

# **Das Korpus *Deutsch in Namibia (DNam)*: Eine Ressource für die Kontakt-, Variations- und Soziolinguistik**

Christian Zimmer<sup>1</sup>, Heike Wiese<sup>2</sup>, Horst J. Simon<sup>1</sup>, Marianne Zappen-Thomson<sup>3</sup>, Yannic Bracke<sup>2</sup>, Britta Stuhl<sup>1</sup> & Thomas Schmidt<sup>4</sup>

Freie Universität Berlin<sup>1</sup>, Humboldt-Universität zu Berlin<sup>2</sup>, University of Namibia (Windhoek)<sup>3</sup>, Institut für deutsche Sprache (Mannheim)<sup>4</sup>

**Abstract:** Dieser Beitrag widmet sich der Beschreibung des Korpus *Deutsch in Namibia (DNam)*, das über die *Datenbank für Gesprochenes Deutsch (DGD)* frei zugänglich ist. Bei diesem Korpus handelt es sich um eine neue digitale Ressource, die den Sprachgebrauch der deutschsprachigen Minderheit in Namibia sowie die zugehörigen Spracheinstellungen umfassend und systematisch dokumentiert. Wir beschreiben die Datenerhebung und die dabei angewandten Methoden (freie Gespräche, „Sprachsituationen“, semi-strukturierte Interviews), die Datenaufbereitung inklusive Transkription, Normalisierung und Tagging sowie die Eigenschaften des verfügbaren Korpus (Umfang, verfügbare Metadaten usw.) und einige grundlegende Funktionalitäten im Rahmen der *DGD*. Erste Forschungsergebnisse, die mithilfe der neuen Ressource erzielt wurden, veranschaulichen die vielseitige Nutzbarkeit des Korpus für Fragestellungen aus den Bereichen Kontakt-, Variations- und Soziolinguistik.

This paper describes the corpus *Deutsch in Namibia (DNam)*, ‘German in Namibia’, which is openly accessible via the *Datenbank für Gesprochenes Deutsch (DGD)*, ‘Database for Spoken German’. This corpus is a new digital resource that comprehensively and systematically documents the language use of the German-speaking minority in Namibia and related language attitudes. We discuss data collection and elicitation methods (conversation groups, “language situations”, semi-structured interviews), data processing including transcription, normalisation and tagging, general corpus characteristics available (size, available metadata etc.) and some basic functionalities within the *DGD*. First research results based on this new empirical

resource illustrate its value for studies on language contact, variation and sociolinguistics.

## 1 Einleitung

Das Deutsche in Namibia hebt sich durch verschiedene Merkmale von anderen Varietäten innerhalb und außerhalb des geschlossenen deutschen Sprachraums ab.<sup>1</sup> Im Unterschied zu anderen Varietäten außerhalb, in denen ein Sprachwechsel bevorsteht oder bereits stattgefunden hat, wird Deutsch hier aktiv gepflegt, konsequent an folgende Generationen weitergegeben und in formellen wie in informellen Situationen verwendet. Dass ein Sprachwechsel nicht unmittelbar bevorsteht, ist dabei durchaus bemerkenswert, da die Community lediglich ca. 20.000 Sprecher/-innen umfasst und sich zudem (im Gegensatz z.B. zu den ebenfalls vitalen *sectarian communities* in Nordamerika) nicht systematisch von anderen Gruppen abschottet.

Im Unterschied zu Varietäten innerhalb des geschlossenen deutschen Sprachraums ist das namibische Deutsche in einen breit mehrsprachigen gesellschaftlichen Kontext eingebettet. Charakteristisch und prägend ist in diesem Kontext ein intensiver Sprachkontakt mit den beiden germanischen Sprachen Englisch und Afrikaans. Hinzu kommt der (weniger stark ausgeprägte) Kontakt mit Bantu- und Khoisansprachen wie Otjiherero oder Khoekhoegowab. Neben diesem multilingualen Setting ist auch die sprachliche Heterogenität der deutschsprachigen Immigrant/-innen kennzeichnend für das Deutsche in Namibia: Da Deutschsprachige aus unterschiedlichen deutschen Dialektgebieten im Südwesten Afrikas aufeinandertrafen und miteinander interagierten, handelt es sich beim heutigen namibischen Deutsch auch um das Resultat von Varietäten- bzw. Dialektkontakt (Zimmer inger.).

<sup>1</sup> Unsere Arbeit wird gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – WI 2155/9-1; SI 750/4-1. Weitere Projektbeteiligte waren Hans C. Boas, Janosch Leugner, Laura Perlitz und Anika Kroll-Tjingaete. Wir danken den zahlreichen Gewährspersonen für die große Kooperationsbereitschaft, ihre Gastfreundschaft und das Interesse an unserem Forschungsprojekt. Dank gebührt weiterhin Jones Anam, Christian Anders, Claudia Czarniak, Jan Gorisch, Philipp Klaußner, Semra Kizilkaya, Jula Kostka, Carina Schüffler und Ray Tjingaete, die bei der Erstellung des Korpus mitgewirkt haben.

Die deutschsprachige Community in Namibia ist damit eine vitale Sprachgemeinschaft, deren Sprachgebrauch maßgeblich durch Sprach- und Varietätenkontakt geprägt ist. Neben grammatischen Spezifika hat diese Konstellation auch ein interessantes Geflecht an einstellungsbezogenen Besonderheiten befördert: Einerseits dient das Deutschlanddeutsche gerade in Bildungskontexten als Prestigevarietät, und nicht-standardsprachliche Strukturen werden generell oft negativ bewertet; andererseits fungieren gerade die namibia-spezifischen Formen als positiv besetzte gruppenidentitätsstiftende Merkmale der deutschsprachigen Namibier/-innen (Wiese & Bracke i. Ersch.).

All dies macht das namibische Deutsche zu einem interessanten Forschungsgegenstand, unter anderem für die Bereiche Kontaktlinguistik, Variationslinguistik, Soziolinguistik und Jugendsprachforschung. Mit dem Korpus *Deutsch in Namibia (DNam)* liegt nun erstmals eine Ressource vor, die den Sprachgebrauch und die Spracheinstellungen der deutschsprachigen Namibier/-innen umfassend und systematisch dokumentiert. Über die *Datenbank für Gesprochenes Deutsch (DGD)* des *Instituts für Deutsche Sprache (IDS)* ist diese neue Ressource frei zugänglich. Das Korpus erlaubt grammatische, pragmatische und lexikalische Analysen des Deutschen in Namibia einschließlich informelleren Sprachgebrauchs sowie die Untersuchung einstellungsbezogener Merkmale innerhalb der Sprechergemeinschaft. Mit einer Gesamtgröße von mehr als 200.000 Tokens erlaubt es das Korpus, auch seltenere Phänomene datenbasiert und mithilfe authentischen Sprachmaterials zu untersuchen und liefert eine Basis für systematische quantitative Analysen.

Flankiert wird das *DNam-Korpus* durch ein weiteres Korpus, das Übersetzungen der klassischen Wenker-Sätze in umgangssprachliches (namibisches) Deutsch erfasst (Wiese 2014). Dieses Korpus enthält Daten von über 200 Teilnehmer/-innen unterschiedlicher Altersgruppen, die mithilfe eines Online-Fragebogens in den Jahren 2013 und 2014 gesammelt wurden. Das Korpus enthält die Wenker-Übersetzungen sowie biographische, soziale und soziolinguistische Daten zu den Teilnehmer/-innen, die mit Hilfe eines Personenfragebogens erfasst wurden. Es steht unter der CC-By 3.0-Lizenz über folgende URL frei zur Verfügung :

<<https://www.linguistik.hu-berlin.de/de/institut/professuren/multilinguale-kontexte/Projekte/Namdeutsch/Korpusdaten/NamDeutsch-Wenker>>.

Auch wegen der Vergleichbarkeit mit der großen Menge an Datensätzen, die mit dem identischen Erhebungsinstrument in anderen Regionen und zu anderen Zeitpunkten gesammelt wurden, handelt es sich hierbei um wertvolles Material für eine Reihe linguistischer Fragestellungen. Für eine Beispielstudie, die auf der Analyse dieser Daten basiert und auf Unterschiede beim Fremdwortgebrauch verschiedener Altersgruppen fokussiert, s. Zimmer (i. Ersch.).

In den weiteren Abschnitten skizzieren wir kurz das soziolinguistische Setting und die Geschichte des Deutschen in Namibia, deren Kenntnis für das Verständnis der Daten wichtig ist (Abschnitt 2).<sup>2</sup> Vor diesem Hintergrund erläutern wir dann das Design des *DNam-Korpus* und gehen dabei auf die Datenerhebung (Abschnitt 3.1), die Datenaufbereitung (Abschnitt 3.2), die Eigenschaften des Korpus (Abschnitt 3.3) und grundlegende Funktionsweisen der *DGD* (Abschnitt 3.4) ein. Die Nutzungsmöglichkeiten der Ressource werden dann anhand erster Studien illustriert (Abschnitt 4).

## 2 Deutsch in Namibia

Namibia ist ein Land, das durch ein hohes Maß an Mehrsprachigkeit gekennzeichnet ist. Als Familiensprachen sind Bantusprachen (vor allem Oshiwambo, aber auch Otjiherero) und Khoisansprachen (z.B. Khoekhoegowab) am weitesten verbreitet. Daneben spielen zwei germanische Sprachen eine wichtige Rolle: Afrikaans wird in vielen Bereichen als Lingua Franca verwendet, und Englisch ist alleinige Amtssprache (als „official language“; neben 13 sogenannten „national languages“, darunter neben einer Reihe autochthoner Sprachen – insbesondere Bantusprachen – auch Deutsch und Afrikaans).

Mitglieder der deutschsprachigen Minderheit sind in aller Regel mindestens dreisprachig und beherrschen neben Deutsch auch Afrikaans und Englisch. Hinzu kommen unterschiedlich stark ausgeprägte, meist jedoch sehr geringe Kenntnisse einzelner Bantu- und/oder Khoisansprachen. Die deutschsprachige Minderheit umfasst etwa 20.000 Sprecher/-innen und damit ungefähr 1% der ca. 2

<sup>2</sup> Für ausführlichere Ausführungen vgl. Pütz (1991, 1995), Gretschel (1995), Schmidt-Lauber (1998), Böhm (2003), Deumert (2009, 2018), Ammon (2014), Dück (2018), Kroll-Tjingaete (2018), Shah & Zappen-Thomson (2018), Stolz & Warnke (2018), Zappen-Thomson (2019) und Zimmer (2019).

Millionen Namibier/-innen. Die Immigration der Deutschsprachigen fand im Wesentlichen im Kontext der Kolonialisierung des heutigen Namibia statt (*Deutsch-Südwestafrika*; 1884 – 1915). Aber auch daran anschließend gab es und gibt es bis heute Zuwanderung aus dem deutschsprachigen Raum in Europa.

Die deutsche Kolonial-Herrschaft endete mit dem Ersten Weltkrieg. Englisch und Afrikaans wurden anschließend als Amtssprachen eingeführt und lösten Deutsch ab. Die Verwaltung Südwestafrikas ging an Südafrika über, das das Gebiet zuvor besetzt hatte, und die 1948 in Südafrika offiziell eingeführte Apartheidspolitik wurde auch auf Südwestafrika übertragen. 1990 erlangte Namibia seine Unabhängigkeit von Südafrika, wodurch auch die Apartheid beendet wurde. Im Zuge dessen wurde Englisch der Status der alleinigen Amtssprache zugesprochen.

Sowohl die Apartheid als auch die Kolonialzeit haben das Land nachhaltig geprägt. Noch heute sind deren Auswirkungen unübersehbar. Auch bei der Wahl der Amtssprache spielte die jüngere Geschichte des Landes eine wichtige Rolle. Die früheren Amtssprachen Afrikaans und Deutsch schieden wegen ihrer Konnotation mit der Kolonial- bzw. Apartheidzeit aus. Weder eine der zahlreichen Bantu- noch eine der Khoisansprachen wurde gewählt, um keine der entsprechenden Gruppen zu bevorteilen. Stattdessen fiel die Wahl auf Englisch, das als neutrale Sprache wahrgenommen wurde. Deutsch ist der Amtssprache untergeordnet, erfährt als eine von 13 Nationalsprachen aber eine gewisse institutionelle Unterstützung. Dies äußert sich zum Beispiel darin, dass das Fach Deutsch als Muttersprache (DaM) nicht nur an Privat-, sondern auch an staatlichen Schulen belegt werden kann (für Details s. Zappen-Thomson 2019).

Generell ist es den Mitgliedern der deutschen Sprechergemeinschaft in Namibia ein wichtiges Anliegen, dass Schüler/-innen das Fach DaM belegen können. Da viele Deutschsprachige auf geographisch abgelegenen Farmen leben und nicht flächendeckend DaM angeboten wird, wohnen viele Schüler/-innen in Internaten, die an Schulen mit DaM-Unterricht angebunden sind (sogenannte *Schülerheime*). Hier wird auch in der Freizeit vorwiegend Deutsch gesprochen.

Darüber hinaus engagieren sich viele Deutschsprachige für den Erhalt von Privatschulen mit DaM-Unterricht. Dieses Engagement wird dadurch erleichtert, dass die deutschsprachige Minderheit in Namibia zu den sozioökonomisch privilegierten Gruppen gehört. Insgesamt ist

die Community gut vernetzt und vergleichsweise aktiv. So werden zahlreiche Veranstaltungen organisiert, die auch dem Erhalt der deutschen Sprache in Namibia dienen sollen. Dazu gehören Karnevalssitzungen, Oktoberfeste, Sportveranstaltungen usw. Zu den Domänen, in denen (auch) Deutsch gesprochen wird, zählen außerdem die Kirchen: In vielen Teilen des Landes werden regelmäßig deutschsprachige Gottesdienste angeboten. Ferner gibt es eine deutschsprachige Tageszeitung (die *Allgemeine Zeitung*) und deutschsprachige Radiosender (das *Hitradio Namibia* sowie ein deutschsprachiges Angebot der *Namibian Broadcasting Corporation*). Ein Teil der Community hat darüber hinaus via Satellitenübertragung Zugang zu deutschen Fernsehsendern.

### 3 Korpus-Design

Das *DNam-Korpus* wurde im Rahmen eines von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts („Namdeutsch: Die Dynamik des Deutschen im mehrsprachigen Kontext Namibias“; Laufzeit: 2016 – 2020) erstellt, das zum einen zunächst an der Universität Potsdam, dann an der Humboldt-Universität zu Berlin (Projektleitung: Heike Wiese) und zum anderen an der Freien Universität Berlin (Projektleitung: Horst J. Simon) und in Kooperation mit der University of Namibia, Windhoek (Marianne Zappen-Thomson), durchgeführt wurde. Neben der Dokumentation hat sich das Projekt der Beschreibung und Analyse grammatischer, pragmatischer und einstellungsbezogener Spezifika des Deutschen in Namibia gewidmet. Die Interpretation der grammatischen Besonderheiten erfolgte aus den Perspektiven von Sprachwandel und Sprachvariation des Deutschen im mehrsprachigen Kontext und fokussierte auf Interferenzen aus den Kontaktsprachen, den Ausbau binnenstrukturell angelegter Tendenzen und Registerdifferenzierungen unter Bedingungen von Sprachkontakt. Die Erfassung, Beschreibung und Analyse der sprachideologischen und einstellungsbezogenen Eigenheiten erfolgte mit Blick auf den Konnex von Sprache und Identität im mehrsprachigen Kontext (Wiese et al. 2017). Einige auf den Korpusdaten basierende Ergebnisse werden in Abschnitt 4 veranschaulicht.

Mit dem im Projekttitel verwendeten Begriff *Namdeutsch* referieren wir auf namibisch geprägten Sprachgebrauch des Deutschen. Dieser

Begriff wird auch in der Community häufig mit dieser Bedeutung verwendet. Alternative Bezeichnungen sind *Nam-Släng*, *Namlish* und *Südwesterdeutsch*. Für die Verwendung von *Namdeutsch* haben wir uns entschieden, da es sich hierbei um eine relativ neutrale Bezeichnung handelt, die weder mit medial stilisiertem Sprachgebrauch assoziiert ist (anders als *Nam-Släng*) noch von vielen jüngeren Community-Mitgliedern wegen ihrer Konnotation mit der Kolonialzeit gemieden wird (im Gegensatz zu *Südwesterdeutsch*), und im Vergleich zu *Namlish*, das auch namibisch geprägtes Englisch bezeichnen kann, terminologisch präziser ist (vgl. hierzu auch Zimmer 2019).

Mit dem Korpus *Deutsch in Namibia* legen wir das Korpus möglichst breit an. Wie unten noch deutlich wird, enthält das Korpus z.B. auch Material, das in formellen Kommunikationssituationen erhoben wurde. Diese Daten sind z.T. sehr eng am Standarddeutschen, das in Deutschland gesprochen wird. Ob sich hierbei eine eigene namibische Variante des Standarddeutschen, also ein Standard-Namdeutsch entwickelt oder ob solche Daten außerhalb des eigentlichen „Namdeutschen“ liegen – wobei sie natürlich immer noch Teil des Deutschen in Namibia sind –, ist Gegenstand aktueller Forschung (vgl. z.B. Ammon, Bickel & Lenz 2016, Kellermeier-Rehbein 2016 und Wiese & Bracke i. Ersch.).

Das Ziel, den deutschen Sprachgebrauch in Namibia möglichst umfassend zu dokumentieren, spiegelt sich in der Wahl der Methoden für die Datenerhebung wider, die im folgenden Abschnitt dargestellt wird.

### **3.1 Datenerhebung**

Die Aufnahmen für das Korpus wurden im Juli/August und November 2017 in drei verschiedenen Set-ups durchgeführt: freie Gespräche, „Sprachsituationen“ und Interviews. Die Datenerhebungen fanden in Unterrichts- und Internatsgebäuden (erste Erhebungsphase) sowie auf den Farmen der Gewährspersonen, sonstigen Privatgebäuden und öffentlichen Räumen statt (zweite Erhebungsphase). In der ersten Erhebungsphase wurden Schüler/-innen aufgenommen, in der zweiten Erhebungsphase schwerpunktmäßig Erwachsene. Die Set-ups wurden in beiden Erhebungsphasen verwendet, allerdings wurde das Vorgehen minimal altersgerecht angepasst. Im Folgenden werden die verwendeten Set-ups näher erläutert:

1) Bei den **freien Gesprächen** unterhielten sich kleinere Gruppen von zwei bis fünf Personen in Abwesenheit der Forscher/-innen in informellem Setting über Themen klassischer soziolinguistischer Interviews (z.B. Kinderspiele) und/oder alltagsrelevante Aspekte des Lebens in Namibia. Für die Durchführung dieses Set-ups mit Schüler/-innen wurde im Vorfeld ein/e Sprecher/-in pro Gruppe als „Assistent/-in“ ausgewählt, der/die als Impulsgeber/-in während des Gruppengesprächs fungierte. Die Aufgabe dieser Person war es, selbst am Gespräch teilzunehmen und neue Themen einzubringen, falls die Unterhaltung ins Stocken geraten sollte. Dafür wurde dem/der Impulsgeber/-in zuvor eine Reihe von Themenvorschlägen an die Hand gegeben: EES (ein deutsch-namibianischer Popmusiker), der Musikgeschmack der eigenen Eltern, Erfahrungen mit Deutschland-Deutschen (namdeutsch: *Jerries*), Erkennungsmerkmale von Deutschland-Deutschen, Kinderspiele (von Namibier/-innen und Deutschland-Deutschen), Erfahrungen mit verschiedenen Aktivitäten (Jagen, Puppenspielen, Steinschleudern [namdeutsch: *Ketti*], Autofahren), gefährliche Situationen und schließlich Unterschiede zwischen Internatsbewohner/-innen (namdeutsch: *Heimer*) und Tagesschüler/-innen (*Städtern*).

Die Motivation für die Themen war es, geeignete Impulse für ein lebendiges Gespräch zu geben, bei dem die Teilnehmer/-innen möglichst natürlich agieren. Wie die so gewonnenen Daten zeigen, sind die Unterhaltungen darüber hinaus auch inhaltlich von soziolinguistischem Interesse.

Ein weiterer Gesprächsimpuls wurde generiert, indem sich die Gruppen einen Ausschnitt eines Musikvideos von EES anschauen konnten, in dem er *Nam-Släng* verwendet: die ersten 80 Sekunden des Videos zum Song *Sundowner*.<sup>3</sup> Die Impulsgeber/-innen wurden gebeten, dieses Video auf einem zur Verfügung gestellten Laptop relativ zu Beginn der Gesprächsrunde abzuspielen. Dies stimulierte nicht nur die Gespräche in den Runden, sondern trug auch zum intendierten informellen Setting bei.

<sup>3</sup> Das Video ist unter folgender URL verfügbar: <[https://www.youtube.com/watch?v=lyq\\_UORMP4Q](https://www.youtube.com/watch?v=lyq_UORMP4Q)> (zuletzt abgerufen am 05.09.2019).

Nach anfänglicher Zurückhaltung entwickelten sich in zahlreichen Gruppen auf diese Weise lebhafte Gespräche, wobei häufig die vorgeschlagenen Themenbereiche zugunsten von (anderen) alltagsrelevanten Gegenständen verlassen wurden. Dieser Teil des Korpus liefert somit freie und authentische Gespräche von befreundeten Schüler/-innen. Für die Transkription ausgewählt wurden die Gesprächsteile, die an das gemeinsame Betrachten des Videos anschließen, da sich die Teilnehmer/-innen zu diesem Zeitpunkt bereits an die Aufnahmesituation gewöhnt und ihre anfängliche Zurückhaltung abgelegt hatten.<sup>4</sup>

Bei der zweiten Erhebungsphase hat sich eine leichte Anpassung des Set-ups für die Datenerhebung mit den überwiegend erwachsenen Teilnehmer/-innen als gewinnbringend herausgestellt: Hier haben wir auf ein Video und Themenvorschläge verzichtet, sondern die Sprecher/-innen gebeten, aktuelle Themen zu besprechen, über die sie sich auch sonst unterhalten würden. In den aus Familienmitgliedern oder Freund/-innen bestehenden Gesprächsrunden wurden folglich von Anfang an selbstgewählte alltagsrelevante Themen besprochen. Auch hier entwickelte sich nach anfänglicher Zurückhaltung in aller Regel ein natürliches Gespräch.

2) Bei den „**Sprachsituationen**“ handelt es sich um naturalistische Daten, die durch gestellte Gespräche elizitiert wurden, die sich auf einen Verkehrsunfall beziehen (zur Methode s. Wiese i. Ersch.). Der Verkehrsunfall wurde nonverbal in Form einer Foto-Geschichte präsentiert. [Abbildung 1](#) enthält sechs der zehn verwendeten Stimuli-Fotos.

<sup>4</sup> Weiteres zur Datenauswahl wird in Abschnitt 3.2 besprochen.



Abbildung 1: Stimuli-Fotos für das „Sprachsituationen“-Set-up (Auswahl)

Für diese Gespräche wurde ein formelles und in ein informelles Setting simuliert, die sich durch die Gesprächspartner/-innen unterschieden. Im formellen Setting sollten sich die Sprecher\*innen vorstellen, sie sprächen mit einer Deutsch-Lehrkraft, die vom Aufnahmeleiter / von der Aufnahmeleiterin gespielt wurde. Für das informelle Setting wählten sie eine nahestehende Person unter den Anwesenden (Familienmitglieder oder Freund/-innen), die dann die/den Gesprächspartner/-in spielte. Mithilfe dieses Set-ups wurden somit gezielt Gespräche in einem vorgegebenen Register eliziert, die möglichst nah an natürlichen Spontandaten sein sollten. Als zusätzliche Kontrolle diente eine Authentizitätsprüfung durch die Peer-group: Die Aufnahmen fanden immer in Kleingruppen statt, sodass nach einem simulierten Gespräch unbeteiligte Sprecher/-innen nach der Repräsentativität des vorangegangenen Gesprächs befragt werden konnten. Falls dabei zur Sprache kam, dass ein Gespräch unnatürlich

verlaufen sei oder z.B. eine Sprecherin, „sehr steif“ oder „übertrieben im Slang“ gesprochen habe, wurden die Aufnahmen wiederholt, bis ein Resultat erzielt wurde, das von der Gruppe akzeptiert wurde.

Die gleiche Methode wurde und wird (mit entsprechend angepassten Stimuli) auch in anderen Sprachkontaktsettings angewendet, sodass eine große und weiterwachsende Bandbreite an vergleichbaren Daten aus unterschiedlichen Regionen verfügbar ist (vgl. Wiese i. Ersch.). Unter anderem wurde das Set-up mit Schüler/-innen in Berlin-Kreuzberg erfolgreich genutzt. Für diese Sprecher/-innengruppe konnte gezeigt werden, dass die mithilfe der „Sprachsituationen“ gesammelten elizitierten informellen Daten qualitativ solchen entsprechen, die aus Spontansprache stammen (vgl. Wiese & Pohle 2016) – gleichzeitig liefern die „Sprachsituationen“ die Möglichkeit, systematisch und kontrolliert vergleichbare Daten zu erfassen, die Registerdifferenzierungen abbilden.

3) Schließlich wurden mit jeweils zwei oder drei Proband/-innen semi-strukturierte **Interviews** geführt. Die von den Forscher/-innen gestellten Impulsfragen bezogen sich dabei auf Spracheinstellungen, Sprachgebrauch, Meinungen zur Sprachpolitik, Sprachbiographisches, Sprachpflege, perzeptionsdiagnostische Aspekte usw. Für die Transkription wurden sieben Interviews aus der zweiten Erhebungsphase ausgewählt.

Die Aufnahmen der ersten Erhebungsphase wurden an allen namibischen Sekundarschulen mit DaM-Unterricht sowie einer Einrichtung für privaten Deutschunterricht und mit der Unterstützung der dortigen Deutsch-Lehrkräfte gemacht. Im Rahmen der ersten Datenerhebungsphase hatten die Teilnehmer/-innen die Möglichkeit, uns freiwillig eine Telefonnummer oder eine Mailadresse für künftige Kontaktaufnahmen zu geben. Über die Schüler/-innen, die uns diese Daten überlassen hatten, waren wir anschließend in der Lage, auch Eltern als Gewährspersonen für die zweite Erhebungsphase zu gewinnen. Mit dem „friend of a friend approach“ (Franke 2008, S. 109–111) konnten wir daraufhin von den ersten erwachsenen Kontaktpersonen ausgehend ein größeres Netzwerk an Teilnehmer/-innen aufbauen.

Die Sekundarschulen mit DaM-Unterricht befinden sich in der Hauptstadt Windhoek (*Deutsche Höhere Privatschule [DHPS]* und

*Delta School Windhoek*), in der Küstenstadt Swakopmund (*Private School Swakopmund* und *Namib High School*) und in der im Landesinneren gelegenen Stadt Otjiwarongo (*Otjiwarongo Secondary School*). Aufnahmen wurden außerdem in den kleineren Orten Otavi und Omaruru sowie auf Farmen in der Umgebung von Windhoek, Witvlei, Omaruru und Otjiwarongo gemacht. Auf diese Weise sind mit Khomas, Otjozondjupa, Erongo und Omaheke alle Regionen abgedeckt, in denen Deutschsprachige in Namibia schwerpunktmäßig leben (vgl. Zimmer 2019). [Abbildung 2](#) gibt einen Überblick über die Erhebungsorte. Um die Anonymität der Teilnehmer/-innen zu gewährleisten, werden die Farmen nicht konkreter lokalisiert.

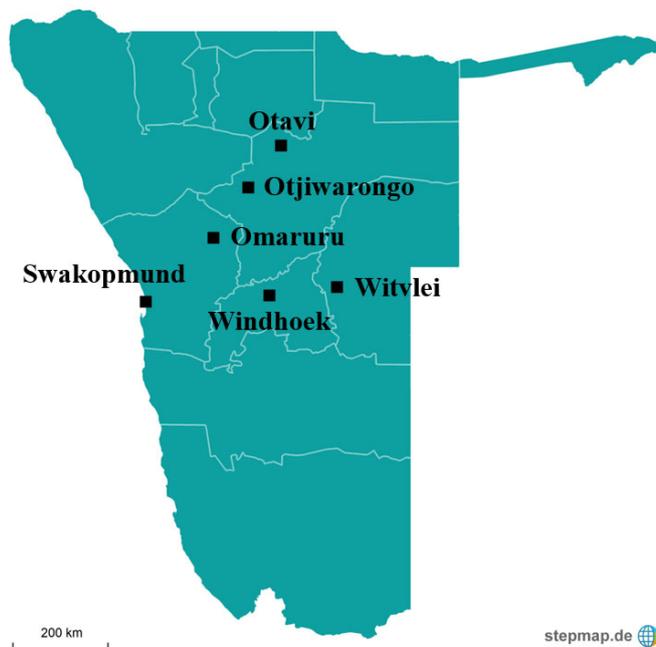


Abbildung 2: Erhebungsorte in Namibia

### 3.2 Datenaufbereitung und -erschließung

Die Gesamtmenge an Daten, die mit der im vorigen Abschnitt beschriebenen Vorgehensweise gesammelt werden konnten, überstieg deutlich die im Projekt vorgesehenen Kapazitäten für die Transkription.

Aus diesem Grund wurden kriteriengeleitet Aufnahmen für die Transkription ausgewählt, während andere Audios bislang noch nicht weiterbearbeitet wurden. Diese Audios werden in einer künftigen Korpus-Version ebenfalls zugänglich.

Bei der **Auswahl der zu transkribierenden Aufnahmen** wurde generell darauf geachtet, ein möglichst ausgewogenes Sample zu erstellen und dabei gleichermaßen Farmer/-innen wie Stadtbewohner/-innen, Schüler/-innen von Privat- und Staatsschulen und Sprecher/-innen aus unterschiedlichen Gebieten Namibias zu berücksichtigen sowie die drei Set-ups möglichst ähnlich zu gewichten. Außerdem wurde im Zweifelsfall den Aufnahmen mit Sprecher/-innen, die in Namibia geboren wurden, Vorrang gegeben. Aufgrund des fortwährenden Zuzugs aus Europa sind allerdings auch Immigrant/-innen der ersten Generation Teil der Community und haben ihren Anteil am zu dokumentierenden Sprachgebrauch, weshalb wir diese Sprecher/-innen nicht kategorisch von der Transkription ausgeschlossen haben. Darüber hinaus ermöglicht dies aufschlussreiche Vergleiche der verschiedenen Gruppen. Insgesamt ist ein breites Spektrum an Bildungsniveaus, Berufen und Altersgruppen abgedeckt, wobei Gespräche von bzw. mit Schüler/-innen der Geburtsjahrgänge 1999 bis 2003 einen Schwerpunkt im Korpus darstellen.

Bei der Auswahl der freien Gespräche spielten darüber hinaus die Natürlichkeit und die Lebendigkeit der Gespräche eine Rolle. Als Indikatoren dafür wurden verschiedene Merkmale herangezogen: So wurden Gespräche tendenziell nicht berücksichtigt, wenn es lange und häufige Gesprächspausen oder viele metasprachliche Kommentare gab, nur einzelne Teilnehmer/-innen sich am Gespräch beteiligten oder die vorgeschlagenen Themen der Reihe nach abgehandelt wurden, ohne dass die Gruppe eigene Gesprächsthemen einbrachte.

Die **Transkription** der ausgewählten Aufnahmen erfolgte nach den *cGAT-minimal*-Richtlinien (Selting et al. 2009; Schmidt, Schütte & Winterscheid 2015) und mithilfe des EXMARaLDA-Partitur-Editors (Schmidt 2016). Folglich wurde nicht phonetisch transkribiert, sondern in literarischer Umschrift. Dieses System folgt in weiten Teilen der Standardorthografie, erlaubt es aber gleichzeitig, typische Phänomene gesprochener Sprache wie Elisionen, Kontraktionen, Wortabbrüche usw. abzubilden (Beispiele folgen weiter unten). Gesprächspausen werden mit Angabe der Dauer (in Sekunden; z.B. (0.6)) ebenso

dokumentiert wie Paraverbales (z.B. *((lacht))*) und Nonverbales (z.B. *((Klingeln))*).

Die ersten Versionen der Transkripte wurden von jeweils einem anderen Mitglied des Transkriptionsteams kontrolliert. Bei der Kontrolle wurden Änderungsvorschläge in Kommentaren festgehalten, anschließend von dem/der ursprünglichen Transkribent/-in überprüft, bei Bedarf besprochen und dann ggf. in das Transkript eingearbeitet. Nach Transkription und Kontrolle der Transkripte durch das Transkriptionsteam erfolgte eine weitere Kontrolle durch die deutschsprachige Namibierin Anika Kroll-Tjingaete (University of Namibia, Windhoek). Dieser Arbeitsschritt war vor allem mit Blick auf namibiaspezifische sowie afrikaanse und auch einige englische Tokens von großem Wert, da diese von den deutschlanddeutschen Transkribent/-in nicht immer richtig erkannt und entsprechend transkribiert werden konnten.

Im Rahmen der Transkription und der Kontrolle der Aufnahmen erfolgte auch die **Anonymisierung** der Transkripte. Um die Anonymität der Sprecher/-innen zu gewährleisten, wurden nicht nur Personennamen und spezifische Ortsangaben (z.B. Farmnamen) anonymisiert, sondern auch alle anderen Aussagen, die Rückschluss auf die Identität der Sprechenden Person erlauben. Dazu wurden die Audio-Dateien an den zu anonymisierenden Stellen akustisch maskiert. Für die Anonymisierung in den Transkripten wurden vier Arten von Siglen angelegt:

- 1) Siglen für die Sprecher/-innen im Korpus. Diese sind folgendermaßen aufgebaut: Der Anfang *NAM* kennzeichnet den Status als Sprecher/-in im Korpus. Anschließend folgt eine dreistellige Zahl zur Identifikation sowie ein *M* oder *W* für das Geschlecht des/der Sprecher/-in. Die abschließende Zahl zwischen 1 und 4 bezeichnet die Altersgruppe des/der Sprecher/-in: 1: 20 Jahre oder jünger; 2: 21-40 Jahre; 3: 41-60 Jahre, 4: über 60 Jahre. Dies ergibt eine vollständige Sigle wie z.B. *NAM001M1*.
- 2) Siglen für die Forscher/-innen, die Interviews geführt haben und/oder an den „Sprachsituationen“ beteiligt waren: *RES1* bis *RES4*. Siglen für die Forscher/-innen und alle anderen Sprecher/-innen im Korpus identifizieren die Sprecher/-innen nicht nur in

den Metadaten, sondern auch in den Transkripten, wenn die jeweilige Person von anderen Sprecher/-innen erwähnt oder mit Namen angesprochen wird.

- 3) Siglen für einzelne Tokens, die anonymisiert wurden (z.B. Ortsnamen und Namen von Personen, bei denen es sich nicht um Sprecher/-innen im Korpus handelt). Diese sind zusammengesetzt aus dem Anfangsbuchstaben des anonymisierten Ausdrucks und einer dreistelligen Zahl, z.B. *N001*.
- 4) Siglen für Äußerungen, die aus mehreren Tokens bestehen und anonymisiert wurden. Solche längeren Passagen mussten z.B. anonymisiert werden, wenn durch das Gesagte eine angesprochene oder erwähnte Person identifizierbar ist. Die entsprechenden Siglen bestehen aus *anonymisierte Äußerung* gefolgt von einer dreistelligen Zahl, z.B. *anonymisierte\_Äußerung001*.

Im Anschluss an die Transkription wurden verschiedene Annotationsschritte durchgeführt. Die Annotationen wurden *inline* umgesetzt, d.h. unmittelbar in den EXMARaLDA-Dateien in Form von weiteren Spuren für jede/-n Sprecher/-innen. Die Transkripte beinhalten dadurch für alle Sprecher/-innen neben der ursprünglichen Transkriptionsebene (*trans*), auch Informationen auf einer tokenisierten Ebene (*trans\_tok*), einer normalisierten Ebene (*norm*), Ebenen mit Wortart- und Lemma-Annotationen (*pos*, *lemma*) und einer Ebene mit Annotationen für Tokens, die aus einer Kontaktsprache stammen (*FW*); (vgl. [Abbildung 3](#)).

	123 [00:36.7]	124 [00:36.9]	125 [00:37.1]	126 [00:37.2]	127 [00:37.4]
<b>NAM162M2 [trans]</b>	müssn wir die			net	rausholn
<b>NAM162M2 [trans_tok]</b>	müssn	wir	die	net	rausholn
<b>NAM162M2 [norm]</b>	<b>müssen</b>	<b>wir</b>	<b>die</b>	<b>net</b>	<b>rausholen</b>
<b>NAM162M2 [pos]</b>	VMFIN	PPER	PDS	PTKMA	VVINF
<b>NAM162M2 [lemma]</b>	müssen	wir	die	net	rausholen
<b>NAM162M2 [FW]</b>				A-S0-I0-L0	

Abbildung 3: Transkriptions- und Annotationsspuren in einem EXMARaLDA-Transkript

Eine tokenisierte Ebene stellte die Grundlage für die weiteren Nachbearbeitungsschritte dar. **Tokenisierung** bedeutete im Kontext unserer Datenbearbeitung die Erstellung einer Spur in der EXMARaLDA-Transkriptdatei, auf der jedes Event genau ein Token enthält. Ein Token ist dabei definiert als Text, der auf der ursprünglichen Transkriptionsebene (trans) zwischen zwei Leerzeichen steht. Für diesen Arbeitsschritt wurde der Tokenisierer des Programms *Pepper* (Zipser & Romary 2010) verwendet.

Beim Bearbeitungsschritt **Normalisierung** fand eine Anpassung des transkribierten Textes statt. Auf der norm-Ebene entspricht die Schreibung der Tokens nicht cGAT, sondern standarddeutscher Orthographie (s. z.B. Zeile 1 vs. Zeile 3 in Abbildung 3). Dies erleichtert zum einen die Suche im Korpus. Zum anderen sind normalisierte Tokens notwendig, um bessere Resultate beim Taggen von Lemmata und Wortarten zu erhalten (s. unten). Wichtig zu erwähnen ist, dass die Normalisierung zwar eine orthographische Anpassung ans Standarddeutsche darstellt, jedoch keine Ergänzungen oder „Korrekturen“ von Gesagtem vorgenommen wurden. Das heißt insbesondere, dass syntaktische Nonstandard-Strukturen in Bezug auf Kasus, Numerus, Genus, Reihenfolge usw. nicht normalisiert wurden. Entscheidungen über Normalisierungen wurden auf Basis der (Online-Version der) Duden-Rechtschreibung<sup>5</sup> und der Richtlinien für die Normalisierung von cGAT-Transkripten für das *Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK)* (Winterscheid et al. 2019) getroffen. Sofern ein Dudeneintrag vorlag, wurden auch Wörter mit den

<sup>5</sup> <www.duden.de> (zuletzt abgerufen am 25.10.2019).

stilistischen Markierungen „umgangssprachlich“ und „salopp“ entsprechend normalisiert. Das gilt beispielsweise für Kontraktionen wie *son* oder *aufm*.

Die Normalisierung der Dateien wurde folgendermaßen umgesetzt: Auf einer Kopie der *trans\_tok*-Spur wurden die Tokens mithilfe eines Python-Skripts (falls notwendig) durch eine normalisierte Variante ersetzt. Dafür wurde ein Ersetzungslexikon aus cGAT-transkribierten Types und ihren normalisierten Entsprechungen verwendet. Dieses Lexikon basiert auf dem Normalisierungs-Lexikon für das *FOLK* und wurde durch korpus-spezifische Einträge ergänzt. Nach der automatischen Ersetzung wurde die Normalisierung überprüft und manuell korrigiert.

Im nächsten Bearbeitungsschritt wurde ein **POS- und Lemma-Tagging** der normalisierten Tokens vorgenommen. POS-Tags ermöglichen die Suche nach allen Types einer Wortart und damit auch nach abstrakteren syntaktischen Konstruktionen. Die Lemmatisierung vereinfacht die Suche, wenn alle Wortformen eines Lexems gefunden werden sollen. Das verwendete POS-Tagset ist *STTS 2.0* (Westpfahl 2014; Westpfahl et al. 2017), eine speziell für die Annotation von Gesprächsdaten abgewandelte Version des *Stuttgart Tübingen Tagset* (Schiller et al. 1999). Ergänzt wurde dieses Set durch drei korpus-spezifische Tags: *SOART* für die im Duden geführte Kontraktion *son* und ihre Flexionsformen,<sup>6</sup> *ATM* für hörbares Atmen sowie *META* für in Doppelklammern auf der Transkriptionsspur notierte paraverbale Äußerungen wie *((lacht))*.

POS- und Lemma-Annotationen wurden automatisch mit *EXMARaLDA (Dulko)*<sup>7</sup> erstellt, einer Version des Partitur-Editors mit integriertem *TreeTagger* (Schmid 1995). Dabei haben wir auf eine Parameterdatei zurückgegriffen, die auf einem Goldstandard aus dem Korpus *FOLK* trainiert wurde (Westphal & Schmidt 2016). Da viele kontaktsprachliche Types in unseren Daten vorhanden sind, die in den Trainingsdaten fehlen dürften, wurde die Annotation in einem weiteren Schritt manuell optimiert. Dafür haben wir Frequenzlisten von Types

<sup>6</sup> Das Label wurde in Anlehnung an APPRART für eine Präposition mit inkorporiertem Artikel wie *am* und *zur* gewählt (Schiller et al. 1999: 67).

<sup>7</sup> S. <https://bitbucket.org/nolda/exmaralda-dulko/src/default/> (letzter Zugriff am 23. September 2019).

und deren Annotation generiert. Wenn ein Type einen POS-Tag mehr als fünf Mal erhalten hat, wurde geprüft, ob es sich dabei um einen Fehler handelte und ggf. per automatischer Ersetzung ein anderer Tag vergeben.<sup>8</sup>

Weiterhin bietet das Korpus auf der Ebene FM eine **Annotation von kontaktsprachlichen Tokens**.<sup>9</sup> In den Daten findet sich, wie bereits angedeutet, ein hoher Anteil von sprachlichem Material aus den lokalen Kontaktsprachen, insbesondere aus Englisch und Afrikaans. Um dessen Identifikation zu erleichtern, wurden entsprechende Tokens gesondert annotiert. Da keine Daten von vorneherein ausgeschlossen werden sollten, wurde dabei eine sehr weite Definition von kontaktsprachlichem Token zugrunde gelegt. So wurden z.B. auch englische Types berücksichtigt, die als Teil des Standarddeutschen in Deutschland zählen können, wie *cool* oder *okay*. Eine Differenzierung unterschiedlicher Arten von kontaktsprachlichen Tokens ermöglicht die Annotation über die folgenden vier Merkmale, die für jedes kontaktsprachliche Token kodiert sind:

- 1) GEBERSPRACHE: Welcher Gebersprache entstammt das Token?
- 2) SEQUENZ ( $\pm$ ): Ist das Token Teil einer Folge mehrerer kontaktsprachlicher Tokens (aus derselben Gebersprache) oder nicht?
- 3) INTEGRATION ( $\pm$ ): Ist das Token morphologisch overt ins Deutsche integriert oder nicht?<sup>10</sup>

<sup>8</sup> Es gab dabei Fälle, in denen Types mehr als einen möglichen POS-Tag haben (z.B. *braaiën* als finite (VVFİN) oder infinite Form (VVİNƑ)). In diesen Fällen wurde nach der Auswertung von Stichproben eine Entscheidung für die häufigere Angabe getroffen.

<sup>9</sup> Diese Annotationsebene ist aus technischen Gründen in der aktuellen Version der *DGD* noch nicht verfügbar, wird aber in einem künftigen Release zugänglich gemacht.

<sup>10</sup> Als Fälle von morphologischer Integration ins Deutsche betrachten wir erstens kontaktsprachliche Tokens, die eine deutsche Flexion aufweisen, die in der Gebersprache an dieser Stelle nicht verwendet würde, und zweitens Komposita, die aus deutschen und kontaktsprachlichen Gliedern bestehen (z.B. *Babystimme*). Alle Tokens, die zu nicht-flektierbaren Wortklassen gehören, sind per Definition nicht overt integriert.

4) LEXIKONEINTRAG ( $\pm$ ): Gibt es für das Token einen Eintrag im Duden oder nicht?

Annotations-Tags sind entsprechend aus vier Teilen aufgebaut wie z.B. *A-S0-I0-L0* in [Abbildung 3](#). Die erste Komponente ist ein Buchstabe (*E*, *A*, *O*, *U* oder *M*) für die Gebersprache des Tokens. Dabei steht *E* für Englisch, *A* für Afrikaans, *O* für *other* (= andere namibische Kontaktsprachen), *U* für *unknown* (= das Token ist höchstwahrscheinlich kontaktsprachlicher Herkunft, die Gebersprache jedoch unbekannt) und *M* für *multiple* (bei Komposita aus mehreren Gebersprachen). Die weiteren Komponenten des Tags bestehen jeweils aus einem Buchstaben gefolgt von *0* oder *1*. Der Buchstabe ist ein Kürzel für jeweils ein Merkmal: *S* für Sequenz, *I* für Integration, *L* für Lexikoneintrag. Die folgende Ziffer gibt an, ob das Merkmal ausgeprägt ist oder nicht: *1* steht für *ja*, *0* für *nein*. Das Beispiel aus [Abbildung 3](#) zeigt also, dass es sich bei *net* um eine Übernahme aus dem Afrikaans handelt (*A*), die hier nicht innerhalb einer Folge mehrerer afrikaanser Tokens auftritt (*S0*), keine overte Integration ins Deutsche aufweist (*I0*) und nicht im Duden steht (*L0*).

Bei den kontaktsprachlichen Tokens ist zusätzlich vermerkt, wenn ein kontaktsprachliches Token im Zuge eines *constructed dialogue*, also einer Wiedergabe von (realer oder fiktiver) wörtlicher Rede, oder metasprachlich verwendet wird. In diesen Fällen beinhaltet das Tag ein Suffix: *-c* für *constructed dialogue* bzw. *-m* für Metasprachliches.

Die Einordnung von Sprachmischungsphänomenen in Kategorien wie Codeswitching, Borrowing etc. ist umstritten (vgl. Poplack 2018). Auf eine solche Einteilung wurde daher bei der Annotation der kontaktsprachlichen Tokens im Korpus zugunsten der genannten theorieneutralen Merkmale verzichtet. Nichtsdestoweniger stellen die Merkmale mögliche Kriterien für eine solche Einteilung dar. So ließen sich bspw. alle Tokens, die Teil einer Sequenz sind, in der keine morphologische Integration zu beobachten ist, als Instanzen von Codeswitching auffassen. Ebenso kann ein Dudeneintrag als Hinweis auf ein etabliertes Lehnwort betrachtet werden.

### 3.3 Eigenschaften des Korpus

Das Korpus umfasst über 18 Stunden transkribiertes Audio-Material und hat einen Umfang von 224.392 Tokens.<sup>11</sup> [Tabelle 1](#) gibt einen Überblick über die Verteilung auf die drei verwendeten Set-ups.

Methode	Tokens	Laufzeit	Sprecher/-innen <sup>12</sup>	Aufnahmen
freie Gespräche	115.004	9:15:00 Stunden	65	21
„Sprachsituationen“	51.509	4:41:30 Stunden	103	198
Interviews	57.879	4:42:15 Stunden	15	7
gesamt	224.392	18:38:45 Stunden	110	226

Tabelle 1: Zusammenstellung des Korpus

Enthalten sind Aufnahmen von 110 Sprecher/-innen im Alter von 14 bis 75 Jahren (58 Sprecher, 52 Sprecherinnen). Das Material verteilt sich auf 226 Aufnahmen. Da mit den Teilnehmer/-innen in der Regel Aufnahmen in allen drei Set-ups durchgeführt wurden, konnte pro Sprecher/-in oft mehr als ein Datentyp transkribiert werden, was intra-individuelle Vergleiche zwischen den verschiedenen Set-ups ermöglicht.<sup>13</sup> Die „Sprachsituationen“ liegen systematisch paarweise vor (für 99 Sprecher/-innen: formelles und informelles Setting).<sup>14</sup>

Komplementiert werden die Korpusdaten durch ausführliche Metadaten. Diese können in der *DGD* angezeigt, exportiert und zum Filtern verwendet werden (s. unten). Neben Informationen zum Set-up

<sup>11</sup> Die hier genannte Anzahl an Tokens bezieht sich auf die Transkriptions-Ebene im Korpus. Annotationen und Normalisierungen wurden ebenso wie Klammern (zur Markierung nonverbaler Handlungen) nicht mitgezählt.

<sup>12</sup> In dieser Spalte werden nur die deutschsprachigen Namibier/-innen aufgeführt. Hinzu kommen die an der Datenerhebung beteiligten Wissenschaftler/-innen mit den Siglen (*RES1* bis *RES4*).

<sup>13</sup> Dementsprechend ergibt sich die Gesamtzahl der Sprecher/-innen in [Tabelle 1](#) nicht aus der Summe an Sprecher/-innen in den einzelnen Set-ups.

<sup>14</sup> Die Zahl von 103 Sprecher/-innen ergibt sich, weil z.T. weitere Personen im informellen Setting als antwortende Gesprächspartner/-in beteiligt waren (s. oben).

der jeweiligen Datenerhebung (z.B. *Interview*) sind vor allem Information zu den Sprecher/-innen verfügbar. Dazu zählen generelle persönliche Angaben wie Geburtsort, Geburtsjahr, Geschlecht, Wohnort, Beruf usw. Als Zusatzmaterial ist darüber hinaus eine ganze Reihe weiterer (sprach-)biographischer Merkmale, etwa eigene Sprachkenntnisse, Sprachkenntnisse der Eltern, Sprachgebrauch im Freundeskreis, Kontakthäufigkeit mit Deutschland-Deutschen sowie Details zur eigenen Immigration oder der Immigration der Vorfahren hinterlegt. Für Schüler/-innen, die an der ersten Erhebungsphase beteiligt waren, wurden darüber hinaus Medienkonsum und der Gebrauch des Deutschen in verschiedenen Domänen erfasst. Eine vollständige Liste aller als Metadaten verfügbaren Variablen und deren Merkmalausprägungen ist in der *DGD* abrufbar. Zentrale Hintergrundinformationen und die Funktionsweise dieser Datenbank werden im folgenden Abschnitt thematisiert.

### **3.4 Die Datenbank für Gesprochenes Deutsch**

Als Forschungsdatenzentrum für Korpora gesprochener Sprache hat das *Archiv für Gesprochenes Deutsch (AGD)* den Auftrag, Daten aus abgeschlossenen Forschungsprojekten zu übernehmen, dauerhaft zu archivieren und für eine Nachnutzung aufzubereiten und zugänglich zu machen. Im Falle des Namdeutsch-Projekts wurde ein neues Modell für solche Datenübernahmen erprobt: Indem das Projekt bereits ab der Konzeptionsphase und über die gesamte Laufzeit hinweg beratend begleitet wurde, konnte der Aufwand zur Aufbereitung nach Datenübernahme am Projektende deutlich reduziert werden. Das primäre Instrument zur Weitergabe und Nutzung aufbereiteter *AGD*-Daten ist die *Datenbank für Gesprochenes Deutsch (DGD* – <https://dgd.ids-mannheim.de>), deren Grundfunktionalität im Folgenden kurz skizziert wird. Eine ausführlichere Darstellung der *DGD* findet sich beispielsweise in Schmidt (2014), praktische Hilfestellungen zur Verwendung (Video-Tutorials, Handreichungen) im Hilfe-Menü der Anwendung selbst.

Wegen der datenschutzrechtlich sensiblen Natur vieler Daten in der *DGD* (so auch beim *DNam*-Korpus) ist deren Nutzung auf Zwecke der akademischen Forschung und Lehre beschränkt, was eine einmalige kostenlose persönliche Registrierung notwendig macht. Interessierte Nutzer/-innen sind aufgefordert, hier Kontaktdaten zu nennen und zu

spezifizieren, für welchen Zweck der Zugang zur *DGD* benötigt wird. Nach entsprechender positiver Prüfung der Angaben erhalten sie Zugangsdaten zum Login.

Die Plattform bietet Funktionalität erstens zum Browsen (Anhören, Ansehen) der Korpusdaten, zweitens zur systematischen Recherche (auf transkribiertem und annotiertem Material, auf Metadaten), drittens zum Download von Ausschnitten oder ausgewählten vollständigen Datensätzen. Dabei sind alle zu einem Korpus gehörigen Daten in vielfältiger Weise miteinander verknüpft, d.h. das Alignment zwischen Transkription und Aufnahme, die Zuordnung von Metadaten zu Gesprächsaufnahmen und Sprecher/-innen sowie die verschiedenen Annotationsebenen des Korpus sind an jeder Stelle der Plattform für Exploration und Analyse der Korpusdaten zugreifbar. Beispielsweise wird die Verknüpfung genutzt, um beim Lesen eines Transkripts in der Browsing-Ansicht den zugehörigen Ausschnitt der Audio-Aufnahme aufzurufen, oder um zu einem Recherchetreffer Informationen zum betreffenden Sprecher bzw. zur betreffenden Sprecherin anzuzeigen.

Für vollständig transkribierte und reichhaltig annotierte Daten, wie sie im *DNam*-Korpus vorliegen, ist vor allem die sog. Token-Recherche (im *DGD*-Menü unter: Recherche > Tokens) ein mächtiges und flexibles Instrument zum Arbeiten mit den Korpusdaten. Die Tokenrecherche erlaubt zunächst gezielte Abfragen auf dem Korpus nach einzelnen Wortformen. Dabei können Informationen auf verschiedenen Annotationsebenen miteinander verknüpft und reguläre Ausdrücke zur Spezifizierung von Zeichenmustern werden. Das Ergebnis einer solchen Recherche wird mit den genannten Verknüpfungen als Keyword in Context (KWIC)-Konkordanz dargestellt (vgl. [Abbildung 4](#)).

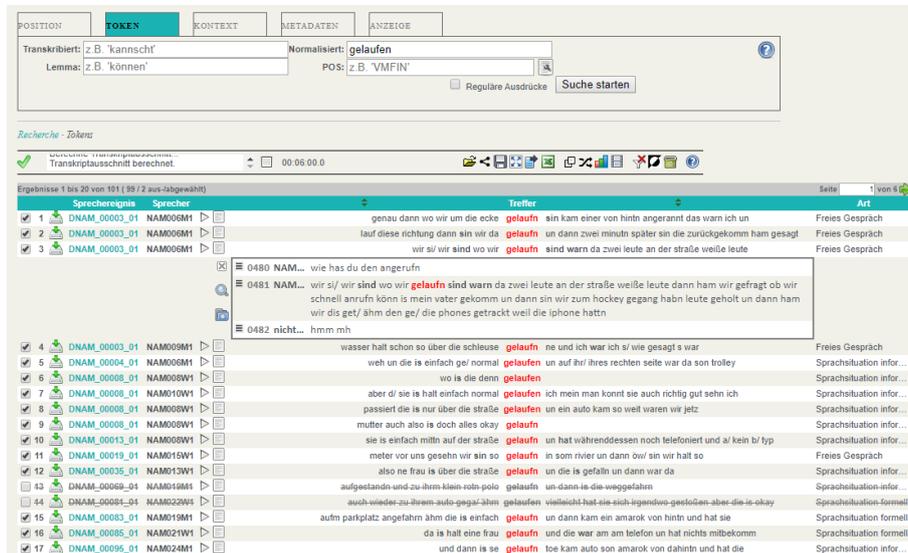


Abbildung 4: Tokenrecherche in der DGD, hier am Beispiel einer Suche nach *gelaufen*

Ausgehend von dieser initialen KWIC-Konkordanz bietet die DGD vielfältige Möglichkeiten, mit Suchergebnissen zu interagieren, sie zu erweitern, zu verfeinern oder weiter zu verarbeiten. Dazu gehören:

- die Möglichkeit, einzelne Treffer im größeren Transkriptkontext anzuzeigen und dabei auch auf das zugehörige Audio zuzugreifen (Zeile 3),
- die Möglichkeit, einzelne Suchergebnisse manuell an- oder abzuwählen, beispielsweise, um falsche Positive auszusortieren (Zeilen 13 und 14),
- Kontextfilter, über die das Suchergebnis nach Vorkommen von spezifischen Tokens im linken oder rechten Kontext weiter eingeschränkt werden kann (fettgedruckte Tokens in der KWIC),
- Metadatenfilter, über die Treffer mit Metadaten zu Aufnahmen und/oder Sprecher/-innen korreliert werden können (zusätzliche Spalte „Art“, die sich auf das Set-up der Aufnahme bezieht),
- die Möglichkeit, den zu einem Treffer gehörenden Audio- und Transkriptausschnitt auf den lokalen Rechner herunterzuladen

und dort weiter zu verarbeiten (z.B. um akustische Messungen mit einem Tool wie *Praat* vorzunehmen).

Auf diese Weise lassen sich mithilfe der *DGD* verschiedene qualitativ oder quantitativ orientierte Fragestellungen an das *DNam*-Korpus computergestützt bearbeiten. Da die *DGD* auch korpusübergreifende Abfragen erlaubt, besteht zudem – bei gebotener methodischer Vorsicht, die insbesondere die unterschiedlichen Erschließungsformen der anderen *DGD*-Korpora gebührend berücksichtigt – die Möglichkeit kontrastierender Untersuchungen, z.B. mit dem *Forschungs- und Lehrkorpus Gesprochenes Deutsch*, das standardnahes gesprochenes Deutsch im deutschsprachigen Kerngebiet erfasst, oder mit Korpora wie *Australiendeutsch*, die andere extraterritoriale Varietäten des Deutschen dokumentieren.

#### 4 Erste Analyseergebnisse

In diesem Abschnitt illustrieren wir kurz die Nutzungsmöglichkeiten des *DNam*-Korpus anhand zweier Bereiche, nämlich der Registerdifferenzierung und der Kasus-Verwendung (für ausführliche Analysen hierzu s. Wiese & Bracke i. Ersch. und Zimmer eing.).

Um Aufschluss über die intra-individuelle Variation von Sprecher/-innen in Form von Registerdifferenzierung zu erhalten, die an die Formalität der jeweiligen Situation geknüpft ist, bieten sich die durch das Set-up „Sprachsituationen“ gewonnenen Daten in besonderer Weise an. Mithilfe der pro Sprecher/-in paarweise vorliegenden Aufnahmen können systematische Vergleiche angestellt werden, um registerspezifische und registerübergreifende Merkmale zu identifizieren, die dann noch weitere Vergleiche innerhalb dieser Daten (z.B. nach Geschlecht oder Altersgruppe von Sprecher/-innen) zulassen und auch mit Daten aus anderen Set-ups abgeglichen werden können. Auf diese Weise können z.B. Befunde zum Varietätengefüge des namibischen Deutschen interpretiert werden, die auch für die Frage nach einer spezifischen namibischen Ausprägung des Standarddeutschen relevant sind (s. oben).

Im Hinblick auf registerspezifischen Sprachgebrauch liefern die Korpusdaten interessante Hinweise auf den differenziellen Gebrauch lexikalischer und grammatischer namdeutscher Charakteristika und ihre Interaktion mit einstellungsbezogenen Befunden (vgl. Wiese & Bracke i. Ersch.). Für die informellen „Sprachsituationen“-Daten sind

insbesondere lexikalische Entlehnungen charakteristisch (vgl. (1) bis (3)), während die formellen Daten typische Gehobenheitsmarker enthalten (vgl. (4) und (5)):

- (1) da war **like** sone alte oma (NAM025M1)
- (2) ich denk ihr **phone** war gebrochn (NAM062W1)
- (3) un dann kommt sone **tannie** da an (NAM019M1)
- (4) alles was da drin war is rausgefalln und **somit** auch ihr telefon (NAM022W1)
- (5) da hat sich ein unfall ereignet wo eine **ältre dame** von ein amarok angefahren wurde (NAM025M1)

Die Häufigkeit lexikalischer Entlehnungen kann auf Grundlage der Korpusdaten als ein konstitutives Merkmal für informelle Register des Namdeutschen identifiziert werden. Zugleich gibt es aber interessante lexemspezifische Unterschiede: So sind zwar Lexeme wie *like* und *phone* mit informellen Daten assoziiert, aber *stampen* und *Trolley* haben einen vergleichsweise hohen Anteil von Vorkommnissen auch in den formellen Bedingungen (vgl. auch Kroll-Tjingaete 2018 zu Entlehnungen in Zeitungstexten). Nichtkanonische grammatische Charakteristika scheinen besonders dann mit dem informellen Register assoziiert, wenn sie auch lexikalischen Transfer involvieren (Wiese & Bracke i. Ersch.): So ist das nichtkanonische Muster *seer/weh kriegen*, das eine semantisch-konstruktionelle Lücke im Deutschen schließt, in der Variante *seer kriegen* mit lexikalischer Entlehnung charakteristisch für das informelle Register; die Variante *weh kriegen* mit nativdeutschen Lexemen findet sich dagegen ebenso in der formellen Bedingung. Weitere nichtkanonische grammatische Phänomene, die auch für formelle Register im Korpus belegt sind, sind das vermutlich kontaktbasierte Muster *spät sein/kommen* sowie zwei Entwicklungen, die zusätzlich durch binnenstrukturelle Dynamiken des Deutschen gestützt sein könnten, nämlich Verwendungen von *haben* anstelle von *sein* bei Perfektbildungen mit Bewegungsverben und Belege für *helfen* mit Akkusativ-Komplement (Wiese & Bracke i. Ersch.).

So ist Beispiel (6) den Aufnahmen aus einer formellen Sprachsituation entnommen, während Beispiel (7) einer informellen Situation entstammt.

(6) da war ne alte frau die **hat** in der straße **gelaufn** und auf einmal kam ein auto (NAM066M1)

(7) die **hat** da auf der straße **gelaufn** mit\_n handy in der hand (NAM064W1)

Dass sowohl lexikalische als auch grammatische Charakteristika im formellen Register auftreten, kann als möglicher Hinweis auf die Entwicklung einer namibischen Standardvarietät des Deutschen interpretiert werden.

Die Befunde zum Sprachgebrauch passen zu einstellungsbezogenen Mustern, die sich in den Korpusdaten aus soziolinguistischen Interviews zeigen. Hier finden sich Hinweise auf ein Spannungsfeld zwischen Standardsprachideologien und sprachpuristischen Einstellungen auf der einen Seite und die Abgrenzung gegenüber Deutschland im Sinne einer lokalen, „namibischen“ Identität auf der anderen Seite (vgl. hierzu etwa auch Schmidt-Lauber 1998; Kellermeier-Rehbein 2015). Während erstere einen Sprachgebrauch stützen, der nah an der deutschlanddeutschen Standardsprache ist, kann die lokale Identifizierung lexikalische Entlehnungen begünstigen. Die im Korpus deutlich werdenden Registerdifferenzierungen reduzieren diese Spannung durch eine deutliche Unterscheidung formeller und informeller Varianten, bei der insbesondere lexikalische Entlehnungen für informelle Register salient zu sein scheinen (Wiese & Bracke i. Ersch.).

Weitere Korpusstudien, die Erkenntnisse illustrieren, die sich aus dem *DNam*-Korpus gewinnen lassen, liegen für den Bereich der Kasus-Verwendung im Namdeutschen vor. Wie in so gut wie allen deutschen Kontakt-Varietäten ist auch im Namdeutschen Variation in diesem Bereich zu beobachten (zum Kasusgebrauch in anderen deutschen Kontakt-Varietäten s. z.B. Franke 2008, Boas 2009, Yager et al. 2015, Rosenberg 2016). Die Variation im Namdeutschen illustrieren die folgenden Beispiele:

(8) dann is sie umgefalln **mit ihr handy** (NAM097M1)

(9) die frau is **mit ihrm handy** gelaufn (NAM123M1)

Mithilfe des *DNam*-Korpus kann dieser Phänomenbereich nun empirisch fundiert analysiert werden (vgl. Zimmer inger.). So zeigt sich z.B.,

- a) dass deutschsprachige Namibier/-innen Kasus in aller Regel in Einklang mit dem deutschen Standarddeutschen verwenden und Strukturen wie in (8) eher eine Ausnahme darstellen,
- b) dass von der Variation vor allem Kontexte betroffen sind, in denen im (deutschlanddeutschen) Standarddeutschen ein Dativ verwendet wird, während Akkusativ-Kontexte im Standarddeutschen und im Namdeutschen nahezu identisch sind,
- c) dass Dative wesentlich stabiler via Personalpronomen kodiert werden als z.B. durch Artikel.

Vor allem die beiden letztgenannten Punkte treffen ebenso auch auf zahlreiche andere deutsche Kontaktvarietäten zu (vgl. z.B. Salmons 1994, Van Ness 1994, de Kadt 2001, Yager et al. 2015, Rosenberg 2018 usw.), was auf sprach- und varietätenübergreifend wirksame Prinzipien hinweist, die sich hier als relevanter herausstellen als z.B. die konkreten Kontaktsprachen des Deutschen.

Auch für solche Bereiche ermöglicht es das *DNam*-Korpus, auch systematisch soziolinguistische Variablen einzubeziehen. Beim Kasus-Gebrauch zeigt sich z.B. ein signifikanter Unterschied zwischen den verschiedenen Schulen, die die Sprecher/-innen besuch(t)en. Hier sticht vor allem die einzige deutsche Auslandsschule, die DHPS, heraus. Auch das Geschlecht der Sprecher/-innen ist bedeutsam ebenso wie die L1 der Elternteile, vor allem die L1 der Mutter (vgl. Zimmer inger.). Dies verdeutlicht, dass die durch das *DNam*-Korpus ermöglichte systematische Berücksichtigung von grammatischen und soziolinguistischen Variablen entscheidend zum Verständnis des deutschen Sprachgebrauchs in Namibia beitragen kann; insbesondere auch für quantitative Analysen kann die Ressource gewinnbringend genutzt werden.

## 5 Fazit

Wie wir im vorliegenden Beitrag gezeigt haben, liegt mit dem *DNam*-Korpus eine Ressource vor, die den Sprachgebrauch und die Spracheinstellungen innerhalb einer deutschsprachigen Minderheit dokumentiert, die sprachwissenschaftlich besonders interessant ist wegen ihrer Vitalität und dem ungebrochenen Ausbau des Deutschen in formelle und informelle Register sowie wegen ihres soziolinguistischen Status als Kennzeichen einer ethnischen Minderheit. Das Korpus erlaubt sowohl Untersuchungen, die sich spezifisch dieser Community widmen, als auch vergleichende Studien. So besteht nun die Möglichkeit, das Deutsche in Namibia bei datenbasierten und varietätenübergreifenden Untersuchungen einzubeziehen und beispielsweise mit Texas German (vgl. z.B. Boas 2009b) oder dem Kiezdeutschen (vgl. Wiese 2012) zu kontrastieren. Das Deutsche in Namibia kann dabei eine wertvolle Vergleichsfolie darstellen, deren sprachliche und sprachexterne Spezifika für kontrastive Herangehensweisen gezielt genutzt werden können (Wiese et al. 2014; zum vergleichenden Ansatz s. z.B. auch Rosenberg 2003). Nicht zuletzt deshalb hoffen wir, mit unserem Korpus eine nützliche Ressource erstellt zu haben, die für eine große Bandbreite an Forschungsfragen genutzt wird und zu einem vertieften Verständnis des Zusammenspiels von Sprachkontakt, -wandel und -variation sowie Spracheinstellungen und -ideologien beitragen kann.

## Literatur

- Ammon, Ulrich (2014): Die Stellung der deutschen Sprache in der Welt. Berlin: de Gruyter.
- Ammon, Ulrich/Bickel, Hans/Lenz, Alexandra N. (Hg.) (2016): Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz, Deutschland, Liechtenstein, Luxemburg, Ostbelgien und Südtirol sowie Rumänien, Namibia und Mennonitensiedlungen. 2. Aufl. Berlin/Boston: de Gruyter.
- Boas, Hans C. (2009a): Case loss in Texas German. In: Barðdal, Jóhanna/Chelliah, Shobhana L. (Hg.): The Role of Semantic, Pragmatic and Discourse Factors in the Development of Case. Amsterdam/ Philadelphia: Benjamins, S. 347–376.

- Boas, Hans C. (2009b): *The Life and Death of Texas German*. (Publications of the American Dialect Society 93). Durham, NC: Duke University Press.
- Böhm, Michael Anton (2003): *Deutsch in Afrika: die Stellung der deutschen Sprache in Afrika vor dem Hintergrund der bildungs- und sprachpolitischen Gegebenheiten sowie der deutschen auswärtigen Kulturpolitik*. (= Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 52). Frankfurt am Main: Lang.
- Deumert, Ana (2009): *Namibian Kiche Duits: the making (and decline) of a neo-African language*. In: *Journal of Germanic Linguistics* 21, S. 349–417.
- Deumert, Ana (2018): *Settler colonialism speaks: Early contact varieties in Namibia during German colonial rule*. *Language Ecology* 2, S. 91–111.
- Dück, Katharina (2018): *Namibia*. In: Plewnia, Albrecht/Riehl, Claudia M. (Hg.): *Handbuch der deutschen Sprachminderheiten in Übersee*. Tübingen: Narr/Francke/Attempto, S. 109–130.
- Duden-Rechtschreibung (2017): *Die deutsche Rechtschreibung*. 27. Aufl. (Duden 1). Berlin: Dudenverlag.
- Franke, Katharina (2008): *„We call it Springbok-German!“: language contact in the German communities in South Africa*. PhD Diss. Monash University.
- Gretschel, Hans-Volker (1995): *The status and use of the German language in independent Namibia: Can German survive the transition?* In: Martin Pütz (Hg.): *Discrimination through Language in Africa? Perspectives on the Namibian Experience*. Berlin & New York: de Gruyter, S. 299–313.
- de Kadt, Elizabeth (2001): *“You still speak German?”: Teenage language skills in a German-speaking community*. In: Tom Ammerlaan (Hg.): *Sociolinguistic and Psycholinguistic Perspectives on Maintenance and Loss of Minority Languages*. Münster et al.: Waxmann, S. 61–72.
- Kellermeier-Rehbein, Birte (2016): *Sprache in postkolonialen Kontexten. Varietäten der deutschen Sprache in Namibia*. In: Stolz, Thomas/Warnke, Ingo H./Schmidt-Brücken, Daniel (Hg.): *Sprache und Kolonialismus*. Berlin & Boston: de Gruyter, S. 213–234.

- Kroll-Tjingaete, Anika (2018): Strukturelle Veränderungen der deutschen Sprache in Namibia: eine textanalytische Untersuchung. (= Hochschulschriften 51). Berlin: trafo.
- Poplack, Shana (2018): Borrowing: Loanwords in the Speech Community and in the Grammar. Oxford University Press.
- Pütz, Martin (1991): 'Südwesterdeutsch' in Namibia: Sprachpolitik, Sprachplanung und Spracherhalt. In: Linguistische Berichte 136, S. 455–476.
- Pütz, Martin (1995): Official monolingualism in Africa. In: Martin Pütz (Hg.): Discrimination through Language in Africa?. (= Contributions to the Sociology of Language 69). Berlin & New York: Mouton de Gruyter, S. 155–173.
- Rosenberg, Peter (2003): Comparative speech island research: Some results from studies in Russia and Brazil. In: Keel, William/Mattheier, Klaus J. (Hg.): Deutsche Sprachinseln weltweit: Interne und externe Perspektiven – German Language Varieties Worldwide: Internal and External Perspectives, S. 199–238.
- Rosenberg, Peter (2016): Regularität und Irregularität in der Kasusmorphologie deutscher Sprachinselvarietäten (Russland, Brasilien). In: Bittner, Andreas/Köpcke, Klaus-Michael (Hg.): Regularität und Irregularität in Phonologie und Morphologie: Diachron, kontrastiv, typologisch. (= Lingua Historica Germanica 13). Berlin & Boston: de Gruyter.
- Rosenberg, Peter (2018): Überflutete Sprachinseln: Sprachvariation, Sprachwechsel und Sprachwandel in deutschen Sprachinseln in Russland und Brasilien. In: Lenz, Alexandra N./Plewnia, Albrecht (Hg.): Variation – Normen – Identitäten. Berlin & Boston: de Gruyter, S. 263–282.
- Salmons, Joseph C. (1994): Naturalness and morphological change in Texas German. In: Berend, Nina/Mattheier, Klaus J. (Hg.): Sprachinselforschung: Eine Gedenkschrift für Hugo Jedig. Frankfurt am Main: Lang, S. 59–73.
- Schiller, Anne/Teufel, Simone/Stöckert, Christine/Thielen, Christine (1999): Guidelines für das Tagging deutscher Textcorpora mit STTS. (Kleines und großes Tagset). Universität Stuttgart, Institut für maschinelle Sprachverarbeitung; Universität Tübingen, Seminar für Sprachwissenschaft.

- Schmid, Helmut (1995): Improvements in part-of-speech tagging with an application to German. Proceedings of the ACL SIGDAT-Workshop. Dublin.
- Schmidt, Thomas (2014): The Database for Spoken German – DGD2. In: Calzolari, Nicoletta/Choukri, Khalid/Declerck, Thierry/Loftsson, Hrafn/Maegaard, Bente/Mariani, Joseph/Moreno, Asunción/Odijk, Jan/Piperidis, Stelios (Hg.): Proceedings of the Ninth Conference on International Language Resources and Evaluation (LREC'14). Reykjavik: European Languages Resources Association (ELRA).
- Schmidt, Thomas (2016): EXMARaLDA Partitur-Editor. Manual. [http://www.exmaralda.org/pdf/Partitur-Editor\\_Manual.pdf](http://www.exmaralda.org/pdf/Partitur-Editor_Manual.pdf).
- Schmidt, Thomas/Schütte, Wilfried/Winterscheid, Jenny (2015): cGat. Konventionen für das computergestützte Transkribieren in Anlehnung an das Gesprächsanalytische Transkriptionssystem 2 (GAT2). [http://agd.ids-mannheim.de/download/cgat\\_handbuch\\_version\\_1\\_0.pdf](http://agd.ids-mannheim.de/download/cgat_handbuch_version_1_0.pdf).
- Schmidt-Lauber, Brigitta (1998): Die verkehrte Hautfarbe: Ethnizität deutscher Namibier als Alltagspraxis. (= Lebensformen 10). Berlin: Reimer.
- Selting, Margret/Auer, Peter/Bart-Weingarten, Dagmar/Bergmann, Jörg/Bergmann, Pia/Birkner, Karin/Couper-Kuhlen, Elizabeth u. a. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 10, S. 353–402.
- Shah, Sheena (2007): German in a contact situation: The case of Namibian German. eDUSA 2, S. 20–45.
- Shah, Sheena/Zappen-Thomson, Marianne (2018): German in Namibia. In Seals, Corinne A./Shah, Sheena (Hg.): Heritage Language Policies around the World. Abingdon & New York: Routledge, S. 128–147.
- Stolz, Thomas/Warnke, Ingo H. (2018): Auf dem Weg zu einer vergleichenden Kolonialtoponomastik. Der Fall Deutsch-Südwestafrika. In: Kellermeier-Rehbein, Birte/Schulz, Matthias/Stolberg, Doris (Hg.): Sprache und (Post)Kolonialismus. Linguistische und interdisziplinäre Aspekte. (= Koloniale und Postkoloniale Linguistik 11). Berlin & Boston: de Gruyter, S. 71–103.

- Van Ness, Silke (1994): Pennsylvania German. In: König, Ekkehard/Van der Auwera, Johan (Hg.): *The Germanic Languages*. London: Routledge, S. 420–438.
- Westpfahl, Swantje (2014): STTS 2.0? Improving the Tagset for the Part-of-Speech-Tagging of German Spoken Data. In: Levin, Lori/Stede, Manfred (Hg.): *LAW VIII – The 8th Linguistic Annotation Workshop*. Dublin: Association for Computational Linguistics and Dublin City University.
- Westpfahl, Swantje/Schmidt, Thomas (2016): FOLK-Gold – A GOLD standard for Part-of-Speech-Tagging of Spoken German. In: Calzolari, Nicoletta/Choukri, Khalid/Declerck, Thierry/Goggi, Sara/Grobelnik, Marko/Maegaard, Bente/Mariani, Joseph/Mazo, Hélène/Moreno, Asunción/Odijk, Jan/Piperidis, Stelios (Hg.): *Proceedings of the Tenth Conference on International Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia. Paris: European Language Resources Association (ELRA). S. 1493-1499.
- Westpfahl, Swantje/Schmidt, Thomas/Jonietz, Jasmin/Borlinghaus, Anton (2017): *STTS 2.0. Guidelines für die Annotation von POS-Tags für Transkripte gesprochener Sprache in Anlehnung an das Stuttgart Tübingen Tagset (STTS)*. Version 1.1, März 2017.
- Wiese, Heike (2012): *Kiezdeutsch. Ein neuer Dialekt entsteht*. München: Beck.
- Wiese, Heike (2014): *Wenker-Namdeutsch*. [www.linguistik.hu-berlin.de/de/institut/professuren/multilinguale-kontexte/Projekte/Namdeutsch/Korpusdaten/NamDeutsch-Wenker](http://www.linguistik.hu-berlin.de/de/institut/professuren/multilinguale-kontexte/Projekte/Namdeutsch/Korpusdaten/NamDeutsch-Wenker).
- Wiese, Heike (i. Ersch.): *Language Situations: A method for capturing variation within speakers' repertoires*. In Yoshiyuki Asahi (Hg.): *Methods in Dialectology XVI*. (Bamberg Studies in English Linguistics). Frankfurt am Main: Lang.
- Wiese, Heike/Bracke, Yannic (i. Ersch.): *Registerdifferenzierung im Namdeutschen: Informeller und formeller Sprachgebrauch in einer vitalen Sprechergemeinschaft*. In: Csaba Földes (Hg.): *Kontaktvarietäten des Deutschen im Ausland*. Tübingen: Narr.
- Wiese, Heike/Pohle, Maria (2016): „Ich geh Kino“ oder „... ins Kino“? *Zeitschrift für Sprachwissenschaft* 35, S. 171–216.

- 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. *Zeitschrift für Dialektologie und Linguistik* 81, S. 274–307.
- Wiese, Heike/Simon, Horst J./Zimmer, Christian/Schumann, Kathleen (2017): German in Namibia: A vital speech community and its multilingual dynamics. *Language & Linguistics in Melanesia*, S. 221–245.
- Winterscheid, Jenny/Deppermann, Arnulf/Schmidt, Thomas/Schütte, Wilfried/Schedl, Evi/Kaiser, Julia (2019): Normalisieren mit OrthoNormal. Konventionen und Bedienungshinweise für die orthografische Normalisierung von FOLKER-Transkripten. Mannheim: Leibniz-Institut für Deutsche Sprache. <https://doi.org/10.14618/ids-pub-9326>.
- Yager, Lisa/Hellmold, Nora/Joo, Hyoun-A/Putnam, Michael/Rossi, Eleonora/Stafford, Catherin/Salmons, Joseph C. (2015): New structural patterns in moribund grammar: case marking in heritage German. *Frontiers in Psychology* 6, S. 1716.
- Zappen-Thomson, Marianne (2019): Stützung des Spracherhalts bei deutschsprachigen Minderheiten: Namibia. In: Ammon, Ulrich/Schmidt, Gabriele (Hg.): *Förderung der deutschen Sprache weltweit*. Berlin & Boston: de Gruyter, S. 517–530.
- Zimmer, Christian (2019): Deutsch als Minderheitensprache in Afrika. In: Herrgen, Joachim/Schmidt, Jürgen Erich (Hg.): *Sprache und Raum – Deutsch. Ein internationales Handbuch der Sprachvariation*. (= Handbücher zur Sprach- und Kommunikationswissenschaft 30.4). Berlin: De Gruyter Mouton, S. 1176–1190.
- Zimmer, Christian (i. Ersch.): Linguistic variation and age of speakers in Namibian German: loan word usage in “Wenker sentences”. In: Hans C. Boas (Hg.): *German Abroad: Comparative Perspectives on language Contact*. Leiden: Brill
- Zimmer, Christian (einger.): *Kasus im Namdeutschen*.
- Zimmer, Christian (einger.): *Siedlungsgeschichte und Varietätenkontakt: Zur Entstehung des Namdeutschen*.
- Zipser, Florian/Romary. Laurent (2010): A model oriented approach to the mapping of annotation formats using standards. In: Calzolari Nicoletta/Choukri, Khalid/Maegaard, Bente/Mariani,

Joseph/Odijk, Jan/Piperdis, Stelios/Rosner, Mike/Tapias,  
Daniel (Hg.): Proceedings of the Workshop on Language  
Resource and Language Technology Standards (LREC 2010).  
Valletta: European Languages Resources Association (ELRA).