (2015) untersucht, wie sich der Sprachgebrauch bei Texasdeutschen unterschiedlicher konfessioneller Herkunft unterscheidet. Boas/Fingerhuth (2017) analysieren den Einfluss von Sprach- und Sprechereinstellungen auf den Erhalt des Texasdeutschen zu Beginn des 21. Jahrhunderts und Jones (2022) zeigt, welchen Einfluss Spracheinstellungen und Geschlechtszugehörigkeit auf den Erhalt des Texasdeutschen haben.

Eine Zusammenarbeit von Angehörigen des TGDP mit einem der Pioniere der Erforschung des Texasdeutschen, Glenn Gilbert (geboren 1936), führte auch zu einer Reihe von Publikationen zur Geschichte der Sprachwissenschaft des Texasdeutschen (siehe Pierce/Boas/Gilbert 2019; Pierce/Boas 2019; Pierce/Boas/Gilbert 2021 und Boas/Somers/Pierce *eing.*). Darüber hinaus wurden die im TGDA-Online enthaltenen Daten in den letzten Jahren für die Entwicklung eines Normalisierungssystems für Kontaktvarietäten des Deutschen (Blevins 2022) genutzt (siehe Abschn. 5.1), und sie bildeten die Grundlage für Warmuth (2023), der die Entwicklung der Diphthonge im Texasdeutschen ausführlich mit unterschiedlichen Methoden untersucht hat. Schließlich sei noch eine Reihe von Untersuchungen zur vergleichenden Sprachinselforschung genannt, die sich mit der Frage auseinandersetzen, wie bestimmte linguistische Phänomene systematisch korpusbasiert über mehrere Kontaktvarietäten hinweg untersucht und verglichen werden können (siehe Saller 2024; Boas 2025; Reher *in Vorb.*).

Außer dem Einsatz der im TGDA-Online enthaltenen Daten für die Forschung und Lehre ist es auch ein besonderes Anliegen des TGDP, die Angehörigen der texasdeutschen Minderheit sowie der allgemeinen Öffentlichkeit über die Forschungsergebnisse des TGDP zu informieren. Im Sinne des von Wolfram/Reaser/Vaughn (2008) entwickelten Prinzips der sog. „Linguistic Gratitude“ besteht das Anliegen darin, breit angelegte Wissenschaftskommunikation zu betreiben, die die Öffentlichkeit u. a. über Sprachvariation, Sprachideologie und Mehrsprachigkeit im texanischen Kontext informiert. So sollen und können Vorurteile gegenüber Sprachminderheiten effektiv aufgeklärt werden, was wiederum zu einer größeren Akzeptanz nicht nur gegenüber Mitgliedern der Sprachminderheit führen kann, sondern auch zu einem größeren Verständnis von sprachwissenschaftlicher Forschung in der breiteren Öffentlichkeit. Um diese Ziele zu verfolgen, halten Mitglieder des TGDP öffentliche Vorträge bei Sitzungen von historischen Gesellschaften und Heimatvereinen in Zentraltexas. Dies dient dazu, die Öffentlichkeit über die Arbeit des TGDP zu informieren und zur Diskussion über weit verbreitete aber nicht immer wissenschaftlich korrekte Ansichten bzgl. Mehrsprachigkeit, Sprachkontakt und Sprachvariation beizutragen.

Die Aktivitäten des TGDP sind in den letzten 25 Jahren auch regelmäßig von den Medien rezipiert worden. Es erschienen über 50 Zeitungsartikel über das TGDP, 12 Radioreportagen, drei für das Fernsehen produzierte Sendungen und ein Dokumentarfilm. Die im TGDA enthaltenen Daten dienten bis jetzt als Grundlage für über 300 akademische Konferenzvorträge. Das TGDP hat auch einen jährlich herausgegebenen Rundbrief, der an alle

vom TGDG aufgenommenen SprecherInnen sowie andere am Texasdeutschen interessierte Personen verschickt wird.³⁷ Dieser Rundbrief klärt auf der Basis der TGDA-Daten über linguistische und historische Aspekte des Texasdeutschen auf, liefert Informationen über die vom TGDG erreichten Meilensteine und informiert über die Forschungs- und Lehraufgaben der Mitglieder des TGDG.

7. Andere Datentypen, künstliche Intelligenz, interdisziplinäre Forschung

Im Jahr 2021 hat das TGDG angefangen, die Dokumentations-, Archivierungs-, und Forschungsaktivitäten über Tonaufnahmen gesprochener Sprache hinaus zu erweitern. Dies betrifft zum einen aus dem 19. und frühen 20. Jahrhundert stammende, handgeschriebene Briefe auf Deutsch, Tagebücher und andere Texte von Vorfahren, die von Gewährspersonen zu Hause aufbewahrt worden sind. Parallel gibt es in Zentraltexas auch zahlreiche historische handgeschriebene Briefe und deutsch-texanische Zeitungen in Dutzenden von Archiven und Bibliotheken. Mitglieder des TGDG haben angefangen, diese zu digitalisieren, um sie systematisch untersuchen zu können (siehe Boas/Levina 2024). Die historischen Texte bzw. deren Inhalte haben zu neuen Forschungsaktivitäten des TGDG in zwei Bereichen geführt.

Erstens fiel auf, dass viele der Themen, Ereignisse und Personen, die in den narrativen Interviews mit heutigen SprecherInnen des Texasdeutschen vorkommen, sich auch in den historischen Briefen finden lassen. Dies sollte eigentlich nicht verwundern, da sich ja viele Gesprächsthemen der narrativen Interviews um die Erfahrungen der Vorfahren bei der Einwanderung nach und der Besiedlung von Texas drehen. Um die Zusammenhänge zwischen den in den narrativen Interviews und den historischen Briefen genannten Themen näher zu untersuchen, und um festzustellen, wie sich die Narrative zu Anfang des 21. Jahrhunderts von denen im 19. Jahrhundert strukturell und inhaltlich unterscheiden, hat sich eine kleine interdisziplinäre Gruppe aus Forschenden der germanistischen Sprachwissenschaft und der Geschichtswissenschaft gebildet. Ziel dieses Vorhabens ist es, systematisch zu vergleichen, wie bestimmte Themen, Ereignisse und Personen unterschiedlich dargestellt und bewertet werden (und warum).

So hat z. B. eine erste interdisziplinäre Pilotstudie (Weiss et al. 2025) eine Reihe unterschiedlicher historischer Textquellen untersucht, um zu sehen, wie der Begriff „Indianer“ (mit welchen Bedeutungen) in unterschiedlichen Kontexten im texasdeutschen Diskurs des 19. Jahrhunderts verwendet wurde. Die Ergebnisse dieser historisch-linguistischen Untersuchung wurden mit einer Analyse der Kontexte in den heutigen narrativen Interviews des TGDG verglichen, in denen auch das Wort „Indianer“ erwähnt wurde. So stellt sich z. B. heraus, dass in vielen historischen texasdeutschen Quellen der Begriff „Indianer“ mit einer eher negativen und exotischen Bedeutung belegt ist, während er in einigen narrativen Interviews des TGDG auch als Opfer der weißen Siedlungs- und Vertreibungs politik charakterisiert wird. Die Ergebnisse dieses Pilotprojekts bilden die Grundlage für die weitere interdisziplinäre Erforschung bzw. den Vergleich von geschriebenen historischen mit gesprochenen aktuellen texasdeutschen Daten, um so zu erkunden, wie sich das Texasdeutsche seit der Mitte des 19. Jahrhunderts entwickelt hat.

³⁷ Siehe <https://tgdp.org/tgdp-newsletter/> (Stand: 18.6.2026).

Zweitens hat das TGDG im Jahr 2022 eine Kooperation mit dem am Deutschen Historischen Institut in Washington, D.C. angesiedelten Projekt „Migrant Connections“ begonnen.³⁸ Dieses Projekt beschäftigt sich mit der Einwerbung und Archivierung von handgeschriebenen Briefen, die im 19. und 20. Jahrhundert von Deutschland nach Nordamerika geschickt worden sind. Parallel zu den an der Forschungsbibliothek in Gotha archivierten Auswandererbriefen aus Nordamerika³⁹ hat das Projekt „Migrant Connections“ das Ziel, die aus Deutschland stammenden Briefe mit Hilfe von Freiwilligen zu transkribieren (und ins Englische zu übersetzen; „Citizen Science“), um sie so für Forschungs- und Lehrzwecke zur Verfügung zu stellen. Seit 2022 gibt das TGDG die von Deutschland nach Texas geschickten Briefe, die digitalisiert worden sind, an das Projekt „Migrant Connections“ weiter, damit diese auch mit in die Sammlung aufgenommen werden können.

Im Kontext dieser Kooperation entstand die Idee zu einer zweiteiligen Untersuchung zur 1) Struktur des Texasdeutschen im 19. Jahrhundert und 2) Erfassung und Modellierung von Themen und Inhalten, die in den Briefen erwähnt werden. Diese zwei Ideen bildeten im Frühling 2024 die Grundlage für einen Projektantrag, der bei der Volkswagenstiftung unter der Förderlinie „Neue Horizonte entdecken“ zusammen von Mitgliedern des Deutschen Historischen Instituts und des TGDG eingereicht und im Dezember 2024 bewilligt wurde. Ziel des im Juni 2025 begonnenen Projekts ist es, mit Hilfe künstlicher Intelligenz auf der Basis der aus dem 19. Jahrhundert stammenden handgeschriebenen Briefe (von Deutschland nach Texas sowie von Texas nach Deutschland) und einer Reihe unterschiedlicher anderer historischer Datenquellen zu deutschen Dialekten wie z. B. Tonaufnahmen aus dem Zwirner-Korpus aus den 1950er Jahren und den Wenkerbögen am Forschungszentrum Deutscher Sprachatlas aus dem 19. Jahrhundert, das gesprochene Texasdeutsch aus den 1880er Jahren zu rekonstruieren.

8. Zusammenfassung und Ausblick

Dieser Beitrag hat gezeigt, wie sich das im Jahr 2001 gegründete Texas German Dialect Project und die aus ihm hervorgegangene digitale Ressource, das Texas German Dialect Archive-Online, in den letzten 25 Jahren entwickelt haben. Das TGDG-Korpus enthält, Stand Dezember 2025, fast 200 Stunden Tonaufnahmen, die den Sprachgebrauch und die Spracheinstellungen der deutschsprachigen Minderheit in Texas dokumentieren. Diese Aufnahmen (mit z. T. dazugehörigen Transkripten) sind aus sprachwissenschaftlicher Sicht aus den folgenden Gründen besonders interessant. Darüber hinaus enthält TGDA-Online ca. 180 Stunden Übersetzungsaufnahmen (nicht transkribiert).

1) Im Gegensatz zu anderen Kontaktvarietäten hat sich das Texasdeutsche nie zu einem einheitlichen Dialekt entwickelt, d. h. die sprachliche Variation ist noch immer extrem hoch (siehe Boas 2016; Boas/Pierce 2025). 2) Joseph Wilson prognostizierte das Aussterben des Texasdeutschen für das Jahr 2036 (Wilson 1986) und diese Vorhersage wird mit an Sicherheit grenzender Wahrscheinlichkeit eintreffen (siehe Boas/Fingerhuth 2017 und Boas/Pierce 2025), d. h. die im TGDA-Online archivierten Daten sind die letzten noch vor dem Aussterben der Kontaktvarietät erhobenen Daten. 3) Das TGDA-Online enthält nicht nur die vom TGDG seit 2001 erhobenen Daten, sondern ermöglicht seit Dezember 2025 auch Zugriff auf die von Glenn Gilbert in den 1960er Jahren erhobenen Daten, die die

³⁸ Siehe <https://www.migrantconnections.org/s/en/page/welcome> (Stand: 18.6.2026).

³⁹ Siehe <https://auswandererbriefe.de> (Stand: 18.6.2026).

Grundlage seines 1972 erschienenen *Linguistic Atlas of Texas German* bilden. Somit erlaubt das TGDA-Online auch diachrone Untersuchungen zur Entwicklung des Texasdeutschen. 4) Das TGDP-Korpus (Version 2.1, Dezember 2025) ist mit fast 1,5 Millionen Token das größte Online-Korpus einer deutschen Kontaktvarietät weltweit und wird kontinuierlich erweitert, bis keine weiteren SprecherInnen des Texasdeutschen mehr gefunden werden können.⁴⁰ 5) Wegen der im Dezember 2024 durchgeführten Integration der texasdeutschen Daten in die ZuMult-Korpusplattform ergeben sich neuartige Möglichkeiten, unterschiedliche deutsche Kontaktvarietäten zu untersuchen und bestimmte Phänomene kontrastiv zu bearbeiten. So lassen sich z. B. dieselben ZuMult Korpusanfragen zum Texasdeutschen und anderen Kontaktvarietäten wie dem Unserdeutschen und dem Namibia-Deutschen stellen. Dadurch ergeben sich für die vergleichende Sprachinselforschung (Rosenberg 2003; Boas 2021, 2025) neue Möglichkeiten, systematisch eine Bandbreite unterschiedlicher Phänomene kontrastiv zu untersuchen.

Literatur

- Austin, Peter K. (2006): Data and language documentation. In: Gippert, Jost/Himmelmann, Nikolaus/Mosel, Ulrike (Hg.): *Essentials of language documentation*. (= Trends in Linguistics 178). Berlin/New York: Mouton de Gruyter, S. 87–112.
- Bathe, Ulrike (2005): *Plural formation in Texas German*. Masterarbeit. Austin, TX: University of Texas at Austin.
- Becker, Kara (2017): The sociolinguistic interview. In: Mallinson, Christine/Childs, Becky/Van Herk, Gerard (Hg.): *Data collection in sociolinguistics: Methods and applications*. 2. Aufl. London/New York: Routledge, S. 91–100.
- Bieseke, Rudolph (1930): *The history of the German settlements in Texas, 1831–1861*. Austin, TX: Von Boeckmann-Jones.
- Blevins, Margaret (2014): *Auxiliary tun in Texas German*. Masterarbeit. Austin, TX: University of Texas at Austin.
- Blevins, Margaret (2018): Towards a constructional analysis of the progressive aspect in Texas German. In: Boas/Höder (Hg.), S. 73–114.
- Blevins, Margaret (2022): *The language-tagging & orthographic normalization of spoken mixed-language data, with a focus on Texas German*. Dissertation. Austin, TX: University of Texas at Austin.
- Boas, Hans C. (2002): The Texas German Dialect Archive as a tool for analyzing sound change. In: Austin, Peter/Aristar Dry, Helen/Wittenberg, Peter (Hg.): *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, Spanien, S. 28.1–28.4.
- Boas, Hans C. (2003): Tracing dialect death: The Texas German Dialect Project. In: Larson, Julie/Paster, (Hg.): *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society, S. 387–398.
- Boas, Hans C. (2005): A dialect in search of its place: The use of Texas German in the public domain. In: Cravens, Craig Stephen/Zersen, David John (Hg.): *Transcontinental encounters: Central Europe meets the American heartland*. Austin, TX: Concordia University Press, S. 78–102.

⁴⁰ Zum Vergleich: Das in der Datenbank Gesprochenes Deutsch am Leibniz Institut für Deutsche Sprache enthaltene DNam-Korpus (Version 2.13 [31-01-2020]) (Zimmer et al. 2020) enthält insgesamt 211.763 Tokens Namibia-Deutsch, das Unserdeutsch Korpus (Version 2.21 [11-01-2024]) (Maitz/Volker 2017) enthält 460.473 Tokens. Für einen Überblick über deutsche Sprachinselporpora, siehe Boas/Fingerhuth (2018).

- Boas, Hans C. (2006): From the field to the web: Implementing best-practice recommendations in documentary linguistics. In: *Language Resources and Evaluation* 40, 2, S. 153–174.
- Boas, Hans C. (2009a): *The life and death of Texas German*. Durham, NC: Duke University Press.
- Boas, Hans C. (2009b): Case loss in Texas German: The influence of semantic and pragmatic factors. In: Barðdal, Jóhanna/Chelliah, Shobhana L. (Hg.): *The role of semantic, pragmatic, and discourse factors in the development of case*. (= *Studies in Language Companion Series* 108). Amsterdam/Philadelphia: Benjamins, S. 347–373.
- Boas, Hans C. (2010): On the equivalence and multifunctionality of discourse markers in language contact situations. In: Harden, Theo/Hentschel, Elke (Hg.): *40 Jahre Partikelforschung*. (= *Stauffenburg Linguistik* 55). Tübingen: Stauffenburg, S. 301–315.
- Boas, Hans C. (2011): Some consequences of case loss in Texas German. In: Yoquelet, Cory (Hg.): *Proceedings of the Twentyninth Annual Meeting of the Berkeley Linguistics Society. Special Session on Minority and Diasporic Languages of Europe*. Berkeley, CA: Berkeley Linguistics Society, S. 43–54.
- Boas, Hans C. (2015): Linguistic splits along religious lines: The role of language maintenance among Catholics and Lutherans in Texas. In: Rosenberg, Peter/Jungbluth, Konstanze/Zinkhahn Rhobodes, Dagna (Hg.): *Linguistic construction of ethnic borders*. [Ausgewählte Vorträge der internationalen Tagung]. Frankfurt a. M./New York: Lang, S. 167–181.
- Boas, Hans C. (2016): Variation im Texasdeutschen: Implikationen für eine vergleichende Sprachinselforschung. In: Lenz, Alexandra N. (Hg.): *German Abroad: Perspektiven der Variationslinguistik, Sprachkontakt- und Mehrsprachigkeitsforschung*. (= *Wiener Arbeiten zur Linguistik* 4). Wien: Vienna University Press, S. 11–44.
- Boas, Hans C. (2018a): Texas. In: Plewnia, Albrecht/Riehl, Claudia Maria (Hg.): *Handbuch der deutschen Sprachminderheiten in Übersee*. Tübingen: Narr, S. 171–192.
- Boas, Hans C. (2018b): A constructional account of the modal particle ‘ja’ in Texas German. In: Boas/Höder (Hg.), S. 253–276.
- Boas, Hans C. (2021): Zur Vergleichbarkeit von Sprachinseldaten: Ein Plädoyer für eine „bottom-up“-Methodologie im Rahmen der Konstruktionsgrammatik und der Frame-Semantik. In: Földes, Csaba (Hg.): *Kontaktvarietäten des Deutschen im Ausland*. (= *Beiträge zur Interkulturellen Germanistik* 14), Tübingen: Narr, S. 66–88.
- Boas, Hans C. (2025): Towards a systematic methodology for comparing extraterritorial German contact varieties. In: Szczepaniak, Renata/Lindenfelser, Siegwalt/Prediger, Angélica (Hg.): *Deutsch als Minderheitensprache in der Welt*. (= *Germanistische Linguistik* 337). Berlin/Boston: De Gruyter, S. 9–49.
- Boas, Hans C./Fingerhuth, Matthias (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. In: *Linguistische Berichte* 249, S. 95–121.
- Boas, Hans C./Matthias Fingerhuth (2018): Deutsche Sprachinselkorpora im 21. Jahrhundert. In: Kupietz, Marc/Schmidt, Thomas (Hg.): *Korpuslinguistik*. (= *Germanistische Sprachwissenschaft um 2020* 5). Berlin/Boston: De Gruyter, S. 151–178.
- Boas, Hans C./Fuchs, Katrin (2018): Zum Einfluss des Standarddeutschen auf das Texasdeutsche im 19. und 20. Jahrhundert: Empirische und methodologische Probleme. In: Lenz, Alexandra N./Plewnia, Albrecht (Hg.): *Variation – Normen – Identitäten*. (= *Germanistische Sprachwissenschaft um 2020* 4). Berlin/Boston: De Gruyter, 283–304.
- Boas, Hans C./Levina, Ekaterina (2024): Zur Dokumentation und Erforschung deutschsprachiger Presstexte in Texas: Welchen Einfluss hatte das Standarddeutsche auf das Texas-Deutsche? In: Földes, Csaba (Hg.): *Auslandsdeutsche Pressesprache in Europa, Asien und Nordamerika*. (= *Deutsch als Fremd- und Fachsprache* 4). Berlin/Boston: De Gruyter, S. 301–320.

25 Jahre Texas German Dialect Project

- Boas, Hans C./Pierce, Marc (2011): Recent lexical borrowings in Texas German. In: Putnam, Michael T. (Hg.): *Studies on German-language islands*. (= *Studies in Language Companion Series* 123). Amsterdam/Philadelphia: Benjamins, S. 129–150.
- Boas, Hans C./Pierce, Marc (2025): Texas German. In: *Yearbook of German-American Studies*. Supplemental Issue 6 (North American Varieties of German), S. 159–190.
- Boas, Hans C./Weilbacher, Hunter (2007): How universal is the pragmatic detachability scale? Evidence from Texas German discourse markers. In: Hoyt, Frederick/Seifert, Nikki/Teodorescu, Alexandra/White, Jessica (Hg.): *Proceedings of the Texas Linguistic Society IX Conference: The Morphosyntax of Underrepresented Languages*. Stanford, CA: CSLI Publications, S. 33–58.
- Boas, Hans C./Weilbacher, Hunter (2009): The unexpected survival of German discourse markers in Texas. In: *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 42, 1, S. 1–15.
- Boas, Hans C./Höder, Steffen (Hg.) (2018): *Constructions in contact: Constructional perspectives on contact phenomena in Germanic languages*. (= *Constructional Approaches to Language* 24). Amsterdam/Philadelphia: Benjamins.
- Boas, Hans C./Schmidt, Thomas/Blevins, Margo (2026): A new corpus platform for the Texas German Dialect Project. In: *Language Resources and Evaluation* 60, article 52. DOI: <https://doi.org/10.1007/s10579-025-09893-6>.
- Boas, Hans C./Somers, Joren/Pierce, Marc (eingereicht): A previously unknown source of 1960s Texas German data. Erscheint in: *Zeitschrift für Dialektologie und Linguistik*.
- Boas, Hans C./Ewing, Karen/Moran, Cheryl/Thompson, Jana (2004): Towards determining the influence of internal and external factors on recent developments in Texas German phonology. In: *University of Pennsylvania Working Papers in Linguistics* 10, 1, S. 47–59.
- Boas, Hans C./Pierce, Marc/Weilbacher, Hunter/Roesch, Karen/Halder, Guido (2010): The Texas German Dialect Archive: A multimedia resource for research, teaching, and outreach. In: *Journal of Germanic Linguistics* 22, 3, S. 277–296.
- Brugman, Henni/Levinson, Stephen/Skiba, Romuald/Wittenburg, Peter (2002): The DOBES archive: Its purpose and implementation. In: Austin, Peter K./Dry, Helen/Wittenburg, Peter (Hg.): *Proceedings of the International LREC Workshop on Resources and Tools in Field Linguistics*. Paris: European Language Resources Association, S. 1–4.
- Campbell, Randolph B. (2003): *Gone to Texas: A history of the Lone Star State*. Oxford/New York: Oxford University Press.
- Clardy, Catherine (1954): *A description and analysis of the German language spoken in New Braunfels, Texas*. Masterarbeit. Austin, TX: The University of Texas at Austin.
- Clyne, Michael (1981): *Deutsch als Muttersprache in Australien: Zur Ökologie einer Einwanderersprache*. (= *Deutsche Sprache in Europa und Übersee* 8). Wiesbaden: Steiner.
- Dux, Ryan (2017): Classifying language contact phenomena: English verbs in Texas German. In: *Journal of Germanic Linguistics* 29, 4, S. 377–428.
- Dux, Ryan (2020): Code-switching and loan translation in German-American. A cognitive-constructional account. In: *Belgian Journal of Linguistics* 34, S. 53–65.
- Eichhoff, Jürgen (1986): *Die deutsche Sprache in Amerika*. In: Trommler, Frank (Hg.): *Amerika und die Deutschen: Bestandsaufnahme einer 300jährigen Geschichte*. Opladen: Westdeutscher Verlag, S. 235–252.
- Eikel, Fred (1949): The use of cases in New Braunfels German. In: *American Speech* 24, 4, S. 278–281.
- Eikel, Fred (1954): *The New Braunfels German Dialect*. Dissertation. Baltimore, MD: John Hopkins University.

- Eshkol-Taravella, Iris/Baude, Olivier/Maurel, Denis/Hriba, Linda/Dugua, Céline/Tellier, Isabelle (2012): Un grand corpus oral ‚disponible‘: Le corpus d’Orléans 1968–2012. In: *Ressources linguistiques libres*, TAL 52, 3/2011, S. 17–46.
- Fandrych, Christian/Schmidt, Thomas/Wallner, Franziska/Wörner, Kai (Hg.) (2023): *Korpora Deutsch als Fremdsprache (KorDaF) 3, 1. Themenschwerpunkt: Zugänge zu mündlichen Korpora für DaF und DaZ: Das ZuMult-Projekt*. Darmstadt: KorDaF.
- Ferger, Anne/Krause, André Frank/Pitsch, Karola (2023): Workflows and methods for creating structured corpora of multimodal human-robot-interaction. In: Cotgrove, Louis/Herzberg, Laura/Lüngen, Harald/Pisetta, Ines (Hg.): *Proceedings of the 10th International Conference on CMC and Social Media Corpora for the Humanities (CMC-Corpora 2023)*, 14–15 September 2023, University of Mannheim, Germany. Mannheim: Leibniz-Institut für Deutsche Sprache (IDS), S. 73–77.
- Fiehler, Reinhard/Wagener, Peter (2005): Die Datenbank Gesprochenes Deutsch (DGD) – Sammlung, Dokumentation, Archivierung und Untersuchung gesprochener Sprache als Aufgaben der Sprachwissenschaft. In: *Gesprächsforschung* 6, S. 136–147.
- Gilbert, Glenn G. (1963): *The German dialect spoken in Kendall and Gillespie counties, Texas*. Dissertation. Cambridge, MA: Harvard University.
- Gilbert, Glenn G. (1965): English loanwords in the German of Fredericksburg, Texas. In: *American Speech* 40, S. 102–112.
- Gilbert, Glenn G. (1972): *Linguistic atlas of Texas German*. Austin, TX: University of Texas Press.
- Guion, Susan (1996): The death of Texas German in Gillespie County. In: Ureland, P. Sture/Clarkson, Iain (Hg.): *Language contact across the North Atlantic*. (= *Linguistische Arbeiten* 359). Tübingen: Niemeyer, S. 443–463.
- Hedeland, Hanna/Schmidt, Thomas (2022): The TEI-based ISO Standard ‘Transcription of spoken language’ as an Exchange Format within CLARIN and beyond. In: Monachini, Monica/Eskevich, Maria (Hg.): *Selected Papers from the CLARIN Annual Conference 2021*. Virtual Event, 2021, 27–29 September. (= *Linköping Electronic Conference Proceedings* 189). Linköping: Linköping University Electronic Press, S. 34–45.
- Jones, Ellen (2022): *Gender roles and language loss: A new perspective from Texas German on language attitudes*. Masterarbeit. Austin, TX: The University of Texas at Austin.
- Jordan, Terry G. (1969): The German Settlement of Texas after 1865. In: *The Southwestern Historical Quarterly* 73, 2, S. 193–212.
- Kaufmann, Göz/Gorisch, Jan/Schmidt, Thomas (2023): Das MEND-Korpus im Archiv für Gesprochenes Deutsch: Entstehung, Möglichkeiten, Grenzen. In: Wolf-Farré, Patrick/Löff Machado, Lucas/Prediger, Angélica/Kürschner, Sebastian (Hg.): *Deutsche und weitere germanische Sprachminderheiten in Lateinamerika: Methoden, Grundlagen, Fallstudien*. (= *MinGLA – Minderheiten Germanischer Sprachen in Lateinamerika* 1). Berlin: Lang, S. 103–147.
- Kloss, Heinz (1977): Die deutsche Schriftsprache bei den Amischen. In: *Deutsch als Muttersprache in Kanada*. Berichte zur Gegenwartslage. (= *Deutsche Sprache in Europa und Übersee* 1). Wiesbaden: Steiner, S. 97–98.
- Lindemann, Luke (2019): When Wurst comes to Wurscht: Variation and koiné formation in Texas German. In: *Journal of Linguistic Geography* 7, 1, S. 33–45.
- Maitz, Péter/Volker, Craig Alan (2017): Documenting Unserdeutsch: Reversing colonial amnesia. In: *Journal of Pidgin and Creole Languages* 32, 2, S. 365–397.
- Nicolini, Markus (2004): *Deutsch in Texas*. (= *Studien und Quellen zur Sprachwissenschaft* 1). Münster: LIT Verlag.

25 Jahre Texas German Dialect Project

- Pierce, Marc/Boas, Hans C. (2019): Where was historical linguistics in 1968 and where is it now? In: Boas, Hans C./Pierce, Marc (Hg.): *New directions for historical linguistics*. (= Brill's Studies in Historical Linguistics 9). Leiden: Brill, S. 1–41.
- Pierce, Marc/Boas, Hans C./Gilbert, Glenn G. (2019): When is a dissertation not a dissertation? On Eikel 1954. In: *Yearbook of German-American Studies* 53, S. 187–196.
- Pierce, Marc/Boas, Hans C./Gilbert, Glenn G. (2021): Fred Eikel and the study of Texas German. In: *Yearbook of German-American Studies* 56, S. 131–148.
- Pierce, Marc/Boas, Hans C./Roesch, Karen (2015): The history of front rounded vowels in New Braunfels German. In: Johannessen, Janne Bondi/Salmons, Joseph C. (Hg.): *Germanic heritage languages in North America: Acquisition, attrition and change*. (= Studies in Language Variation 18). Amsterdam/Philadelphia: Benjamins, S. 117–131.
- Reher, Caroline (in Vorb.). *Satzverknüpfung im Kontakt: Ein Vergleich des Deutschen in Chile, Texas und Namibia*. Dissertation. Dortmund: Technische Universität Dortmund.
- Roesch, Karen A. (2009): *Texas Alsatian: Henri Castro's legacy*. Dissertation. Austin, TX: The University of Texas at Austin.
- Roesch, Karen A. (2012): *Language maintenance and language death: The decline of Texas Alsatian*. (= Culture and Language Use: Studies in Anthropological Linguistics 6) Amsterdam/Philadelphia: Benjamins.
- Rosenberg, Peter (2003): *Vergleichende Sprachinselforschung: Sprachwandel in deutschen Sprachinseln in Russland und Brasilien*. In: *Linguistik Online* 13, 1, S. 273–323.
- Rybarski, James (2006): *On the distribution of cases in present-day Texas German*. Senior honors thesis. Austin, TX: The University of Texas at Austin.
- Saller, Anna (2024): *Der Einfluss sprachlicher Vitalität und des Standards auf die Frequenz der ,tun'-Periphrase in verschiedenen deutschen Sprachminderheiten: Texas, Australien, Namibia und Siebenbürgen*. In: Wildfeuer, Alfred/Franz, Sebastian (Hg.): *Standardisierung und (Re-)Vitalisierung von kleinen Sprachen*. (= Germanistische Linguistik 55, 1). Hildesheim: Olms, S. 275–314.
- Salmons, Joseph C./Lucht, Felicia A. (2006): *Standard German in Texas*. In: Thornburg, Linda L./Fuller, Janet M. (Hg.): *Studies in contact linguistics: Essays in honor of Glenn G. Gilbert*. Frankfurt a. M. u. a.: Lang, S. 165–186.
- Santorini, Beatrice (1990): *Part-of-speech tagging guidelines for the Penn Treebank Project*, Pennsylvania State University. <https://catalog.ldc.upenn.edu/docs/LDC99T42/tagguid1.pdf> (Stand: 27.6.2026).
- Schmid, Helmut (1994): *Probabilistic part-of-speech tagging using decision trees*. In: *Proceedings of the International Conference on New Methods in Language Processing*. Manchester: University of Manchester, S. 44–49.
- Schmidt, Thomas (2011): *A TEI-based approach to standardising spoken language transcription*. In: *Journal of the Text Encoding Initiative* 1. <https://journals.openedition.org/jtei/142> (Stand: 27.6.2026).
- Schmidt, Thomas/Wörner, Kai (2014): EXMARaLDA. In: Durand, Jacques/Gut, Ulrike/Kristoffersen, Gjert (Hg.): *The Oxford handbook of corpus phonology*. (= Oxford Handbooks in Linguistics). Oxford: Oxford University Press, S. 402–419.
- Schmidt, Thomas (2016a): *Good practices in the compilation of FOLK, the Research and Teaching Corpus of Spoken German*. In: Kirk, John M./Andersen, Gisle (Hg.): *Compilation, transcription, markup and annotation of spoken corpora*. *International Journal of Corpus Linguistics* 21, 3, S. 396–418.
- Schmidt, Thomas (2016b): *Construction and dissemination of a corpus of spoken interaction – tools and workflows in the FOLK project*. In: Kupietz, Marc/Geyken, Alexander (Hg.): *Corpus linguistic software tools*. *Journal for Language Technology and Computational Linguistics* 31.1, S. 117–144.

- Schmidt, Thomas (2017): DGD – die Datenbank für Gesprochenes Deutsch: Mündliche Korpora am Institut für Deutsche Sprache (IDS) in Mannheim. In: *Zeitschrift für germanistische Linguistik* 45, 3, S. 451–463.
- Schmidt, Thomas (2025): Représenter et accéder à la parole dans les corpus oraux: diversification et adaptation des méthodes et technologies. In: Kanaan-Caillo, Laya/Dugua, Céline/Gerstenberg, Annette (Hg.): *Représenter la parole: Apports à une dimension fondamentale de la linguistique.* (= *Études de linguistique française* 6). Berlin/Boston: De Gruyter, S. 185–208.
- Schmidt, Thomas/Ferger, Anne (2025): Putting things on top of other things – The ZuMult platform for multimodal corpora and its ecosystem. Paper presented at the CLARIN Annual Conference 2025, Vienna, 30 September – 2 October 2025.
- Schmidt, Thomas/Blevins, Margaret/Boas, Hans C./Gilbert, Glenn G. (2026): The Texas German Dialect Project Corpus as a diachronic resource for investigating language contact. In: *Proceedings of Workshop on Dialects in NLP – A Resource Perspective (DialRes) at LREC 2026*. Paris: European Language Resources Association (ELRA), S. 221–229.
- Schmidt, Thomas. (in Vorb.): Die ZuMult-Plattform für audiovisuelle Sprachkorpora. In: *Zeitschrift für Germanistische Linguistik*.
- Seyfeddinipur, Mandana/Krifka, Manfred/Ameka, Felix/Kung, Susan/Bolton, Lissant/Monroig, Miyuki/Blumtritt, Jonathan/Neba, Ayu'nwi/Carpenter, Brian/Nordhoff, Sebastian/Cruz, Hilaria/Pakendorf, Brigitte/Drude, Sebastian/Epps, Patience/Rau, Felix/Ferreira, Vera/Rice, Keren/Galucio, Ana/Rießler, Michael/Prince, Kilu von (2019): Public access to research data in language documentation: Challenges and possible strategies. In: *Language Documentation and Conservation* 13, S. 545–563.
- Sloetjes, Han/Wittenburg, Peter (2008): Annotation by category – ELAN and ISO DCR. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Marrakesch: European Language Resources Association (ELRA), S. 816–820.
- Thompson, Jana (2005): *Code-switching in Texas German*. Masterarbeit. Austin, TX: The University of Texas at Austin.
- Trudgill, Peter (2004): *New-dialect formation: The inevitability of colonial Englishes*. Oxford/New York: Oxford University Press.
- Warmuth, Matthias (2023): *Phonological convergence and variation in a dying language variety: The case of Texas German*. Dissertation. Austin, TX: The University of Texas at Austin.
- Weilbacher, Hunter (2011): Texas German discourse pragmatics. A preliminary study of the English-origin discourse markers *of course*, *see*, and *now*. In: Putnam, Michael T. (Hg.): *Studies on German-language Islands.* (= *Studies in language companion series* 123). Amsterdam/Philadelphia: Benjamins, S. 455–474.
- Weiss, Jana/Blevins, Margo/Boas, Hans C./Sauer, Verena (2025): Towards a roadmap for exploring past and present Texas German voices: A case study of German-Indigenous relations. In: *Yearbook of German American Studies* 58, S. 173–214.
- Westpfahl, Swantje/Schmidt, Thomas (2013): POS für(s) FOLK – Part of Speech Tagging des Forschungs- und Lehrkorpus Gesprochenes Deutsch. In: Zinsmeister, Heike/Heid, Ulrich/Beck, Kathrin (Hg.): *Das Stuttgart-Tübingen Wortarten-Tagset – Stand und Perspektiven*. *Journal for Language Technology and Computational Linguistics* 28.1, S. 139–153.
- Westpfahl, Swantje/Schmidt, Thomas (2016): FOLK-Gold – A GOLD standard for Part-of-speech-tagging of Spoken German. In: Calzolari, N./Choukri, K./Declerck, T./Goggi, S./Grobelnik, M./Maegaard, B./Mariani, J./Mazo, H./Moreno, A./Odijk, J./Piperidis, S. Hg.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Paris: European Language Resources Association (ELRA), S. 1493–1499.

25 Jahre Texas German Dialect Project

- Whitworth, Leslie (2005): Remnants of the German language in Mason county, Texas. Masterarbeit. Austin, TX: The University of Texas at Austin.
- Wilson, Joseph (1986): Texas German and other American immigrant languages: Problems and prospects. In: Gish, Theodore/Spuler, Richard (Hg.): *Eagle in the New World: German immigration to Texas and America*. College Station, TX: Texas A & M University Press, S. 221–240.
- Wolfram, Walt (2011): Fieldwork methods in language variation. In Wodak, Ruth/Johnstone, Barbara/Kerswill, Paul (Hg.): *The SAGE handbook of sociolinguistics*. Los Angeles: Sage, S. 296–311.
- Wolfram, Walt/Reaser, Jeffrey/Vaughn, Charlotte (2008): Operationalizing linguistic gratuity: From principle to practice. In: *Language and Linguistics Compass* 2, 6, S. 1109–1134.
- Zimmer, Christian/Wiese, Heike/Simon, Horst J./Zappen-Thomson, Marianne/Bracke, Yannic/Stuhl, Britta/Schmidt, Thomas (2020): Das Korpus Deutsch in Namibia (DNam): Eine Ressource für die Kontakt-, Variations- und Soziolinguistik. In: *Deutsche Sprache* 48, S. 210–232.
- Zwirner, Eberhard/Bethge, Wolfgang (1958): *Erläuterungen zu den Texten*. (= Lautbibliothek der deutschen Mundarten 1). Göttingen: Vandenhoeck & Ruprecht.

Dr. Hans C. Boas
Raymond Dickson, Alton C. Allen, and Dillon Anderson Centennial Professor
Department of Germanic Studies
2505 University Ave, C3300
Burdine Hall 336, 3rd Floor
The University of Texas at Austin
Austin, TX 78712-1802
U.S.A.
E-Mail: hcb@mail.utexas.edu

Dr. Margo Blevins
Department of Germanic Studies
2505 University Ave, C3300
Burdine Hall 336, 3rd Floor
The University of Texas at Austin
Austin, TX 78712-1802
U.S.A.

Dr. Thomas Schmidt
LinguisticBits / MusicalBits
Dromersheimerstrasse 24b
55411 Bingen

Dr. Ellen J. Schoedler
Department of Germanic Studies
2505 University Ave, C3300
Burdine Hall 336, 3rd Floor
The University of Texas at Austin
Austin, TX 78712-1802
U.S.A.

Hans C. Boas / Margo Blevins / Thomas Schmidt / Ellen J. Schoedler / Joren Somers

Dr. Joren Somers
Department of Germanic Studies
2505 University Ave, C3300
Burdine Hall 336, 3rd Floor
The University of Texas at Austin
Austin, TX 78712-1802
U.S.A.