

Peter Meyer/Doris Stolberg

Eine Plattform für deutschsprachige Varietäten im Kontakt

Korpusabfrage, Sprachanalyse und Wörterbuch zur Untersuchung lexikalischer Kontaktphänomene

Abstract Within the project *Lexical Dynamics of German Varieties in Contact* at the Leibniz Institute for the German Language (IDS), lexical structures of German varieties spoken outside Central Europe's main German-speaking area are analyzed through corpus-based linguistic methods. A central question is whether lexical changes arise predominantly from particular language constellations or more broadly from general contact conditions. In addition, borrowability across word classes is investigated. This paper introduces a web-based research and analysis platform currently under development. It allows linguistic researchers to annotate corpus data flexibly, produce structured lexicographic documentation, and interactively explore linguistic phenomena. Content is presented through systematically interconnected units covering individual lexical items, word combinations, and abstract lexical patterns (e. g., discourse markers, modal particles, auxiliary selection). Navigation is facilitated by full-text searches and a thematic tagging system. The individual units of information provide dynamic access to corpus excerpts, query results, and interactive statistical analyses.

Keywords spoken language corpora, variational linguistics, language contact, lexicographical databases

1. Grundsätzliches: Das Projekt *Lexikalische Dynamik deutschsprachiger Varietäten im Kontakt*

Sprachkontaktprozesse sind durch eine besondere Dynamik gekennzeichnet, speziell im mündlichen und informellen Sprachgebrauch. Im Rahmen unseres Projekts zu deutschsprachigen Kontaktvarietäten untersuchen wir komparativ die Lexik ausgewählter Varietäten des Deutschen, insbesondere solcher, die außerhalb des geschlossenen deutschen Sprachgebiets in Mitteleuropa gesprochen werden.

Beim Vergleich der Lexik unterschiedlicher extraterritorialen Varietäten des Deutschen ergibt sich aus methodischer Sicht die Herausforderung, Kontaktphänomene auf der Grundlage von Daten aus verwandten, aber unterschiedlichen deutschen Ausgangsvarianten zu vergleichen. Extraterritoriale Varietäten des Deutschen basieren auf verschiedenen Ausgangsvarianten, z. B. im Fall von neuseeländischem Deutsch (Bairisch) oder pennsylvanischem Deutsch (Pfälzisch) im Vergleich zu australischem Deutsch und namibischem Deutsch, die dem heutigen Standarddeutsch näherstehen. Es fehlt also ein zuverlässiges *Tertium Comparationis*. Dieser Herausforderung kann zumindest partiell dadurch begegnet werden, dass bei einer anscheinenden Parallelität der Kontaktfolgen die tatsächliche Ausgangsvariante bestmöglich berücksichtigt wird und die unterschiedlichen dialektalen

Ausgangslagen in Betracht gezogen werden. Des Weiteren können die außersprachlichen Bedingungen, z. B. die Sprachkontaktsituationen, sehr unterschiedlich sein. So ist z. B. in Namibia das Standarddeutsche medial und in der schulischen Bildung im Umfeld der deutschsprachigen Gemeinschaft sehr präsent (Wiese et al. 2017). Demgegenüber findet die Weitergabe des Deutschen in Australien, wenn überhaupt, innerfamiliär und primär mündlich statt (Clyne 1981; Riehl 2015). Diese Faktoren wirken sich auf Entlehnungen und andere lexikalische Kontakteinflüsse in unterschiedlicher Weise aus.

Zu den (mikro-)diachronen Sprachkontaktprozessen im Lexikon, die das Projekt untersucht, gehören u. a. Phänomene wie lexikalische Entlehnungen in verschiedenen Stadien der Integration, Verschiebungen der Bedeutung und der Verwendungsmuster bestehender Lexeme oder neue Formen und Funktionen lexikalisch-pragmatischer Elemente. Neben der Untersuchung von einzelwortbezogenen Entlehnungen stehen dabei insbesondere Phänomene im Bereich der lexikalischen Syntagmatik und Pragmatik im Fokus.

Gerade bei mündlichen Daten ist es oft schwierig, zwischen spontanen Phänomenen des mehrsprachigen Sprachgebrauchs (z. B. Code-Switching oder Ad-hoc-Entlehnungen) und etablierten lexikalischen Kontaktprodukten zu unterscheiden (Poplack 2018). Hier können Unterschiede in der Verwendungshäufigkeit wichtige Anhaltspunkte liefern, in dem Sinne, dass spontane Verwendungen nur vereinzelt auftreten, etablierte Entlehnungen dagegen in höherer Zahl und personenübergreifend verwendet werden. Große, digital durchsuchbare Korpora stellen somit ein wichtiges Instrument in Hinblick auf den Entlehnungsstatus von Lexemen oder ihren Verwendungsweisen dar, da solche Unterschiede erst bei einer ausreichend großen Datenmenge zuverlässig identifizierbar sind.

Die zunehmende Standardisierung (vgl. dazu Hedeland/Schmidt 2022) und Verfügbarkeit größerer mündlicher Korpora ermöglicht es daher, den Untersuchungsschwerpunkt auf gesprochensprachliche Daten zu legen und darauf aufbauend internetlexikographische Dokumentationsformate zu entwickeln. Eine wesentliche Datengrundlage des Projekts sind gesprochensprachliche Korpora zu Diaspora-Varietäten in der Datenbank für Gesprochenes Deutsch (DGD) des Leibniz-Instituts für Deutsche Sprache (IDS), zunächst zum Australiendeutschen, Unserdeutschen (Papua-Neuguinea, Australien), Namibiadeutschen und zukünftig zum Kaukasiendeutschen. In der projektbezogenen Forschung geht es primär um die Untersuchung und Erfassung unterschiedlicher Arten von lexikalischen Strukturmustern, die in ihrer Sprachkontaktabhängigkeit, ihrer Entwicklung und ihren (De-)Stabilisierungstendenzen in verschiedenen Typen deutscher Varietäten untersucht und verglichen werden (vgl. für einschlägige Untersuchungen zu Australiendeutsch u. a. Clyne 1981; Clyne 2003; Riehl 2015; Riehl/Plewnia 2018; zu Namibiadeutsch u. a. Shah 2007; Shah/Zimmer 2023; Zimmer et al. 2020; Wiese et al. 2014; Wiese et al. 2017; zu Unserdeutschem u. a. Lindenfelser 2021; Maitz 2016, 2019; Maitz/Volker 2017). Eine der zentralen Forschungsfragen ist, inwieweit Sprachwandel von spezifischen Kontaktsprachkonstellationen geprägt oder aber allgemeiner durch die Kontaktsituation an sich bedingt ist.

Ein weiteres Forschungsinteresse ist darin begründet, dass unterschiedliche lexikalische Elemente verschieden häufig entlehnt werden. Es wird seit langem angenommen, dass lexikalische Entlehnbarkeit in engem Zusammenhang mit der Zugehörigkeit zu einer bestimmten Wortart steht. Diese Annahme beruht auf der Beobachtung, dass in Sprachkontaktumgebungen, die lexikalische Entlehnungen begünstigen, Substantive in vergleichsweise höherer Zahl entlehnt werden, gefolgt von Verben und Adjektiven, während Funktionswörter deutlich seltener entlehnt zu werden scheinen (vgl. bereits Haugen 1950; Weinreich 1953). Entlehnbarkeitsskalen, die der Abstraktion dieser Beobachtungen dienen, können lexikalische Entlehnungsmuster beschreiben und bis zu einem gewissen Grad vorhersagen.

Zusätzlich werden soziolinguistische Faktoren zur Feinabstimmung der Skalen in entsprechende Modelle integriert und zur Erklärung von beobachteten Abweichungen argumentativ herangezogen, wie bereits z. B. Thomason/Kaufman (1988) darlegen.

Als mögliche Ursache für die beobachteten Unterschiede in der lexikalischen Entlehnbarkeit wird, neben sozialen Faktoren, häufig eine unterschiedliche grammatische Komplexität der Wortarten angenommen, die z. B. verhindert, dass Verben mit der gleichen Häufigkeit entlehnt werden wie Substantive, da die Verwendung von Verben morphosyntaktisch komplexer sei als die von Substantiven (vgl. Haspelmath 2009 für eine kurze Darstellung dieser Annahme). Matras/Adamou (2023, S. 5) hingegen argumentieren, dass Entlehnbarkeit eher als ein direktes Produkt kommunikativer, sozialer und pragmatischer Motivationen zum Entleihen betrachtet werden sollte und weniger als eine Frage formaler Zwänge,¹ da die Sprachkontaktforschung gezeigt hat, dass grundsätzlich alle lexikalischen Elemente entlehnbar sind. Aus ihrer Sicht stellt eine Wortart zudem keine einheitliche Kategorie dar, die eine hohe oder niedrige Entlehnbarkeit vorhersagt. Vielmehr kann es innerhalb einer Wortart ‚gespaltene Kategoriewerte‘ („split category values“, Matras/Adamou 2023, S. 6) geben. So zeigen verschiedene Gruppen von Funktionswörtern eine unterschiedliche Entlehnungshäufigkeit; Diskursmarker beispielsweise werden wesentlich häufiger entlehnt als Pronomina. Jedoch auch innerhalb von Wortklassen im engeren Sinne gibt es Abstufungen, so dass z. B. Präpositionen, die Distanz und komplexe Bezugspunkte oder kontrastive Bedeutung kodieren, häufiger sprachübergreifend entlehnt werden als andere Arten von Präpositionen.

Eines der Ziele der zu entwickelnden Online-Ressource, die im Folgenden genauer beschrieben wird, ist es, einen Beitrag zur Beantwortung der oben umrissenen Forschungsfragen zu leisten. Die vergleichende Untersuchung extraterritorialer Varietäten des Deutschen ermöglicht es, nicht nur allgemeine Entlehnungstendenzen von Partikularentwicklungen zu unterscheiden, sondern erlaubt auch einen Einblick in Entlehnungspräferenzen in vergleichbaren Kontexten, nämlich der informellen mündlichen Interaktion in verwandten Varietäten, in jeweils unterschiedlichen Kontaktkonstellationen.

2. Eine Forschungsplattform für Diaspora-Varietäten des Deutschen

Der vorliegende Beitrag stellt eine derzeit in Entwicklung befindliche webbasierte Forschungs- und Analyseplattform vor. Ihr Ziel ist es, einem sprachwissenschaftlich informierten Publikum innovative Möglichkeiten zur Exploration von Untersuchungsergebnissen des Projekts zu bieten, indem Beschreibungstexte mit interaktiven Zugängen zu den zugrundeliegenden Korpusdaten verknüpft werden. Die Plattform wird voraussichtlich im Laufe des Jahres 2026 als eigenständiges Onlineangebot des IDS zugänglich werden und nach einer kostenfreien Registrierung kostenfrei nutzbar sein.

Die verschiedenen Informationsangebote und Funktionalitäten der Plattform werden im Folgenden jeweils aus drei Perspektiven vorgestellt:

1 „ ‚Borrowability‘ should [...] be considered a direct product of communicative, social and pragmatic motivations to borrow rather than a matter of formal constraint.“ (Matras/Adamou 2023, S. 5).

- a. Zunächst wird die Verwendung der Online-Forschungsplattform, die über einen Browser aufgerufen wird, aus der Sicht der Nutzer:innen betrachtet.
- b. Aus der Perspektive der am Aufbau der Plattform beteiligten Linguist:innen werden anschließend die redaktionellen Abläufe, die für die Erstellung der Inhalte erforderlich sind, sowie die dabei verwendeten Softwarewerkzeuge skizziert.
- c. In einem letzten Schritt wird dann aus Entwickler:innensicht jeweils die zugrundeliegende Datenmodellierung, die Umsetzung durch Datenbanken und Schnittstellen sowie die technische Umsetzung in der Webapplikation diskutiert. Der Fokus der Betrachtungen liegt dabei auf der Konzeption und grundsätzlichen Architektur des Systems; Hinweise auf konkrete Implementierungsstrategien sind vorläufiger Natur.

Als konkretes Fallbeispiel sollen im Folgenden Untersuchungen zu Neubedeutungen (Sinnerweiterungen und -verschiebungen) deutscher lexikalischer Einheiten², insbesondere Verben, dienen, die zumindest potenziell aus Kontaktsprachen übernommen wurden.

2.1 Informationseinheiten

Das Informationsangebot der Plattform ist ganz allgemein in Einzeltexte gegliedert, die hier als *Informationseinheiten* bezeichnet werden sollen. Informationseinheiten können zum einen lexikologische Phänomene einer Einzelsprache beschreiben, von konkreten Lexemen bis hin zu bestimmten Lexemklassen, syntagmatischen Mustern und Lexikalisierungs- und Univerbierungsprozessen. Ein konkretes Beispiel wäre eine Informationseinheit zu Verwendungsweisen des Verbs *nehmen* im Australiendeutschen (DGD-Korpus: AD), die mutmaßlich durch Sprachkontakt mit dem Englischen entstanden sind, etwa die Lesart ‚jemanden/etwas an einen Ort bringen‘ wie in

- (1) Ich mußte erst deutsch lernen, dann hat meine Tante mich zum Adelaide genommen [...]

(IDS, DGD, AD--_E_00057_SE_01_T_01, [00:05:04.08]) [<http://dgd.ids-mannheim.de>, Stand: 3.6.2025]

Entsprechend sind auch Informationseinheiten denkbar, die das Auftreten solcher Lehnbedeutungen für mehrere australiendeutsche Lexeme gemeinsam beschreiben.

Informationseinheiten können zum anderen aber auch kontrastiv bzw. mit den Mitteln der Variationslinguistik sprachübergreifende Untersuchungsergebnisse präsentieren, in Bezug auf (1) beispielsweise allgemein Lehnbedeutungen, die in Varietäten wie dem Australien- und Namibiadeutschen auf englischen Sprachkontakt zurückgehen können. Auf einer noch allgemeineren Ebene sind Übersichtseinheiten zu bestimmten Themen angesiedelt; so könnte es etwa eine allgemeine Informationseinheit zu Lehnbeziehungen geben. Durch Querverweise auf Einheiten, in denen konkrete Aspekte des Themas behandelt werden, können solche Überblicksartikel auch als Navigationshilfen im Gesamtangebot der Plattform dienen.

Aus redaktioneller Sicht sind Informationseinheiten XML-Dokumente, die in einem auf der Plattform bereitgestellten Redaktionssystem von autorisierten Linguist:innen erstellt

2 Lexikalische Einheiten können Einzelwörter, aber auch Mehrworteinheiten, d. h. mehr oder weniger feste Kombinationen aus mehreren Wörtern, sein.

und kollaborativ bearbeitet werden. Dies wirft die Frage nach der Modellierung und internen Struktur dieser Dokumente auf. Anders als bei einem Wörterbuch werden, wie oben beschrieben, auf der Plattform sehr heterogene Klassen von Inhalten verwaltet. Je nach linguistischer Thematik stehen aber auch schon bei Informationseinheiten, die beispielsweise „nur“ Einzelllexeme einer Sprache behandeln, recht unterschiedliche lexikalische Aspekte im Vordergrund, so dass nicht einmal in dieser Untergruppe von Informationseinheiten eine einheitliche, granulare mikrostrukturelle Modellierung praktikabel erscheint. Informationseinheiten werden daher überwiegend als hierarchisch nach Abschnitten und Unterabschnitten gegliederter Freitext gestaltet. Dennoch sind für bestimmte Klassen von Einheiten jeweils unterschiedliche formalisierte Abschnitte sinnvoll. Im Falle von Einheiten zu Einzelwörtern entsprechen diese beispielsweise typischen lexikographischen Angaben zur Wortart, Morphosyntax, Etymologie usw. Das Redaktionssystem muss also verschiedene Typen von Informationseinheiten anbieten. Jedem Typ ist – zur Validierung der Eingaben – ein eigenes XML-Schema sowie ggf. zusätzlich eine eigene Dokumentation mit redaktionellen Vorgaben zugeordnet.

Auf der Ebene der Softwarearchitektur ist eine dokumentenorientierte Datenbank die zentrale Komponente des Plattform-Backends. Jede Informationseinheit wird zusammen mit Verwaltungsinformationen, u. a. einem volltextindizierbaren Auszug aus dem XML-Quelltext, als separates Dokument in der Datenbank abgelegt. Neue Typen von Informationseinheiten können jederzeit modular dadurch hinzugefügt werden, dass ein zugehöriges XSLT-Stylesheet und ein Schemadokument in entsprechende Verzeichnisse abgelegt werden.

2.2 Zugriffsstrukturen

Die inhaltlichen Beziehungen zwischen Informationseinheiten der Plattform sind im Allgemeinen nicht systematischer oder hierarchischer Natur, auch wenn es natürlich „vertikale“ thematische Verhältnisse von über- zu untergeordneten Einheiten geben kann. Offensichtlich lassen sich Informationseinheiten auch nicht auf naheliegende Weise als sortierte Liste anbieten, abgesehen vielleicht von Einheiten zu Einzelllexemen. Es gibt also keine für die Plattform geeignete elementare Makrostruktur, die einer Lemmaliste im Wörterbuch vergleichbar wäre. Das System muss aus diesem Grund geeignete Zugriffsstrukturen und Suchoptionen für Endnutzer:innen anbieten.

- Grundlegend sind Verweisverfahren: Die Informationseinheiten sind über Querverweise systematisch miteinander vernetzt. Im laufenden Text einer Einheit funktionieren diese Verweise als Verlinkungen, mit denen jeweils eine andere Informationseinheit bzw. darin enthaltene Abschnitte aufgerufen werden können. Zusätzlich kann jede Einheit separate Abschnitte anbieten, die, soweit möglich und sinnvoll, thematisch über- und untergeordnete sowie anderweitig inhaltlich zugehörige Einheiten auflistet, ähnlich wie dies von Wiki-Datenbanken vertraut ist. Spezielle Informationseinheiten können als Einstieg in eine bestimmte Thematik dienen und liefern in strukturierter Weise kommentierte Verweise auf zugehörige Einheiten der Plattform. Auch die Startseite des Gesamtsystems kann letztlich eine solche – im Weiteren als *navigational* bezeichnete – Informationseinheit sein, die ihrerseits auf wichtige thematische Übersichtseinheiten, aber auch auf einzelne aktuell neu hinzugekommene oder anderweitig relevante Einheiten verweist.

- Wesentliche Zugriffsstruktur der Plattform ist ein dynamisch erweiterbares System von ‚semantischen‘ Etiketten (Tags), die Informationseinheiten beispielsweise hinsichtlich der in ihnen behandelten deutschen Varietäten und Kontaktsprachen sowie linguistischen Themen einordnen. Durch Auswählen eines oder mehrerer Tags können die Nutzer:innen der Plattform die Gesamtliste der Einheiten effektiv filtern, wobei navigationelle Einheiten in der Resultatliste besonders hervorgehoben werden. Die einer angezeigten Einheit zugeordneten Tags erscheinen visuell prominent in ihrer Online-Präsentation. In unserem Beispiel (1) einer Einheit zu kontaktinduzierten Lesarten des australiendeutschen *nehmen* sind naheliegende Tags „Varietät: Australiendeutsch“, „Kontaktsprache: Englisch“, „Thema: Lehnbeziehungen“ und ggf. auch „Thema: Argumentstrukturen“.
- Gefiltert werden können Informationseinheiten aber auch durch eine Volltextsuche. Dies kann die Möglichkeit einschließen, nach „Metawörtern“ zu suchen, die im eigentlichen Text der betreffenden Einheit womöglich gar nicht auftauchen, aber für die Einheit inhaltlich einschlägig sind, ähnlich den Keywords, die für Webseiten zur Suchmaschinenoptimierung hinterlegt werden können. Das können beispielsweise standarddeutsche Metalemmata sein, wenn das in der Einheit beschriebene Wort auf eine dialektale Form der deutschen Ausgangsvarietät zurückgeht (z.B. das diskursstrukturierende Element „weißt du“ als Metalemma zur dialektalen Ausprägung „weischt“).

Die beschriebenen Zugriffsstrukturen lassen sich in der Benutzeroberfläche als responsive Seitenleiste gestalten, die Texteingaben bzw. Dropdown-Listen für die auswählbaren Filter zur Verfügung stellt und eine entsprechend gefilterte und sortierbare Liste von Informationseinheiten anbietet. Das Anklicken eines Elements der gefilterten Liste öffnet dann im zentralen Bereich die ausgewählte Einheit. Abbildung 1 zeigt einen Entwurf der für alle Nutzer:innen der Plattform zugänglichen Such- und Ansichtsseite für Informationseinheiten.

Im Redaktionsprozess werden Tags und Metawörter als Metadaten, d.h. letztlich durch XML-Elemente in einem ‚Header‘-Bereich, realisiert. Dazu gehört auch der Status einer Einheit als navigationell. Verweise auf andere Einheiten werden über eindeutige IDs ermöglicht und können im Redaktionssystem durch direkte Auswahl der Zieleinheit erzeugt werden. Implementierungsseitig wird die Suche durch Indizierung der Metadatenfelder der Datenbankdokumente sowie durch einen Volltextindex umgesetzt.

Lexikalische Dynamik deutschsprachiger Varietäten im Kontakt RESSOURCEN PROJEKT ?

Volltextsuche

Australiendeutsch ×

Lehnbeziehungen ×

+ TAG ▾

6 Suchergebnisse

Australiendeutsch geben
Wortartikel

Australiendeutsch naja
Wortartikel

Australiendeutsch nehmen
Wortartikel

Australiendeutsch well
Wortartikel

Australiendeutsche Diskurspartikeln
einzelsprachlicher Übersichtsartikel

Australiendeutsche Verben mit entlehnten Bedeutungen
einzelsprachlicher Übersichtsartikel

Wortartikel: Australiendeutsch *nehmen*

Australiendeutsch Argumentstrukturen Lehnbeziehungen

Kontaktsprache Englisch

nehmen + Zeit

Englisch: sth takes (Zeiteinheit)
Standarddeutsch: etw dauert (Zeiteinheit)

Belege im AD-Korpus

Bigramme im AD-Korpus

☐ 'nehmen' an Position 1 ☒ 'nehmen' an Position 2

Wort	Wortart	Wort ↑	Wortart	Häufigkeit
– Jahre	NN	genommen	VVPP	1

KWICs für dieses Bigramm

das Land wieder reine machen , und es hat viele Jahre **genommen** , viel Arbeit . Und denn äh ((...)) durch das ((...)) die ändert - Zeiten

Wort	Wortart	Wort ↑	Wortart	Häufigkeit
+ Monat	NN	genommen	VVPP	1

Abb. 1: Entwurf der Nutzeroberfläche des Onlinesystems. Gezeigt wird eine Informationseinheit zu kontaktinduzierten Lesarten des australiendeutschen Verbs *nehmen*. Im linken Seitenbereich erscheint eine Liste von Informationseinheiten, die hier bereits nutzerseitig nach den Tags „Australiendeutsch“ sowie „Lehnbeziehungen“ gefiltert wurde.

2.3 Interaktive Komponenten und Korpuszugriff

Die Informationseinheiten der Plattform sind nicht als vorgefertigte, statische lexikografische Beschreibungen konzipiert, sondern sollen einen konfigurierbaren, auf den jeweils relevanten Forschungskontext zugeschnittenen Zugriff auf zugehörige Daten aus den verwendeten Korpora gewähren. Eine besondere Rolle spielen für das Gesamtprojekt dabei die mündlichen Korpora zu deutschen Diasporavarietäten in der *Datenbank für Gesprochenes Deutsch* (DGD) des IDS, zu denen man nach kostenfreier Registrierung Zugang erhält. Die Ein- oder Anbindung weiterer, auch schriftsprachlicher, Korpora ist jedoch möglich und vorgesehen.

Ähnlich wie in LeGeDe („Lexik des gesprochenen Deutsch“, www.owid.de/legede/, Stand: 17.12.2025), einem am IDS entwickelten Prototyp eines Wörterbuchs zur Lexik des gesprochenen Deutsch, können insbesondere, nach Anmeldung mit den DGD-Zugangsdaten, zum jeweiligen Kontext in der Informationseinheit passende Transkriptausschnitte aus den DGD-Korpusdaten präsentiert und dabei dergestalt mit einer Audiowiedergabe kombiniert werden, dass die aktuell zu hörende Position im Transkript unterlegt ist. Dabei wird der unter dem Namen ‚ZuViel‘ firmierende Visualisierungszugang der für audiovisuelle Korpora entwickelten und u. a. für die DGD genutzten ZuMult-Softwarearchitektur (zumult.org) genutzt; vgl. Abbildung 2. Im Gegensatz zu LeGeDe soll es die hier beschriebene Plattform

aber ermöglichen, nicht nur einzelne redaktionell ausgewählte Transkriptausschnitte anzeigen zu lassen, sondern sämtliche der im Kontext passenden Korpusbelege in tabellarischer, durchsuchbarer, filterbarer und sortierbarer Form als Keywords in Context (KWICs), zu denen dann per Mausklick interaktive Transkripte erscheinen.

The screenshot shows the DGD-Visualisierungszugang interface. On the left, there is a list of KWICs (Keywords in Context) for the verb 'nehmen'. The list includes entries like '0012 MC Was für Bäume waren denn ((???)) ? Was für Bäume waren denn?', '0013 FS Was wir, was wir Rotgummi nennen.', '0014 MC Und, äh, wie ist es jetzt was haben die hier angefangen?', '0015 FS Ja, well, seitdem die haben es angefangen, wieder zu farmerieren, da haben wer denn die Bäume haben müssen wieder aus, wegmachen und denn das Land wieder reine machen, und es hat viele Jahre **genommen**, viel Arbeit. Und denn äh ((...)) durch das ((...)) die ändert- Zeiten verändern sich da haben wir an- angefangen Frucht und Wein zu bauen. Denn durch die Weinbauerei ist das- der ganze District hier richtig in Gange gekommen.', '0016 MC Und was für Obstbäume haben Sie denn hier in der Gegend?', and '0017 FS Alle, alle Sorten Obst- Äpfel, Birnen, Quitten und Pfirsichen, Feigen, alle, alle Sorten, äh ((...)) äh,'. On the right, there is a detailed transcript view for the selected entry '0015 FS'. It shows a timeline from 01:37 to 02:07 and a play button. The transcript text is: 'Ja, well, seitdem die haben es angefangen, wieder zu farmerieren, da haben wer denn die Bäume haben müssen wieder aus, wegmachen und denn das Land wieder reine machen, und es hat viele Jahre **genommen**, viel Arbeit. Und denn äh ((...)) durch das ((...)) die ändert- Zeiten verändern sich da haben wir an- angefangen Frucht und Wein zu bauen. Denn durch die Weinbauerei ist das- der ganze District hier richtig in Gange gekommen.'

Abb. 2: Audiovisuelle Präsentation eines Transkriptausschnittes mit dem DGD-Visualisierungszugang ZuViel. Der Ausschnitt enthält den in Abbildung 1 gezeigten KWIC zum Verb *nehmen* in der entlehnten Bedeutung ‚dauern‘.

In unserem Beispiel einer Informationseinheit zu Lehnbedeutungen der australiendeutschen Verwendung von *nehmen* heißt dies, dass zu jeder einzelnen Lehnbedeutung alle relevanten KWICs aus dem AD-Korpus der DGD angezeigt werden können. Neben der oben genannten Bedeutung ‚jemanden/etwas an einen Ort bringen‘ ist dies etwa die Bedeutung ‚eine bestimmte Zeitspanne in Anspruch nehmen‘, wie in

- (2) [...] das hat sie sechs Monat genommen, bis sie sind hergekommen hier.

(IDS, DGD, AD--_E_00039_SE_01_T_01, [00:08:48.17]) [<http://dgd.ids-mannheim.de>, Stand: 3.6.2025]

Um eine solche Funktionalität zu ermöglichen, muss das betreffende Korpus mithilfe eines für autorisierte Linguist:innen zugänglichen Redaktionswerkzeugs annotiert werden. Dazu wird für die Lehnbedeutungen von *nehmen* eine eigene Annotationsschicht angelegt, die nicht in der DGD, sondern in der Datenbank der Plattform verwaltet und gespeichert wird. Mithilfe einer DGD-Abfrage in der Abfragesprache CQP (vgl. Frick/Helmer/Wallner 2023, S. 50 mit Verweisen auf weitere Informationen) kann die bearbeitende Person im Annotationstool z. B. alle KWICs im Korpus AD anzeigen lassen, die eine Form des Lexems *nehmen* enthalten. Jedem *nehmen*-Token kann dann, im elementarsten Fall mit einem Dropdown-Auswahlmenü, von dem/der Annotator:in eine der im Tool vordefinierten Lesarten zugeordnet werden. Um sodann in der Informationseinheit eine KWIC-Liste für *nehmen*-Token in einer bestimmten Lesart zu generieren, muss von der Autorin bzw. dem Autor der Einheit im Editor lediglich ein einzelnes, für diesen Zweck vordefiniertes, XML-Element eingefügt werden, das über Attribute die Annotationsschicht und den annotierten Wert (Lesart) spezifiziert.

In der technischen Realisierung des Annotationswerkzeuges entspricht jedes annotierte Token einem eigenen Dokument in der Datenbank der Plattform. Dieses Dokument nimmt

zum einen mit bestimmten IDs auf das betreffende annotierte Token in der DGD Bezug, z. B. mit einer ID für das jeweilige DGD-Transkript sowie einer ID für das konkrete Token darin. Zum anderen speichert das Dokument die jeweils zugehörige Annotationsinformation, im konkreten Beispielfall also die Zuordnung zu einer Annotationsschicht für Lehnbedeutungen eines bestimmten Verbs sowie die dem Token zugeordnete Lehnbedeutung. Durch dieses Vorgehen entfällt die Notwendigkeit, Daten der DGD-Datenbank zu duplizieren. Das Annotationswerkzeug kommuniziert über leistungsfähige Schnittstellen der bereits erwähnten ZuMult-Architektur mit der DGD (vgl. Frick/Helmer/Wallner 2023). Lässt ein:e Endnutzer:in der Plattform die KWICs einer Lesart anzeigen, werden diese ebenfalls in Echtzeit über die ZuMult-Schnittstelle abgerufen.

Aus dem XML-Element generiert ein XSLT-Stylesheet die browserseitige dynamische KWIC-Listendarstellung für die Nutzer:innen. Dazu wird eine HTML-Komponente sowie eine Serverschnittstelle für die erforderlichen komponentenspezifischen Datenbankabfragen implementiert. Im Projekt kann für verschiedene interaktive Elemente eine ganze Bibliothek solcher HTML-Elemente erstellt werden, die dann von Autor:innen in jeder Informationseinheit nach Belieben genutzt werden kann.

Auf einer abstrakteren Ebene können aber auch aggregierende, quantitative Auswertungen über dem Korpusdatenbestand in eine Informationseinheit integriert werden. Für derzeit zwei DGD-Korpora ist eine externe Webanwendung, der Lexical Explorer (Lemmenmeier-Batinić 2020), entwickelt worden, die vorausberechnete Häufigkeitslisten und tabellarische Auswertungen zu den in den betreffenden Korpora verfügbaren Annotationsmerkmalen und Metadaten bietet. Vergleichbare Funktionalität soll, direkt eingebettet in die Informationseinheiten der neuen Plattform und jeweils kontextuell begrenzt auf relevante Wörter, insbesondere auch für die plattformspezifischen Annotationen zur Verfügung stehen.

Im Beispiel des australiendeutschen *nehmen* können beispielsweise bei der Behandlung einer bestimmten Lehnbedeutung interaktive Tabellen von Bi- und Trigrammen sowie Kookkurrenzen von *nehmen* in der betreffenden Lesart angezeigt werden, einschließlich statistischer Angaben zu Frequenz und Signifikanz der jeweiligen Wortverbindung und der Möglichkeit, sich alle KWICs mit der jeweiligen Wortverbindung anzeigen zu lassen. Die Lesarten können auch mit Metadaten wie dem Sprecher-geschlecht oder dem Typ der mündlichen Interaktion kreuztabelliert werden. Visualisierungen, z. B. Balken- oder Tortendiagramme zur Häufigkeitsverteilung der einzelnen Lehnbedeutungen von *nehmen* im Korpus, lassen sich ebenfalls realisieren.

Obwohl derlei quantitative Auswertungen im Prinzip automatisch über Abfragen im Korpus sowie in der Plattform-Datenbank gewonnen werden können, ist in vielen Fällen eine Echtzeitberechnung nicht möglich. So müssen für Signifikanzmaße bei Bigrammen, wie etwa der Log Likelihood Ratio, für jedes in einer Informationseinheit zu behandelnde Bigramm sämtliche anderen Bigramme des Korpus ausgezählt werden, die dasselbe Erst- bzw. Zweitelement aufweisen. Solche vorzuberechnenden Auswertungen müssen nach jeder Veränderung am Korpus- oder Annotationsbestand erneut programmgesteuert durchgeführt und die Resultate in der Datenbank der Plattform hinterlegt werden. Im Falle einer Bigrammtabelle würde jedes einzelne Bigramm mitsamt zugehörigen statistischen Maßen durch ein eigenes Datenbankdokument repräsentiert.

Abschließend sei auf die Möglichkeit hingewiesen, dass eine Informationseinheit über ein speziell zu diesem Zweck implementiertes benutzerdefiniertes HTML-Element Zugriff auf eine gesamte, in die Plattform eingebundene, Ressource gewährt und dabei auf diese Ressource zugeschnittene Suchfunktionen bereitstellt. Ein einfaches Beispiel wäre ein Wör-

terbuch zu einer bestimmten Diasporavarietät mitsamt filterbarer Lemmaliste. Auch kleinere schriftsprachliche Korpora könnten auf diese Weise eingebunden werden. Im Falle einer integrierten lexikographischen Ressource kann es von der Anzahl der Lemmata abhängen, ob es sinnvoller ist, jedes einzelne Lemma als separate Informationseinheit anzubieten, statt das gesamte Wörterbuch in einer einzigen Einheit zu präsentieren.

3. Fazit

Die hier vorgestellte, in der Entwicklung befindliche Plattform trägt dazu bei, mündliche Korpora deutschbasierter Kontaktvarietäten systematisch in Bezug auf lexikalische Kontaktphänomene zu untersuchen. Sie bietet vielfältige und neuartige technische Möglichkeiten, um kontaktinduzierte Entwicklungen in verschiedenen extraterritorialen Varietäten des Deutschen vergleichend zu betrachten und zu beschreiben. Dabei kann der Fokus je nach Forschungsinteresse auf bestimmte Wortarten, auf Einzellexeme oder auf andere Ordnungskategorien gelegt werden, um unterschiedliche Kontaktkonstellationen und deren lexikalische Auswirkungen strukturiert zu untersuchen. Dadurch ermöglicht sie es, sich der Frage nach unterschiedlichen Entlehnbarkeiten zu nähern und zu deren theoretischer Modellierung beizutragen. Ebenso erlaubt sie es, Unterschiede zwischen allgemeinen Sprachwandeltendenzen in deutschen Kontaktvarietäten einerseits und konstellationsspezifischen Entwicklungen andererseits schärfer herauszuarbeiten. Indem die traditionellen Grenzen zwischen Korpusabfrage, Sprachanalyse und Wörterbuch zunehmend aufgelöst werden, wird so aus einem digitalen Wörterbuchportal ein ebenso komplexes wie flexibles lexikologisches Forschungsinstrument.

Literatur

Clyne, Michael (1981): Deutsch als Muttersprache in Australien. (= Deutsche Sprache in Europa und Übersee 8). Wiesbaden: Steiner.

Clyne, Michael (2003): Dynamics of language contact. (= Cambridge Approaches to Language Contact). Cambridge: Cambridge University Press.

DGD = Leibniz-Institut für Deutsche Sprache (o. D.): Datenbank für Gesprochenes Deutsch (DGD). https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome (Stand: 15.12.2025).

Frick, Elena/Helmer, Henrike/Wallner, Franziska (2023): ZuRecht: Neue Recherchemöglichkeiten in Korpora gesprochener Sprache für Gesprächsanalyse und Deutsch als Fremd- und Zweitsprache. In: Korpora Deutsch als Fremdsprache 3, 1, S. 44–71.

Haugen, Einar (1950): The analysis of linguistic borrowing. In: Language 26, S. 210–231.

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.): Proceedings of the CLARIN Annual Conference 2021. Virtual Event, 2021, 27–29 September. (= Linköping Electronic Conference 189). Linköping: Linköping University, S. 34–45.

Lindenfelser, Siegwalt (2021): Kreolsprache Unserdeutsch. Genese und Geschichte einer kolonialen Kontaktvarietät. (= Koloniale und Postkoloniale Linguistik/Colonial and Postcolonial Linguistics [KPL/CPL] 17). Berlin/Boston: De Gruyter.

Maitz, Péter (2016): Unserdeutsch. Eine vergessene koloniale Varietät des Deutschen im melanesischen Pazifik. In: Lenz, Alexandra N. (Hg.): German abroad – Perspektiven der Variationslinguistik, Sprachkontakt- und Mehrsprachigkeitsforschung. (= Wiener Arbeiten zur Linguistik 4). Göttingen: V&R unipress, S. 211–240.

- Maitz, Péter (2019): Deutsch als Minderheitensprache in Australien und Ozeanien. In: Herrgen, Joachim/Schmidt, Jürgen E. (Hg.): Sprache und Raum. Ein internationales Handbuch der Sprachvariation. Bd. 4: Deutsch. (= Handbooks of Linguistics and Communication Science 30.4). Berlin/Boston: De Gruyter, S. 1191–1209.
- Maitz, Péter/Volker, Craig A. (2017): Documenting Unserdeutsch: Reversing colonial amnesia. In: Journal of Pidgin and Creole Languages 32, 2, S. 365–397.
- Matras, Yaron/Adamou, Evangelia (2023): Word classes in language contact. In: van Lier, Eva (Hg.): The Oxford handbook of word classes. (= Oxford Handbooks in Linguistics). Oxford: Oxford University Press, S. 887–898.
- Poplack, Shana (2018): Borrowing. Loanwords in the speech community and in the grammar. Oxford u. a.: Oxford University Press.
- Riehl, Claudia M. (2015): Language attrition, language contact, and the concept of relic variety: The case of Barossa German. In: International Journal of the Sociology of Language 236, S. 1–33.
- Riehl, Claudia M./Plewnia, Albrecht (2018): Australien. In: Plewnia, Albrecht/Riehl, Claudia M. (Hg.): Handbuch der deutschen Sprachminderheiten in Übersee. Tübingen: Narr, S. 9–32.
- Shah, Sheena (2007): German in a contact situation: The case of Namibian German. In: eDUSA 2, 2, S. 20–45.
- Shah, Sheena/Zimmer, Christian (2023): Grammatical innovations in German in multilingual Namibia: The expanded use of linking elements and *gehen* ‚go‘ as a future auxiliary. In: Journal of Germanic Linguistics 35, 3, S. 205–265.
- Thomason, Sarah G./Kaufman, Terrence (1988): Language contact, creolization and genetic linguistics. Berkeley, CA: University of California Press.
- Weinreich, Uriel (1953): Languages in contact. (= Publications of the Linguistic Circle of New York 1). Den Haag: De Gruyter.
- Wiese, Heike/Simon, Horst J./Zappen-Thomson, Marianne/Schumann, Kathleen (2014): Deutsch im mehrsprachigen Kontext: Beobachtungen zu lexikalisch-grammatischen Entwicklungen im Namdeutschen und im Kiezdeutschen. In: Zeitschrift für Dialektologie und Linguistik 81, 3, S. 274–307.
- Wiese, Heike/Simon, Horst J./Zimmer, Christian/Schumann, Kathleen (2017): German in Namibia: A vital speech community and its multilingual dynamics. In: Maitz, Péter/Volker, Craig A. (Hg.): Language & Linguistics in Melanesia (Sonderheft: Language Contact in the German Colonies: Papua New Guinea and Beyond), S. 221–245.
- 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, S. 210–232.

Kontaktinformation

Dr. Peter Meyer
Leibniz-Institut für Deutsche Sprache (IDS)
R 5, 6-13
68161 Mannheim
E-Mail: meyer@ids-mannheim.de

Dr. Doris Stolberg
Leibniz-Institut für Deutsche Sprache (IDS)
E-Mail: stolberg@ids-mannheim.de

Bibliografische Angaben

Dieser Text ist Teil der Publikation: Brunner, Annelen/Hansen, Sandra/Lang, Christian/Tu, Ngoc Duyen Tanja/Wolfer, Sascha (Hg.) (2026): Deutsch im Wandel – Tools, Methoden, Ressourcen. Beiträge zur Methodenmesse der IDS-Jahrestagung 1. (= *IDSopen* 16). Mannheim: IDS-Verlag.
<https://doi.org/10.21248/idsopen.16.2026.63>.