RegioLingua

Zeitschrift für regionale Sprache und Literatur

Herausgegeben von
Doreen Brandt, Marina Frank, Robert Langhanke, Nicole Palliwoda und Jeffrey Pheiff

Jahrgang 3 (2026)

Heft 1

Magnus Breder Birkenes, Robert Engsterhold, Hanna Fischer und Jeffrey Pheiff (Hg.)
Digitale Dialektologie. Plattformen und Apps zu den deutschen Regionalsprachen

Hans C. Boas, Margaret Blevins & Thomas Schmidt

Texas-Deutsch Digital

Zitationsvorschlag:

Boas, Hans C./Blevins, Margaret/Schmidt, Thomas (2026): Texas-Deutsch Digital. In: Birkenes, Magnus Breder/Engsterhold, Robert/Fischer, Hanna/Pheiff, Jeffrey (Hg.). *Digitale Dialektologie. Plattformen und Apps zu den deutschen Regionalsprachen*. = *RegioLingua. Zeitschrift für regionale Sprache und Literatur* 3 (1), 377–391. DOI: 10.17192/regiolingua.2026.3.1.9123.

https://journals.uni-marburg.de/regiolingua/
Kontakt: regiolingua@uni-marburg.de
ISSN: 2944-2133

Die Zeitschrift *RegioLingua* erscheint online über den Publikationsserver der Universitätsbibliothek Marburg unter der Creative Commons Lizenz CC-BY 4.0.

HANS C. BOAS, MARGARET BLEVINS & THOMAS SCHMIDT

Texas-Deutsch Digital

Texas German Digital

Zusammenfassung: Texas-Deutsch ist ein seit den 1840er Jahren gesprochener gemischter Auswandererdialekt (vgl. Gilbert 1972). Seit 2001 erhebt und archiviert das *Texas German Dialect Project* (TGDP) Sprachdaten des akut vom Aussterben bedrohten Texas-Deutschen. Dieser Artikel stellt die digitale Infrastruktur des TGDP vor und gibt einen kurzen Überblick über die Daten, die es erfasst. Seit 2024 sind die Daten mit mehreren Annotationsebenen über die Korpusplattform ZuMult zugänglich. Anschließend wird gezeigt, wie das System komplexe Korpusanalysen ermöglicht.

Schlagwörter: Texas-Deutsch, Sprachkontakt, Sprachinseln, extraterritoriale Varietäten des Deutschen, Korpusaufbau, ZuMult.

Abstract: Texas German is a mixed immigrant dialect that has been spoken since the 1840s (see Gilbert 1972). Since 2001, the *Texas German Dialect Project* (TGDP) has been collecting and archiving linguistic data on Texas German, a language that is critically endangered. This article presents the TGDP's digital infrastructure and provides a brief overview of the data it collects. Since 2024, the data has been accessible via the ZuMult corpus platform with multiple levels of annotation. It is then demonstrated how the system enables complex corpus analyses.

Keywords: Texas German, language contact, speech islands, extraterritorial varieties of German, corpus building, ZuMult.

1. Gegenstand

Texas-Deutsch ist ein seit den 1840er Jahren gesprochener gemischter Auswandererdialekt (siehe Gilbert 1972), der aus dem Kontakt von mindestens fünf unterschiedlichen Dialekten des Deutschen entstanden ist und akut vom Aussterben bedroht ist, da er seit den 1950er Jahren nicht mehr an jüngere Generationen weitergegeben wurde (siehe Moore 1979, Salmons 1983, Wilson 1986, Nicolini 2004, Boas 2009).[1] Im Jahr 2001 gründete Hans C. Boas im *Department of Germanic Studies* an der *University of Texas at Austin* das *Texas German Dialect Project* (TGDP), um so viele SprecherInnen des Texas-Deutschen wie möglich zu interviewen und die Tonaufnahmen der Interviews sowie die daraus resultierenden Transkripte und Übersetzungen ins Englische online für Lehr- und

[1] Für einen Überblick über die Ursprungsdialekte und die unterschiedlichen strukturellen Merkmale des Texas-Deutschen siehe Boas (2018) und Boas/Pierce (2025). Das Texas-Deutsche zeichnet sich durch einen hohen Grad an (inter- und intraindividueller) Variation aus, siehe Boas (2016).

RegioLingua. Zeitschrift für regionale Sprache und Literatur, 3 (2026), H. 1, 377–391.
DOI: 10.17192/regiolingua.2026.3.1.9123

Forschungszwecke zur Verfügung zu stellen (siehe Boas 2002, Boas 2006).[2] In den letzten 25 Jahren haben Mitglieder des TGDP drei unterschiedliche Arten von Daten aufgenommen (siehe Methode).

Die von Mitgliedern des TGDP aufgenommenen und verarbeiteten Interviewdaten werden im Texas German Dialect Archive Online (TGDA-Online, tgdp.org/dialect-archive) archiviert, um so für eine dauerhafte Sicherung der Aufnahmen mit den letzten noch verbliebenen SprecherInnen des Texas-Deutschen für künftige Generationen zu sorgen. Ursprünglich wurde das TGDA-Online primär für Unterrichtszwecke konzipiert, d. h. es sollte ab dem Jahr 2002 für Studierende der Germanistischen Linguistik an der *University of Texas at Austin* ein online Zugang zu gesprochenen Daten des Texas-Deutschen ermöglicht werden, um diese im Unterricht bei der Analyse von Daten und der Überprüfung von abstrakten linguistischen Konzepten bzw. außerhalb der Lehrveranstaltungen im Rahmen von Hausaufgaben nutzen zu können. Erst im Rahmen der Einbindung von texas-deutschen Daten aus dem TGDA-Online in Lehrveranstaltungen in den Jahren 2002–2003 wurde deutlich, dass die im Archiv deponierten Daten auch die Grundlage für sprachwissenschaftliche Forschungen bilden können. So werden die seit 2002 aufgenommenen Daten mittlerweile nicht nur in Lehrveranstaltungen eingesetzt, sondern auch für Forschungszwecke.[3]

2. Umsetzung

2.1. Methode

Die vom TGDP durchgeführten Interviews bestehen aus drei Teilen: Übersetzungen, einem soziolinguistischen Interview und einem biografischen Fragebogen.

Übersetzungen: Die Gewährspersonen werden gebeten, Wörter, Mehrworteinheiten und Sätze vom Englischen ins Texas-Deutsche zu übersetzen. Dieser kontrollierte Datensatz beruht auf den Übersetzungslisten von Gilbert (1972) (148 Wörter, Mehrworteinheiten und Sätze), Guion (1996) (20 Sätze) und Eikel (1954) (191 Sätze).[4] Diese vom TGDP erhobenen Daten erlauben einen Vergleich anzustellen, wie unterschiedliche SprecherInnen des Texas-Deutschen dieselben Wörter, Mehrworteinheiten oder Sätze unterschiedlich realisieren (siehe Boas 2009 und Boas/Pierce 2011).[5]

Soziolinguistische Interviews: Diese Interviews, in denen in offenen Gesprächen unterschiedliche Themen besprochen werden, werden so weit wie möglich auf Deutsch geführt.[6] Obwohl den Informanten nicht identische Fragen gestellt werden, gibt es viele Themen, die in diesen Interviews häufig vorkommen, wie z. B. Kindheit, Familie,

[2] Das TGDP hat über die Jahre hinweg erhebliche Unterstützung beim Auf- und Ausbau der dem Archiv zu Grunde liegenden Datenbank und Website durch die Programmierer der *Liberal Arts Instructional Technology Services* (LAITS) an der *University of Texas* erhalten.
[3] Siehe Boas/Fingerhuth (2018) für einen Vergleich der Arten von Daten des TGDA-Online mit den Arten von Daten in anderen Korpora für deutsche Sprachinseldaten.
[4] Siehe Pierce et al. (2019) und Pierce et al. (2021) zur Evaluation der wissenschaftlichen Forschungen von Fred Eikel zum Texas-Deutschen.
[5] Aufgrund von Zeitbeschränkungen haben weniger als 100 SprecherInnen alle drei Übersetzungslisten komplett abgefragt.
[6] Die InterviewerInnen selbst sind L1-, L2- und L3-SprecherInnen des Standarddeutschen.

Vorfahren, Ausbildung, Schulerfahrungen, Traditionen, Politik, Wetter, Sprachgebrauch, Essen und Trinken, Lokalgeschichte und Gebräuche. Ziel der Fragen ist es, die Gewährspersonen in ein Gespräch zu verwickeln, in dem sie so natürlich wie möglich spontan Texas-Deutsch sprechen und nicht nur kurze Antworten auf einzelne Fragen geben (siehe Boas 2003).

Biografischer Fragebogen: Dieser Fragebogen erfragt unter anderem genaue Informationen über Alter, Herkunft, Religion, Bildungsgrad und Bereiche des Sprachgebrauchs sowie der Sprach- und SprecherInnen-Einstellungen (siehe Boas 2005, Boas/Fingerhuth 2017, Boas 2021a). Während des Interviews wird der Fragebogen vom Interviewer schriftlich ausgefüllt. Das biografische Interview wird auf Englisch geführt, um u. a. festzustellen, ob etwaige Sprechunflüssigkeiten beim Sprechen des Texas-Deutschen wie Wortfindungsprobleme, Stottern, häufige Reparatur von Satzstrukturen usw. auch beim Sprechen des Englischen auftreten und herauszuarbeiten, wie die Gewährspersonen Englisch sprechen, d. h. ob sie mit einem von ihrer Erstsprache Texas-Deutsch verursachten Akzent sprechen oder nicht (siehe Boas et al. 2010 und Boas 2021b).

Zusätzlich zu diesen drei Kern-Datentypen verfügt das TGDP auch über weitere Materialien, die dem Projekt gegeben wurden, wie z. B. Audioaufnahmen aus den 1960er- und 1990er-Jahren. Das *Texas German Dialect Archive* (TGDA**)** ist die vollständige Sammlung aller texas-deutschen Materialien des TGDPs. Es umfasst schriftliche Daten (Briefe, Tagebücher, Zeitungen usw.), Audio- und Videomaterialien (sowohl die vom TGDP erstellten als auch die dem TGDP gespendeten Materialien) und weitere Materialien. Die verschiedenen Ressourcen des TGDA und die Art und Weise sie miteinander in Beziehung stehen, sind in Abbildung 1 dargestellt.[7]

Man kann auf den öffentlich zugänglichen Teil des TGDA online über das Texas German Dialect Archive Online (TGDA-Online)[8] oder ZuMult[9] zugreifen. Im *TGDA-Online* kann man sich sowohl Audioausschnitte von Übersetzungslisten, die auf den Listen von Gilbert (1972) und Eikel (1954) basieren, als auch das TGDP-Korpus anhören. Das *TGDP-Korpus* enthält alle derzeit öffentlich zugänglichen soziolinguistischen Interviewausschnitte zusammen mit ihren Transkriptionen und englischen Übersetzungen.

Über *ZuMult* (siehe 2.2 unten) kann man auf eine Version des TGDP-Korpus mit ausführlichen Annotationen zugreifen. Es ist möglich, Interviews anhand der SprecherInnen-ID-Nummer, der Interviewer-ID-Nummer und dem Interview-Ort zu suchen (*Browse by Interview*). Es gibt auch eine CQP-Abfrageoberfläche,[10] die komplexe Abfragen der vielen Annotationsebenen in den Transkriptionsdateien ermöglicht (*Query*, s. Ergebnisse unten). Mit der Funktion *Browse by section titles* können Benutzer nach bestimmten Schlüsselwörtern in den beschreibenden Titeln der Interviewsegmente suchen (z. B. „school", „Christmas"), um Gespräche zu bestimmten Themen zu finden.

[7] Das TGDP umfasst das gesamte Forschungsprojekt mit all seinen Bestandteilen, einschließlich Daten aus anderen Gebieten (z. B. Pennsylvania-Deutsch, Wisconsin-Deutsch).
[8] https://tgdp.org/dialect-archive-viewer/#/login (02.11.2026).
[9] https://tgdp-zumult.la.utexas.edu/index.jsp (02.11.2026).
[10] CQP ist ein Akronym für Corpus Query Processor.

Neben dem Zugriff auf das TGDP-Korpus ermöglicht ZuMult auch den Zugriff auf den soziolinguistischen Teil des GTXG-Korpus. Das *GTXG-Korpus* ist eine Sammlung von Tonaufnahmen, die Glenn Gilbert in den 1960er und 1970er Jahren angefertigt hat. Ähnlich wie die vom TGDP gesammelten Daten enthält das GTXG-Korpus soziolinguistische Interviews und Übersetzungsaufgaben. Derzeit ist online nur der Teil dieses historischen Korpus verfügbar, der soziolinguistische Interviews enthält.[11]

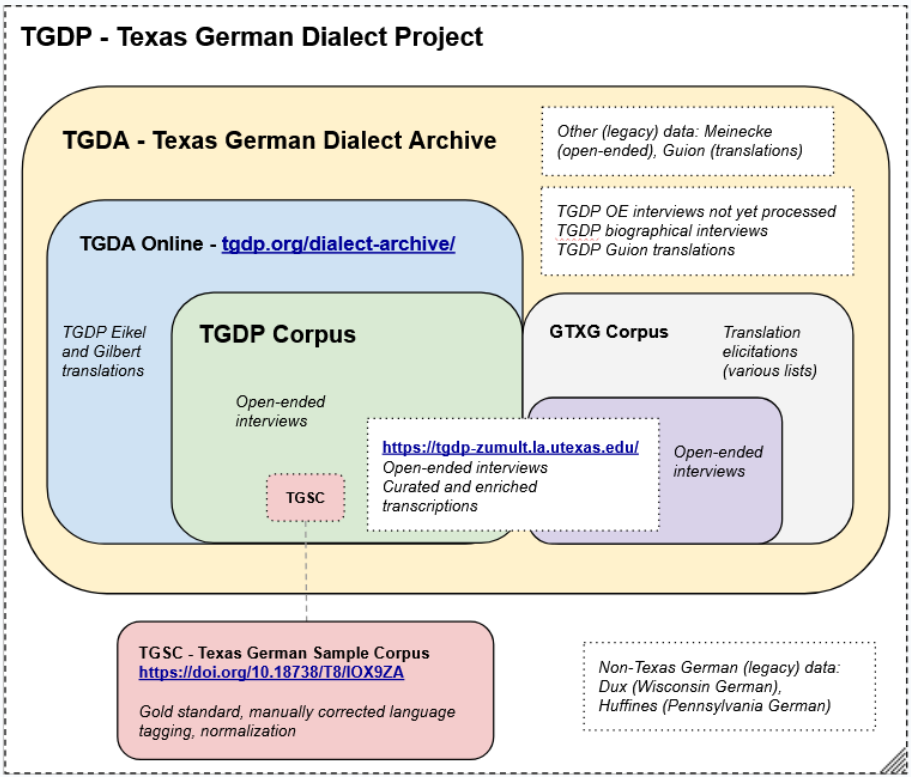


Abbildung 1: Vom Texas German Dialect Project erstellte Ressourcen

In ZuMult wird die manuelle Transkription in literarischer Umschrift automatisch tokenisiert, und es werden automatisch zusätzliche Annotationsebenen hinzugefügt (Abb. 2). Somit stehen zur Verfügung:

- *Original* = die ursprüngliche, manuell erstellte Transkription (Span-Annotation gemäß den projektinternen Transkriptionsrichtlinien)
- *Translation* = die ursprüngliche, manuell erstellte wortwörtliche englische Übersetzung (Span-Annotation)
- *ID* = eine eindeutige Token-ID für jedes Token innerhalb einer Transkription
- *Token* = eine tokenisierte Version der ursprünglichen Transkription
- *Language* = Sprach-Tags für jedes Token (deu = Deutsch, eng = Englisch, amb = ambig, xxx = nicht-lexikalisches Material, z. B. Häsitationsmarker) (automatisch generiert)
- *Normalization* = orthografisch normalisierte Tokens (automatisch generiert)
- *Lemma* = Lemma (automatisch generiert)

[11] Derzeit wird daran gearbeitet, historische Übersetzungsdaten auch über ZuMult zugänglich zu machen, siehe Schmidt et al. (2026a).

- *POS* = Wortart-Tags (STTS 2.0-Tag-Set, Westpfahl/Schmidt, 2016) (automatisch generiert)
- *UD-POS* = Wortart-Tags (Universal Dependencies-Tag-Set, de Marneff et al. 2021) (automatisch generiert)
- *Phon* = kanonische phonologische Umschrift in IPA, mittels G2P (Reichel/Kisler 2014) automatisch aus der literarisch transkribierten Form berechnet
- *Speech-rate* = Artikulationsrate, berechnet als Silben pro Sekunde (nicht in Abbildung 2 dargestellt)
- *Transfer* = potentiell von Sprachtransfer betroffene Tokens, während der ursprünglichen Transkription von Transkribenden mit eckigen Klammern annotiert (in Abbildung 12 nicht dargestellt)[12]

Außerdem stellt ZuMult auch die Metadaten zu Interviews (z. B. Ort, Jahr) und zu den beteiligten SprecherInnen (z. B. Geschlecht, Geburtsjahr, religiöse Affiliation) für Abfragen zur Verfügung.

Annotations for TRS_1-33-1-14-a / ab_a9 ×

Original	SO der hat in ein [store] gearbeit un ich hab in anern [store] gearbeit. Da ham wir ... denn i											
Translation	*So he had in a store worked and I have in another store worked. There had we ... then is he ... ge- sch- finally is he gone-awe*											
ID	a9_w1	a9_w3	a9_w5	a9_w7	a9_w9	a9_w11	a9_w13	a9_w15	a9_w17	a9_w19	a9_w21	a9_w23
Token	SO	der	hat	in	ein	store	gearbeit	un	ich	hab	in	anern
Language	deu	deu	deu	deu	deu	eng	deu	deu	deu	deu	deu	deu
Normalisation	so	der	hat	in	ein	store	gearbeitet	und	ich	habe	in	anderen
Lemma	*so*	*die*	*haben*	*in*	*ein*	*store*	*arbeiten*	*und*	*ich*	*haben*	*in*	*ander*
POS	NGIRR	PDS	VAFIN	APPR	ART	FM	VVPP	KON	PPER	VAFIN	APPR	PIDAT
UD-POS	INTJ	PRON	AUX	ADP	DET	NOUN	VERB	CCONJ	PRON	AUX	ADP	DET
Phon	[z'O]	[d'e:6]	[h'at]	[?'In]	[?'aIn]	[st'O:R]	[g@.?'ar.baIt]	[?'Un]	[?'IC]	[h'a:p]	[?'In]	[?'an.6n]

Abbildung 2: Auszug aus dem annotierten Transkript des TGDP-Interviews 1-33-1-14-a

Die Sprach-Tags und Normalisierungen sind die Grundlage, auf der die Ebenen Lemma, POS, UD-POS und Phon basieren. Die Sprach-Tags und das Normalisierungsschema basieren auf dem in Blevins (2022) entwickelten System zur orthografischen Normalisierung und Sprachkennzeichnung. Die Sprach-Tags weisen eine Fehlerquote von 6 % auf, die Normalisierungsebene eine Fehlerquote von 7 % und die POS-Tags eine Fehlerquote von 14,6 %.[13] Die Verwendung der Normalisierungs- und Sprachkennzeichnungsebenen als Grundlage für die Lemmatisierung und POS-Ebenen bedeutet, dass Fehler aus der

[12] In den Originaltranskriptionen werden die Transkribierenden gebeten, Elemente, die möglicherweise auf Sprachkontakt zurückzuführen sind, in eckigen Klammern zu kennzeichnen, z. B. englische Wörter, gemischte Wörter, nicht standardmäßige Wortstellung, nicht standardmäßige Genus/Kasus usw.

[13] Die POS-Fehlerquote ist zwar relativ hoch, für Korpusabfragen in einem qualitativen Ansatz, d. h. bei denen es darum geht, gute Beispiele zu finden und nicht größere, vollständige Trefferlisten von echten Positiven, ist sie dennoch akzeptabel.

automatischen Normalisierung möglicherweise auf die beiden anderen Schichten übertragen werden können. Die Fehlerquoten für die anderen Ebenen wurden nicht berechnet.

2.2. Technologie

ZuMult steht für *Zugänge zu multimodalen Korpora gesprochener Sprache*. Es handelt sich um eine Architektur und Plattform, mit der audiovisuelle Sprachkorpora webbasiert zur Verfügung gestellt werden können. Die Architektur ist so konzipiert, dass sie flexibel für verschiedene Formen der Datenhaltung im Backend konfiguriert und im Frontend für verschiedene Nutzungsszenarien angepasst werden kann (Fandrych et al. 2023).

Den Kern der ZuMult-Architektur bildet ein Datenmodell, mit dem sich Design, Primärdaten und Sekundärdaten verschiedenartiger Korpora gesprochener Sprache auf einer gemeinsamen Basis abbilden lassen. Die Implementierung der Architektur bedient sich, wo immer möglich, etablierter Standards (wie etwa *ISO 24624:2016 Language resource management — Transcription of spoken language*) und bewährter Technologien (wie etwa Lucene zur Indizierung der Annotationen für die CQP-Suche). Dies trägt zur Interoperabilität und Anschlussfähigkeit der Plattform innerhalb digitaler Infrastrukturen bei und reduziert den langfristigen Aufwand für Entwicklung und Wartung des Codes (vgl. Schmidt et al. 2026b).

Für die Daten des TGDA wurde zunächst eine Verarbeitungs-Pipeline eingerichtet, die ausgehend von den ursprünglich in ELAN[14] erstellten Transkripten die oben beschriebenen zusätzlichen Informationen (Tokenisierung und verschiedene Annotationsebenen) automatisch berechnet und als ISO-konforme TEI-Dateien[15] speichert. Metadaten wurden aus der projektinternen Korpus-Management-Plattform speechislands.org in das XML-Format des EXMARaLDA Corpus-Managers (COMA, Wörner 2012) überführt. In dieser Form können sie von der ZuMult-Architektur verarbeitet werden.

Die Plattform selbst wurde in Details der Benutzeroberfläche auf die Texas German-Daten hin optimiert. Dies umfasst beispielsweise die standardmäßige Anzeige der englischen Übersetzungen als Interlinear-Annotationen im Transkript oder die farbliche Hervorhebung von Passagen, die als potentielle Transfer-Phänomene markiert wurden (siehe Abb. 3).

Für die Besonderheit der Unterteilung der Interviews in thematische Abschnitte wurde ein eigenes Such-Interface hinzugefügt, das vor allem solchen Nutzungsszenarien entgegenkommt, die mehr den Inhalt als die linguistische Form der Daten im Blick haben (Abb. 4).

[14] „ELAN (EUDICO Linguistic Annotator) ist ein Annotationswerkzeug, mit dem Sie Annotationen für Video- und Audiodaten erstellen, bearbeiten, visualisieren und durchsuchen können“ (Tacchetti 2018).
[15] Die Text Encoding Initiative (TEI) ist ein internationales Konsortium, das Standards und Richtlinien für die Kodierung maschinenlesbarer Texte (vor allem in den Geistes- und Sozialwissenschaften) entwickelt.

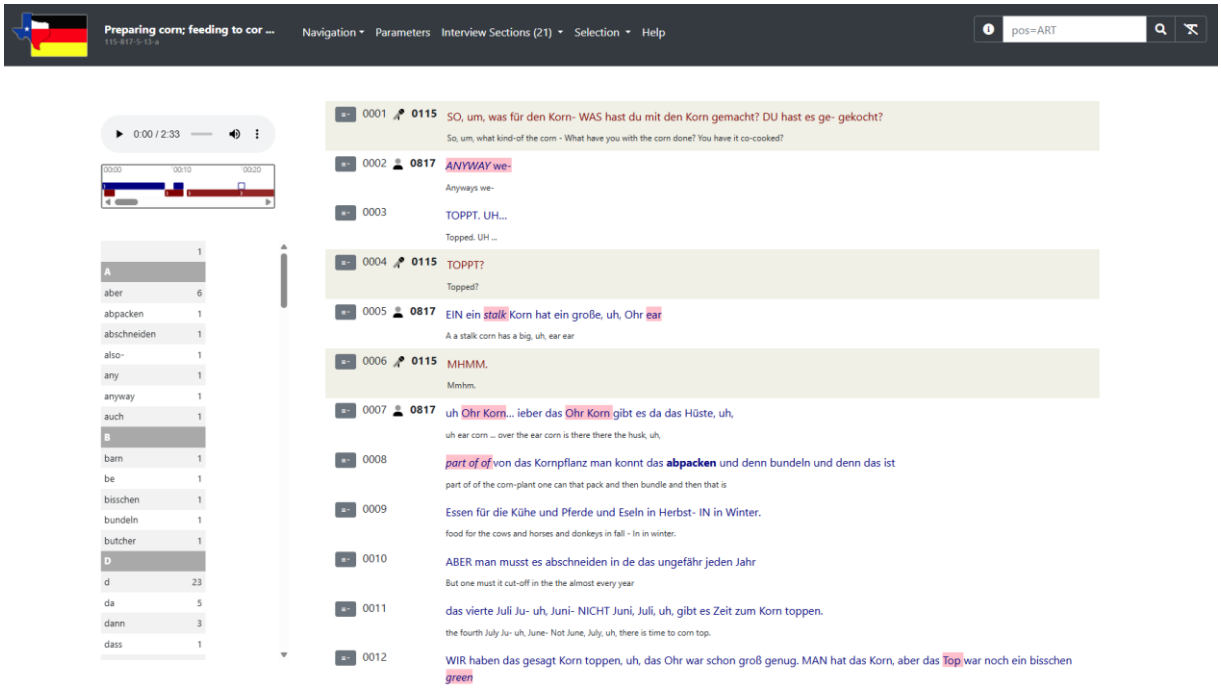


Abbildung 3: ZuMult Transkript

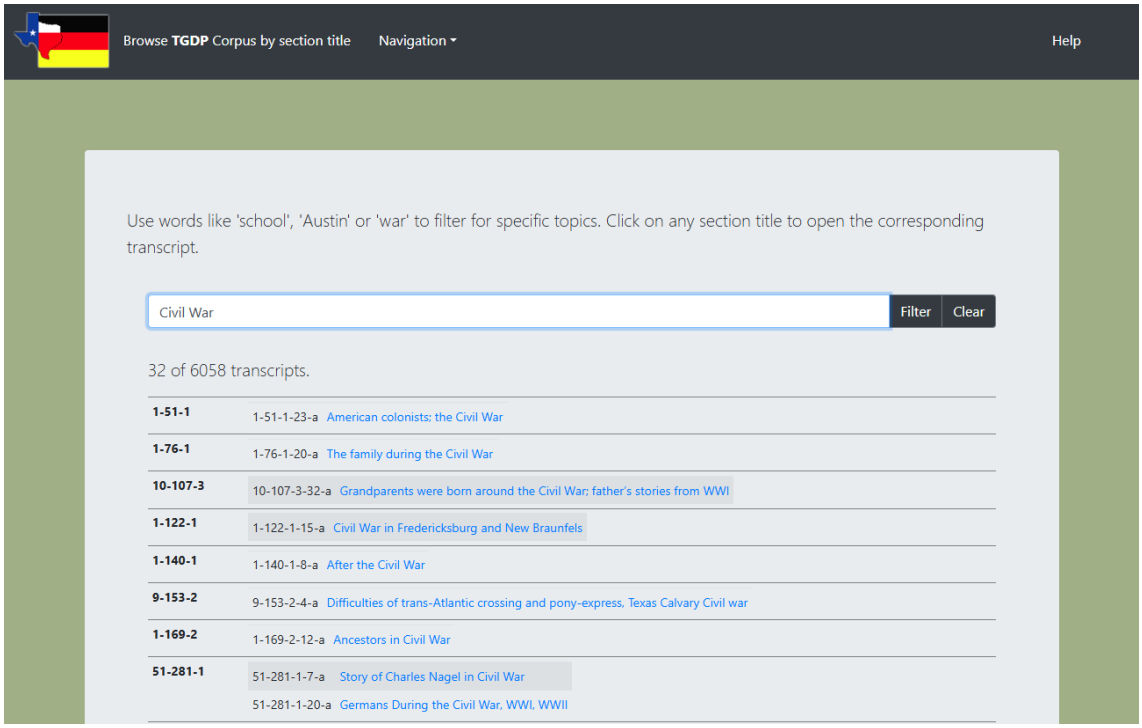


Abbildung 4: Stichwortsuche zu bestimmtem Thema (“Civil War”) mit ZuMult.

Die erste Version der Plattform wurde im Dezember 2024 veröffentlicht. Mit der zweiten Version im Januar 2026 kamen einige Verbesserungen der Funktionalität, zusätzliche Interviews aus dem TGDP-Korpus sowie die soziolinguistischen Interviews aus den Gilbert-Beständen (GTXG) hinzu.

3. Ergebnisse

ForscherInnen können alle Annotationsebenen sowie die grundlegenden Metadaten, die mit jedem Interview verbunden sind (z. B. Ort des Interviews, Jahr des Interviews) und/oder die Metadaten der SprecherInnen (z. B. Geschlecht, Geburtsjahr, Bildungsniveau, Religionszugehörigkeit) abfragen. Weitere Informationen zu Abfragen finden sich im Hilfedokument, das in ZuMult verfügbar ist (https://tgdp-zumult.la.utexas.edu/documentation.jsp).

Wenn man beispielsweise untersuchen möchte, wie SprecherInnen, aus New Braunfels englischen Substantiven ein Geschlecht zuweisen, könnte man folgende Abfrage stellen:

```
[pos="ART"][udpos="NOUN" & lang="eng"]
within <Interview_Location="New Braunfels"/>
```

Dadurch werden alle Vorkommen in Interviews aus New Braunfels gefunden, in denen auf ein Token, das als Artikel gekennzeichnet ist, direkt ein Token folgt, das mit einem englischen Sprach-Tag und einem Substantiv gemäß dem Universal Dependencies Tagset gekennzeichnet ist.[16]

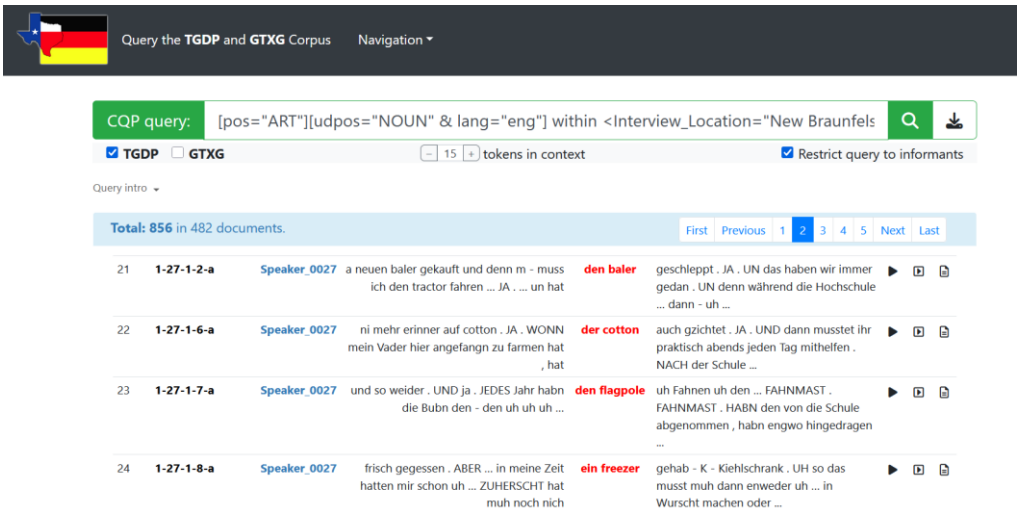


Abbildung 5: Beispiel Suchanfrage nach der Kombination deut. Artikel + engl. Nomen in New Braunfels

Zusätzlich zur Analyse über die KWIC-Konkordanz kann man auch Suchergebnisse exportieren sowie auf einzelne Treffer klicken, um zum entsprechenden Transkript zu gelangen und um sich die Original-Audiodatei anzuhören. Dies kann besonders hilfreich

[16] Wenn man versuchen würde, englische Substantive mit der POS-Ebene zu suchen, würde man null Treffer erhalten, da die POS-Ebene auf dem STTS 2.0-Tagset basiert und somit alle fremdsprachigen Materialien als FM markiert sind. Es kann hilfreich sein, mehrere Suchanfragen zu versuchen, um eine Frage auf unterschiedliche Weise anzugehen. In diesem Beispiel wäre eine weitere nützliche Suchanfrage: `[word="d.*"][udpos="NOUN" & lang="amb"] within <Interview_Location="New Braunfels"/>`. Diese Suche sucht nach allen Tokens, die mit einem „d“ beginnen und direkt darauf folgt ein Token, das als sprachlich mehrdeutiges Substantiv markiert ist.

sein, wenn man beispielsweise zwischen englischer und deutscher Aussprache (s. u.) unterscheiden möchte.

Ein weiteres Beispiel dafür, wie gut sich die ZuMult Plattform eignet, um systematisch zu untersuchen, wie Entlehnungen aus dem Englischen im Texas-Deutschen integriert worden sind bzw. welche Grundlage sie für erweiterte sprachliche Muster bieten, wie sie für Sprachkontaktsituationen typisch sind (siehe z. B. Clyne 2003, Boas/Höder 2025, Riehl 2026), ist die ins Texas-Deutsche entlehnte englische koordinierende Konjunktion *or*. Diese Konjunktion kommt u. a. in demselben Kontext wie die deutsche Konjunktion *oder* vor (z. B. *Ich habe nie for die Jungs or die Manns Zeuge genäht* (41-215-1-29-a)). Eine genauere Untersuchung der Ergebnisse der Korpusabfrage für *or* (insgesamt 1.212 Belege, siehe Abb. 6) zeigt darüber hinaus auch weitere interessante Verwendungen der aus dem Englischen entlehnten Konjunktion, wie die folgenden Beispiele zeigen.

Erstens wird *or* in Verbindung mit *so* als Chunk verwendet, um eine Annäherung auszudrücken, z. B. *Didn't know that's what I did but das das hab ich getan für viele Jahren uh zwanzig or so* (112-462-1-16-a). In dieser Diskurskonstruktion (siehe Östman/Trousdale 2013) kann *or* wahlweise mit englischer (/oɹ/) oder texas-deutscher (/oə/) Aussprache ausgesprochen, gefolgt von einer texas-deutschen bzw. englischen Aussprache von *so* (/zo:/ bzw. /so:/). Die genauen Unterschiede zwischen den Aussprachevarianten lassen sich aus dem Transkript per se nicht direkt erschließen, dazu kann über einen Klick auf den kleinen Pfeil auf der rechten Seite des Transkriptausschnitts die zugehörige Tonaufnahme abgespielt werden.

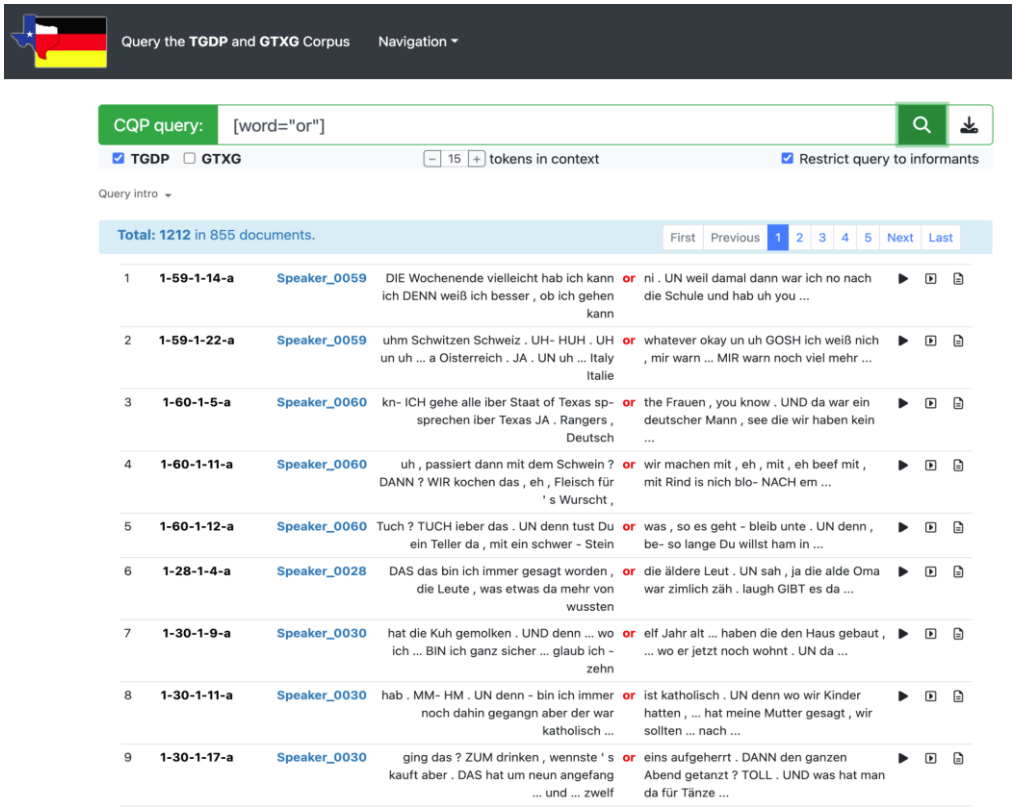


Abbildung 6: Beispiel Suchanfrage nach Kontexten, in denen die entlehnte englische koordinierende Konjunktion *or* ('oder') vorkommt

Zweitens wird *or* im Texas-Deutschen als Häsitationsmarker verwendet, wie in den Beispielen *... and for Spaß sind mir nach Landa Park, am bicycle or ... gefahren* (1-82-1-4-a) und *Das haben, or, haben ein Masse getan und denn musst mir die Eier waschen* (115-727-1-10-a). Hier benutzen die SprecherInnen die entlehnte englische Konjunktion *or*, um zu signalisieren, dass sie etwas Zeit brauchen, um nachzudenken bzw. sich an ein Wort zu erinnern (häufig erscheint diese Verwendung von *or* auch in Verbindung mit anderen Häsitationsmarkern wie *uh* (/ʌ/), z. B. *die hat uh Dressing gemacht und uh uh i weiss ni was fr ein uh- ich tu corn or- or uh- uh… What else do i make? I cant think right now* (115-586-1-3-a)).

Die dritte Funktion von *or* im Texas-Deutschen ist entweder das Signalisieren der Korrektur von etwas vorher Gesagtem, was nicht korrekt war wie in (1), oder einer Spezifizierung von etwas vorher Gesagtem wie in (2).

1. Korrektur
   a. *Und da haben wir wir wohnen an den camp an den camp… **or** in den camp* (112-533-1-10-a)
   b. *Un wenn duh etwas- uh pork **or** Schweinefleisch rindust, denn dut das das nich* (1-42-1-10-a)
2. Spezifierung
   a. *mein nächste Schwester is ein Jahre jünger wie ich bin **or** achtzehn Monat* (115-610-1-4-a)
   b. *da war' n Leut in… in die… Gemeinschaft **or** Community, was nicht Deutsch gesprochen haben* (1-55-1-10-a)

In (1a) korrigiert der Sprecher seine Präposition und ersetzt *an* durch *in*. In (1b) ersetzt der Sprecher ein englisches Wort durch das entsprechende deutsche Wort, da er erkennt, dass er in diesem Teil des Interviews (Texas-)Deutsch sprechen soll. In (2b) scheint der Sprecher nach dem richtigen Wort zu suchen (wie die Pause vor *Gemeinschaft* andeutet); anschließend scheint er entweder unsicher zu sein, ob er das richtige Wort gewählt hat, oder ob der Interviewer seine beabsichtigte Bedeutung richtig versteht, weshalb er das englische Wort hinzufügt, um Klarheit zu schaffen.

4. Ausblick

Das seit fast 25 Jahren online verfügbare TGDA-Online hat sich im Laufe der Zeit erheblich verändert. Was als überschaubare Online-Ressource für die Lehre in germanistischer Linguistik an der University of Texas at Austin geplant war, hat sich seit 2002 zu dem weltweit umfangreichsten Onlinearchive einer deutschen Sprachkontaktvarietät entwickelt (siehe Boas 2021b, Boas et al. 2026). Wie mittlerweile gut 5.500 Registrierungen für das TGDA-Online (Stand: März 2026) bezeugen, besteht seitens Studierenden, Lehrenden und Forschenden ein durchaus großes Interesse am Texas-Deutschen. Die seit 2024 implementierte Korpusplattform ZuMult bietet darüber hinaus viele umfangreiche Recherchemöglichkeiten an, die auch komplexe Korpusanfragen bedienen kann (siehe Abschnitt 3 oben).

Durch regelmäßige Updates (im Schnitt einmal jährlich) werden die Korpusbestände des TGDA-Online kontinuierlich erweitert. In der ZuMult-Version vom Dezember 2025

sind ca. 196 Stunden bearbeitete, transkribierte und übersetzte soziolinguistischer Interviews veröffentlicht. Stand März 2026 verfügt das TGDP aber noch immer über weitere 100 Stunden zusätzlicher Tonaufnahmen soziolinguistischer Interviews der letzten 25 Jahre, die aus Mangel an finanzieller Unterstützung, Personal und Zeit noch nicht bearbeitet werden konnten. Aktuell rechnen wir mit einem regelmäßigen jährlichen Update des TGDP-Korpus, bei dem nicht nur neue Tonaufnahmen mit den dazugehörigen Transkripten und Übersetzungen veröffentlicht werden sollen, sondern auch die Funktionalität der ZuMult Plattform weiterhin ausgebaut werden soll. Darüber hinaus hoffen wir auch, in den kommenden zehn Jahren noch mehrere hundert zusätzliche SprecherInnen des Texas-Deutschen zu finden und aufzunehmen, bevor der Dialekt nach fast 200 Jahren ausgestorben sein wird (siehe Wilson 1986). Diese Tonaufnahmen, zusammen mit ihren Transkripten und Übersetzungen, sollen dann auch sukzessive veröffentlicht werden.

Die Korpusplattform ZuMult eröffnet auch neue Wege für die vergleichende Sprachinselforschung (siehe Rosenberg 2005, Boas 2021b, Boas 2025, Riehl 2026). So gibt es am Archiv für Gesprochenes Deutsch (Stift/Schmidt 2014) am Leibniz Institut für Deutsche Sprache (Mannheim) andere Korpora von deutschen Kontaktvarietäten wie dem Namibia-Deutschen (Zimmer et al. 2020) und dem Unserdeutschen (Maitz/Volker 2017), die ebenfalls mit ZuMult abgefragt werden können. Damit eröffnet sich u. a. die Möglichkeit, dieselbe Korpusanfrage für drei unterschiedliche Kontaktvarietäten des Deutschen zu stellen, um so dieselben Phänomene systematisch kontrastiv zu erforschen (siehe Boas 2025).

Referenzen

Blevins, Margaret (2022): *The Language-tagging & Orthographic Normalization of Spoken Mixed-language Data, with a Focus on Texas German.* The University of Texas at Austin/Doktorarbeit.

Boas, Hans C. (2002): The Texas German Dialect Archive as a Tool for Analyzing Sound Change. In: Austin, Peter/Dry, Helen A./Wittenburg, 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, Spain.

Boas, Hans C. (2003): Tracing Dialect Death: The Texas German Dialect Project. In: Larson, Julie/Paster, Mary (Hg.): *Proceedings of the 28th Meeting of the Berkeley Linguistics Society*, 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/Zersen, David (Hg.): *Transcontinental Encounters. Central Europe Meets the American Heartland*. Austin: Concordia University Press, 78–102.

Boas, Hans C. (2006): From the Field to the Web. Implementing Best-practice Recommendations in Documentary Linguistics. In: *Language Resources and Evaluation* 40 (2), 153–174.

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

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

Boas, Hans C. (2018): Texas. In: Plewnia, Albrecht/Riehl, Claudia Maria (Hg.): *Handbuch der deutschen Sprachminderheiten in Übersee*. Tübingen: Narr, 171–192.

Boas, Hans C. (2021a): 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*. Tübingen: Narr Francke Attempto, 66–88.

Boas, Hans C. (2021b): Zwei Jahrzehnte digitale Dokumentation und Erforschung eines aussterbenden deutschen Auswandererdialekts. Das Texas German Dialect Project (2001–2021). In: *Zeitschrift für deutschsprachige Kultur und Literatur* 30, 229–268.

Boas, Hans C. (2025): Towards a systematic methodology for comparing extraterritorial German contact varieties. In: Lindenfelser, Siegwalt/Prediger, Angélica/Szczepaniak, Renata (Hg.): *Deutsch als Minderheitensprache in der Welt*. Berlin/Boston: De Gruyter, 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, 95–121.

Boas, Hans C./Fingerhuth, Matthias (2018): Deutsche Sprachinselkorpora im 21. Jahrhundert. In: Kupietz, Marc/Schmidt, Thomas (Hg.): *Germanistische Sprachwissenschaft um 2020: Korpuslinguistik. Vol. 5*. Berlin/Boston: De Gruyter, 125–150.

Boas, Hans. C./Höder, Steffen (2025): Diasystematic Construction Grammar at work. The need for anon-modular, data-driven approach to multilingual grammar. In: Boas, Hans C. Boas/Höder, Steffen (Hg.). *Constructions in Contact 3. Constructional schemas and patterns in language contact.* Amsterdam/Philadelphia: John Benjamins, 22–78.

Boas, Hans C./Pierce, Marc (2011): Lexical Developments in Texas German. In: Putnam, Michael (Hg.): *Studies on German Language Islands*. Amsterdam/Philadelphia: John Benjamins, 129–150.

Boas, Hans C./Pierce, Marc (2025): Texas German. In: Louden, Mark/Keel, William (Hg.): *North American Varieties of German. Yearbook of German American Studies Supplemental Issue* 6, 159–190.

Boas, Hans C./Pierce, Marc/Roesch, Karen/Halder, Guido/Weilbacher, Hunter (2010): The Texas German Dialect Archive: A Multimedia Resource for Research, Teaching, and Outreach. In: *Journal of Germanic Linguistics* 22 (3), 277–296.

Boas, Hans C./Schmidt, Thomas/Blevins, Margo (2026). A new corpus platform for the Texas German Dialect Project. In: *Language Resources and Evaluation*, 60: 52. DOI: 10.1007/s10579-025-09893-6.

Clyne, Michael (2003): *Dynamics of Language Contact*. Cambridge: Cambridge University Press.

de Marneffe, Marie-Catherine/Manning, Christopher/Nivre, Joakim/Zeman, Daniel (2021): Universal Dependencies. In: *Computational Linguistics* 47 (2), 255–308.

Eikel, Fred (1954): *The New Braunfels German Dialect. Manuscript*. Johns Hopkins University/Doktorarbeit.

Fandrych, Christian/Schmidt, Thomas/Wallner, Franziska/Wörner, Kai. (Hg.) (2023): Zugänge zu multimodalen Korpora gesprochener Sprache. In: *Themenheft der Zeitschrift Korpora Deutsch als Fremdsprache* (3) 1. https://kordaf.tujournals.ulb.tu-darmstadt.de/issue/92/info/ [Stand: 02.11.2026].

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

Guion, Susan G. (1996): The Death of Texas German in Gillespie County. In: Ureland, P. Sture/Clarkson, Iain (Hg.): *Language Contact Across the North Atlantic*. Tübingen: Niemeyer, 443–463.

ISO (2016): ISO 24624:2016 Language resource management — Transcription of spoken language.

Maitz, Peter/Volker, Craig (2017): Documenting Unserdeutsch. Reversing Colonial Amnesia. In: *Journal of Pidgin and Creole Languages* 32 (2), 365–397.

Moore, Barbara Joane Reeves (1979): *A sociolinguistic longitudinal study, 1969–1979, of a Texas German community, including curricular recommendations*. The University of Texas at Austin/Unveröffentlichte Doktorarbeit.

Nicolini, Markus (2004): *Deutsch in Texas*. Münster: LIT Verlag.

Östman, Jan-Ola/Trousdale, Graeme (2013): Dialects, Discourse, and Construction Grammar. In: Hoffmann, Thomas/Trousdale, Graeme (Hg.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press, 476–490.

Pierce, Marc/Boas, Hans C./Gilbert, Glenn (2019): When is a Dissertation Not a Dissertation? On Eikel 1954. In: *Yearbook of German-American Studies* 53, 187–196.

Pierce, Marc/Boas, Hans C./Gilbert, Glenn (2021): Fred Eikel and the Study of Texas German. In: *Yearbook of German-American Studies* 56, 131–147.

Reichel, Uwe D./Kisler, Thomas (2014): Language-Independent Grapheme-Phoneme Conversion and Word Stress Assignment as a Web Service. In: Hoffmann, Rüdiger (Hg.): *Elektronische Sprachverarbeitung.* Dresden: TUDpress (Studientexte zur Sprachkommunikation, 71), 42–49.

Riehl, Claudia (2026): *German as a Heritage Language*. Cambridge: Cambridge University Press.

Rosenberg, Peter (2005): Dialect Convergence in the German (Sprachinseln) Speech Islands. In: Auer, Peter/Hinskens, Frans/Kerswill, Paul (Hg.): *Dialect Change: Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, 221–235.

Salmons, Joseph (1983): Issues in Texas German Language Maintenance and Shift. In: *Monatshefte* 75 (2), 187–196.

Schmidt, Thomas/Blevins, Margaret/Boas, Hans C./Gilbert, Glenn G. (2026a): The Texas German Dialect Project Corpus as a Diachronic Resource for Investigating Language Contact. In: *Proceedings of Workshop on Dialects in Natural Language Processing – A Resource Perspective (DialRes)*, Mallorca, Spain, 221–229.

Schmidt, Thomas/Ferger, Anne/Frick, Elena (2026b): Putting Things on Top of Other Things. The ZuMult Platform for Multimodal Corpora and its Ecosystem. In: Grisot, Cristina/Kontino, Thalassia (Hg.): *Selected Papers from the CLARIN Annual Conference 2025, Vienna*. Linköping: Linköping University Electronic Press, 16–20.

Stift, Ulf-Michael/Schmidt, Thomas (2014): Mündliche Korpora am IDS: Vom Deutschen Spracharchiv zur Datenbank für Gesprochenes Deutsch. In: Institut für Deutsche Sprache (Hg.): *Ansichten und Einsichten. 50 Jahre Institut für Deutsche Sprache.* Redaktion: Melanie Steine, Franz Josef Berens. Mannheim: IDS Mannheim. 360–375.

Tacchetti, Maddalena (2018): User Guide for ELAN Linguistic Annotator. Version 5.0.0. https://www.mpi.nl/corpus/html/elan_ug/pr01.html [Stand: 02.11.2026].

Westpfahl, Swantje/Schmidt, Thomas (2016): FOLK-Gold – A GOLD Standard for Part-of-Speech Tagging of Spoken German. In: *Proceedings of the Tenth Conference on International Language Resources and Evaluation (LREC 16)*, Paris, 4215–4220.

Wilson, Joseph (1986): The English Spoken by German Americans in Central Texas. In: Paul Schach (Hg.): *Languages in Conflict: Linguistic Acculturation on the Great Plains* Lincoln: University of Nebraska Press, 157–173.

Wörner, Kai (2012). Finding the balance between strict defaults and total openness – Collecting and managing metadata for spoken language corpora with the EXMARaLDA Corpus Manager. In: Schmidt, Thomas/Wörner, Kai (Hg.): *Multilingual Corpora and Multilingual Corpus Analysis* 14, Amsterdam; Philadelphia: John Benjamins Pub. Co (Hamburg Studies on Multilingualism), 383–400.

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 (3), 210–232.

Abkürzungen

COMA = (EXMARaLDA) Corpus Manager
ELAN = EUDICO Linguistic Annotator
EXMARaLDA = Extensible Markup Language for Discourse Annotation
GTXG = Gilbert Texas German Corpus
ISO = International Standards Organisation
TEI = Text Encoding Intiative
TGDA = Texas German Dialect Archive
TGDP = Texas German Dialect Project
ZuMult = Zugänge zu multimodalen Korpora gesprochener Sprache