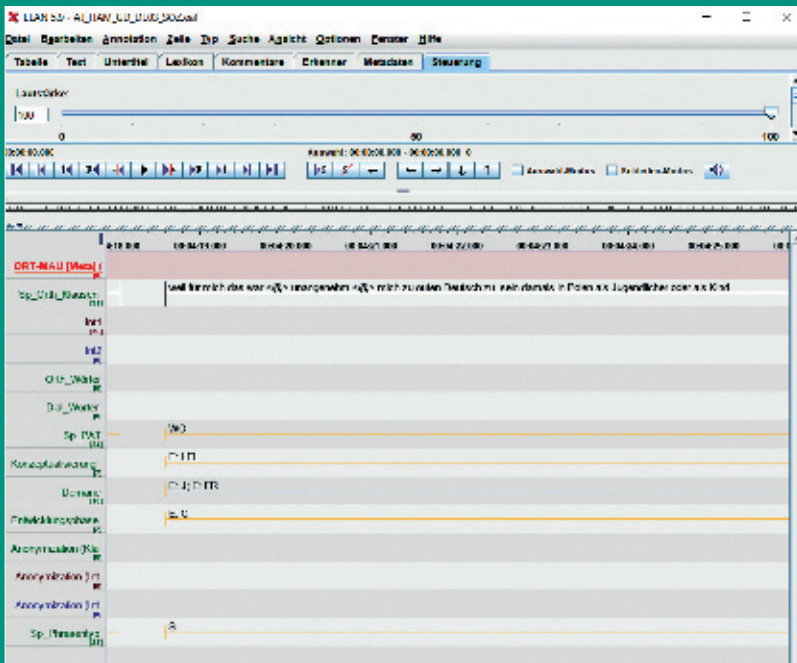


BJÖRN HANSEN
ANNA ZIELIŃSKA (Hg.)

Soziolinguistik trifft Korpuslinguistik

Deutsch-polnische und
deutsch-tschechische Zweisprachigkeit



Universitätsverlag
WINTER
Heidelberg



SLAVICA

Monographien, Hand-, Lehr- und Wörterbücher

Herausgegeben von

Marek Nekula

Klaus Trost

Band 9



Soziolinguistik trifft Korpuslinguistik

Deutsch-polnische und
deutsch-tschechische Zweisprachigkeit

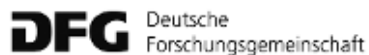
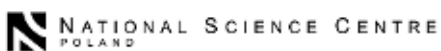
Herausgegeben von

BJÖRN HANSEN
ANNA ZIELIŃSKA

Universitätsverlag
WINTER
Heidelberg

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet
über <http://dnb.d-nb.de> abrufbar.

The paper was written within the framework of a research project entitled „Language across generations: contact induced change in morpho-syntax in German-Polish bilingual speech“ financed by the NCN, National Science Centre, Poland, project no. 2016/23/G/HS2/04369, carried out at the Institute of Slavic Studies, Polish Academy of Sciences and by the DFG, German Research Foundation, project no. HA 2659/9-1, carried out at the University of Regensburg.



Universitätsverlag Winter GmbH
Dossenheimer Landstraße 13
D-69121 Heidelberg
www.winter-verlag.de

TEXT: © 2022 Björn Hansen, Anna Zielińska

GESAMTHERSTELLUNG: Universitätsverlag Winter GmbH, Heidelberg, 2022

ISBN (Hardback): 978-3-8253-4894-6

ISBN (PDF): 978-3-8253-8591-0

DOI: <https://doi.org/10.33675/2022-82538591>



Dieses Werk ist lizenziert unter einer
Creative Commons Namensnennung – Nicht kommerziell – Keine Bearbeitungen
4.0 International Lizenz.
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Die Bedingungen der Creative-Commons-Lizenz gelten nur für Originalmaterial.
Die Wiederverwendung von Material aus anderen Quellen (gekennzeichnet mit
Quellenangabe) wie z. B. Schaubilder, Abbildungen, Fotos und Textauszüge erfordert
ggf. weitere Nutzungsgenehmigungen durch den jeweiligen Rechteinhaber.

Inhaltsverzeichnis

Vorwort	7
Iga Kościółek: LangGener als internationales Forschungsprojekt	11
<i>Konzeptuelle Überlegungen</i>	
Carolin Centner: Der Generationsbegriff	23
Barbara Jańczak – Marek Nekula – Anna Zielińska: Sprachbiographische Annotation im deutsch-polnischen Korpus	31
Aneta Bučková – Carolin Centner – Felicja Księżyk – Irena Prawdzic: Sprachstrukturelle Annotation der LangGener-Korpora: Typologie und Abgrenzungsprobleme	53
<i>Praktische Lösungen</i>	
Anna Jorroch – Irena Prawdzic: Praktische Durchführung der Feldforschung im deutsch-polnischen Team	93
Aneta Bučková – Irena Prawdzic: Transkriptionskonventionen	105
Aneta Bučková – Carolin Centner: Vom Gespräch zu annotierten Daten	115
Michał Woźniak: Von der annotierten Datei zum Korpus	135
<i>Dokumentation und Analysen</i>	
Anna Jorroch – Felicja Księżyk – Anna Zielińska: Dialektale Merkmale in den sprachbiographischen Interviews der älteren Generation in Polen	153

Björn Hansen – Marek Nekula: Die LangGener-Korpora als multifunktionale Ressourcen der Mehrsprachigkeits- forschung zwischen Sozio- und Korpuslinguistik	175
Aneta Bučková – Marek Nekula: Immigrantinnen und Immigranten aus der Tschechoslowakei in Deutschland: Musterentlehnungen in ihren sprachbiographischen Interviews	193
<i>Anhänge</i>	
Anhang 1: Interviewleitfaden für die sprachbiographischen Gespräche	247
Anhang 2: Einwilligungserklärung zur Erhebung und Verarbeitung personenbezogener Daten für Forschungszwecke	251
Anhang 3: Formularz zgody na przetwarzanie danych osobowych i rozpowszechnianie wizerunku	255
Anhang 4: Liste der Informantinnen und Informanten im LangGener-Korpus (deutsch-polnisches Subkorpus)	259
Anhang 5: Liste der Respondentinnen und Respondenten im deutsch-tschechischen Korpus Jazyky v migraci (Sprachen in Migration)	265

Vorwort

Das Projekt

Der vorliegende Sammelband enthält Beiträge, die im Rahmen des gemeinsamen Projekts *Language across generations: contact induced change in morphosyntax in German-Polish bilingual speech* (LangGener) der Institute für Slavistik an der Polnischen Akademie der Wissenschaften und der Universität Regensburg verfasst wurden.¹ Finanziell wurde es vom binationalen Förderprogramm Beethoven 2 unterstützt. Das Projekt und die Monographie schließen auch ein Teilprojekt zu deutsch-tschechischen Bilingualen ein, dessen Ergebnisse vergleichend einbezogen werden.

Im Rahmen des Projekts wurden Interviews mit zweisprachigen Personen aufgenommen, die unterschiedlichen Generationen und Gruppen zuzuordnen sind:

- 1) Die in den vor dem Zweiten Weltkrieg zu Deutschland gehörenden Gebieten des heutigen Polens geborenen und heute dort lebenden Personen bilden die ältere Generation;
- 2) Zur jüngeren Generation gehören die in den 50er und 60er Jahren des 20. Jahrhunderts ebenfalls dort Geborenen, die nach Deutschland ausgewandert sind;

Im deutsch-tschechischen Teilprojekt wurden zwei Gruppen interviewt:

- 1) Sudetendeutsche, die in der Tschechoslowakei um 1955 geboren und zwischen 1967 und 1989 als Spätaussiedlerinnen und -aussiedler nach Deutschland ausgewandert sind;
- 2) Tschechische Migrantinnen und Migranten, die in der Tschechoslowakei um 1955 geboren und zwischen 1967 und 1989 nach Deutschland ausgesiedelt wurden.

Die Untersuchungen erfassen also deutsch-polnische zweisprachige Personen aus zwei Generationen aus der gleichen Umgebung, aber mit unterschiedlichen Lebenserfahrungen. Im deutsch-tschechischen Teilprojekt handelt es sich um zwei Gruppen, bei denen man auf Grund unterschiedlicher Herkunft unterschiedliche Vorstellungen in Bezug auf die Sprachen Deutsch und Tschechisch voraussetzen kann, die man der Einfachheit halber auch Sprachideologien nennt. Die Sprachbiographien der zweisprachigen Personen wurden von den historischen Umständen entscheidend bestimmt, d. h. vom Zweiten Weltkrieg und von den grundlegenden gesellschaftspolitischen Veränderungen im

¹ Dieses Projekt wurde von der Deutschen Forschungsgemeinschaft unter der Projektnummer HA 2659/9-1 und vom Narodowe Centrum Nauki mit der Nummer 2016/23/G/HS2/04369 gefördert.

Nachkriegspolen bzw. in der Nachkriegstschechoslowakei sowie von der Aussiedlung aus dem Ostblock.

Das im Rahmen des Projekts erstellte Korpus dient nicht nur der Dokumentation und Präsentation der gesammelten Audioaufzeichnungen und ihrer Transkription, sondern vor allem soll es bei der Beantwortung der Frage helfen, ob die durch den Sprachkontakt hervorgerufenen morphosyntaktischen Veränderungen zweisprachiger Personen generationsspezifische oder – beim tschechischen Teilprojekt – sprachideologisch bedingte Unterschiede aufweisen. Untersucht wird daher der Zusammenhang zwischen den Sprachbiographien und der Verwendung der Sprache bei Vertretern zweier Generationen bzw. zweier Gruppen mit unterschiedlichen soziodemographischen Hintergründen.

Das Interesse richtet sich auf die zweisprachigen Sprecher sowie auf den Gebrauch der morphosyntaktischen Modelle im Polnischen und Deutschen bzw. im Tschechischen und Deutschen. Hauptanliegen des Bandes ist es, die Herausforderungen und Schwierigkeiten bei der Korrelation soziolinguistischer und korpuslinguistischer Fakten zu zeigen.

Die Annotation der morphosyntaktischen Replikationen werden im Band erläutert. Ein Novum für die Korpuslinguistik ist beim deutsch-polnischen Korpus die Annotation der soziolinguistischen Fakten, welche den sprachlichen Entwicklungsstand der Sprecher, ihre Sprachideologien, Sprachbeherrschung, Erfahrungen und Erlebnisse mit der Sprache reflektieren. Die Korrelation der sprachlichen mit den soziolinguistischen Fakten soll eine Antwort auf die oben genannte Untersuchungsfrage geben. Darauf zielt auch das tschechische Teilprojekt ab.

Die Leitfrage des gesamten Projektes ist also, ob es hinsichtlich von kontaktbedingten Musterreplikationen zum einen Unterschiede zwischen den beiden jeweils untersuchten Generationen und zum anderen Unterschiede zwischen den jeweiligen Sprachen gibt. Gibt es in der Sprache der Bilingualen die gleichen Typen von Musterreplikationen im Deutschen und im Polnischen bzw. Tschechischen? Damit hängt auch die Frage der Direktionalität der Replikationen (Deutsch \leftrightarrow Slavisch) zusammen. Ferner strebt das Projekt an, eine exhaustive Typologie von Musterreplikationen zu entwickeln.

Im Sammelband stellt ein internationales Wissenschaftlerteam von Germanistinnen und Germanisten und Slavistinnen und Slavisten die Herausforderungen vor, die mit der Erstellung des soziolinguistischen Korpus verbunden waren: die Entwicklung des Forschungskonzepts und die wissenschaftliche Diskussion darüber, die Ausarbeitung der Methoden zur Durchführung der Feldforschung und die Gewinnung soziolinguistischer Interviews nach der Klärung allgemeiner technischer Fragen.

Wir erhoffen uns, dass die vorliegende Arbeit zum einen als ein nützlicher und praktischer Wegweiser für die nachfolgenden internationalen Forschungsgruppen und zum anderen als eine detaillierte und informative Dokumentation der dreijährigen inhaltlich und organisatorisch komplexen wissenschaftlichen Kooperation und Teamarbeit fungieren wird.

Der Band bietet somit eine praktische Anleitung für die Erstellung soziolinguistischer Korpora. Daher hofft das Team, dass die Monographie das Interesse der Leserinnen und Leser weckt und Linguistinnen und Linguisten zur Erstellung weiterer soziolinguistischer Korpora anregt.

Struktur des Sammelbandes

Die vorliegende Monographie umfasst elf Kapitel und fünf Anhänge. Darin geben die Projektteilnehmenden einen Einblick in die verschiedenen Phasen der gemeinsamen Projektarbeit. Die im Band enthaltenen Artikel reflektieren zudem sehr gut den projektinternen Arbeitsstil der beiden Teams.

Der Sammelband spiegelt die Einzelschritte der Projektarbeit in inhaltlicher und teilweise auch chronologischer Sicht wider. Die Kapitel gliedern sich in die Themenbereiche konzeptuelle Überlegungen, praktische Lösungen sowie Dokumentation und Analysen.

Das erste Kapitel zeigt den Forschungskontext auf und beschreibt die Teamarbeit in unserem internationalen Projekt. Die anschließenden zwei Kapitel führen einige grundlegende konzeptuelle Überlegungen ein. Zunächst wird der für das LangGener-Projekt zentrale Begriff der Generation aus linguistischer Perspektive beleuchtet. Anschließend werden als Besonderheit unserer Korpora die soziolinguistische Annotation und die ebenfalls neue kontaktlinguistische Annotation vorgestellt. Die folgenden vier Kapitel dokumentieren den recht komplexen Weg von der Anbahnung des Gesprächs zum annotierten Korpus. So wird in Jorroch & Prawdzc im Detail die praktische Durchführung der Feldforschung im deutsch-polnischen Team vorgestellt. Das Kapitel Bučková & Prawdzc legt die Transkriptionskonventionen dar. Das folgende Kapitel Bučková & Centner zeigt den Weg vom Gespräch zur annotierten Datei und das Kapitel von Woźniak den weiteren Weg zum Korpus. Der letzte Abschnitt des Sammelbandes ist der Dokumentation und einer ersten empirischen Analyse gewidmet. Der Beitrag von Jorroch, Księżyk & Zielińska stellt die dialektalen Merkmale in den sprachbiographischen Interviews der älteren Generation in Polen vor. Das Kapitel Hansen & Nekula verortet die LangGener an der Schnittstelle von Korpus- und Soziolinguistik und zeigt zukünftige Anwendungsfelder der Korpora auf. Den Abschluss bildet der Beitrag „Immigrantinnen und Immigranten aus der Tschechoslowakei in Deutschland: Musterentlehnungen in ihren sprachbiographischen Interviews“. Im Anhang 4 befinden sich Tabellen, aus denen die Herkunft und der Wohnort der Respondentinnen und Respondenten zu ersehen sind.

Der Sammelband wendet sich an linguistisch interessierte Leserinnen und Leser und setzt im Prinzip keine Kenntnisse des Polnischen bzw. Tschechischen voraus. Bei der Notation des für uns so zentralen authentischen Datenmaterials haben wir uns um ein weitgehend einheitliches Format bemüht. Slavisches Material wird durchgängig übersetzt. Teilweise wurden zusätzlich die Leipzig Glossing Rules angewandt, wenn es der Autorin bzw. dem Autor explizit um sprachliche Strukturen geht.

Danksagungen

An einem solch großen Projekt sind sehr viele Personen beteiligt und viele Personen haben uns substantiell unterstützt. Wir sind sehr dankbar für die fachliche Unterstützung durch unsere Konsultantinnen und Konsultanten Prof. Dr. Sandra Birzer (Universität Bamberg), Prof. Dr. Bernhard Brehmer (Universität Konstanz) und Dr. Christoph Draxler (LMU München). Ebenfalls dankbar sind wir den zwei anonymen Gutachterinnen bzw. Gutachtern, die uns eine Vielzahl substantieller Anregungen gegeben haben. Ganz herz-

lich möchten wir uns bei Frau Annett Reisinger für die umfassende sprachliche Korrektur der vorliegenden Monographie bedanken. Großer Dank geht auch an den Reihenherausgeber Prof. Dr. Marek Nekula für seine professionelle Unterstützung bei der Edition des Buchs.

Ferner bedanken wir uns bei Dr. Marie Kopřivová für die Hilfe beim Erstellen der Transkriptionskonventionen sowie dem Team des Tschechischen Nationalkorpus (Dr. Marie Kopřivová, Michal Křen und David Lukeš), das die Integration des deutsch-tschechischen Korpus ins Tschechische Nationalkorpus möglich machte. Dr. Zuzana Finger danken wir für die Unterstützung bei der Suche von Probandinnen und Probanden aus der Gruppe der Spätaussiedlerinnen und Spätaussiedler. Ein besonderer Dank gilt allen, die uns bei der Suche nach Angehörigen der „Generation Polen“ behilflich waren: Frau Dr. Magdalena Lemańczyk und Frau Teresa Wiśniewska von der Sozial-Kulturellen Gesellschaft der Deutschen Minderheit Ortsgruppe Arnswalde, Herrn Piotr Nycz und Herrn Erhard Grünbauer von der Sozial-Kulturellen Gesellschaft der Deutschen Minderheit Ortsgruppe Stargard, Herrn Peter Jeske von der Sozial-Kulturellen Gesellschaft der Deutschen Minderheit „Pomerania“ in Köslin, Herrn Edwin Kemnitz und Frau Helga Hilda Kuczowska von der Deutschen Sozial-Kulturellen Gesellschaft in Schneidemühl, Herrn Klemens Mrela von der Deutschen Sozial-Kulturellen Gesellschaft Ortsgruppe Flatow, Frau Irena Szubzda und Frau Barbara Rużewicz vom Verband der Deutschen Gesellschaften in Ermland und Masuren, Herrn Roland Hau von der Danziger Deutschen Minderheit, Frau Róża Kańkowska von der Gesellschaft der Deutschen Minderheit Stadt und Kreis Elbing, Frau Danuta Krause von der Sozial-Kulturellen Gesellschaft der Deutschen Minderheit in Stettin, Frau Lidia Przepiórka vom Sozial-Kulturellen Bund der Deutschen Minderheit und Frau Gertruda Cicha, Prof. Dr. Patrick Zeller von der Universität Hamburg, Frau Aleksandra Jeszke-Zillmer von der Deutsch-Polnischen Gesellschaft Hamburg, der Gesellschaft der Deutschen Minderheit „Masuren“ in Lyck, dem Haus der Deutsch-Polnischen Zusammenarbeit und der Redaktion des „Wochenblatts“. Ohne diese Personen und Institutionen, die uns ihr Vertrauen geschenkt haben, hätten wir das Projekt nicht realisieren können.

An dieser Stelle danken wir für die geleistete Arbeit unseren Transkribentinnen und Transkribenten: Dr. Maksim Duškin, Dr. Gabriela Augustyniak-Žmuda, Maria Svojanovská, Vladimír Svojanovský, Iveta Patáková, Louisa Polowy, Michael Wastl, Katarzyna Olszewska, Nikola Koczuba, Katarzyna Laszuk, Anna Janusz, Szymon Pogwizd, Daria Šemberová, Agnieszka Paździor, Maria Walasek, Alicja Fiedler, Katarzyna Goszcz, Sandra Przywara, Natalia Prüfer, Judyta Klimkiewicz, Kamil Czaiński und Piotr Jorroch. Des Weiteren gilt ein besonderer Dank für die Arbeit mit ELAN unserem Mitarbeiter Roman Fisun vom Institut für Slavistik der Universität Regensburg sowie für die Unterstützung während des gesamten Verarbeitungsprozesses unserem – bereits oben erwähnten – Konsultanten Christoph Draxler vom Institut für Phonetik und Sprachverarbeitung der LMU München.

Schließlich möchten wir uns von ganzem Herzen bei allen unseren Gesprächspartnerinnen und Gesprächspartnern bedanken, die uns ihre Geschichten anvertraut haben!

LangGener als internationales Forschungsprojekt

Iga Kościółek

Abstract:

LangGener as an international research project

This chapter describes the institutional background of the binational research project LangGener. The project is funded within the joint funding scheme “Beethoven 2” aimed to strengthen international academic cooperation between Germany and Poland. The projects are supported by the Polish National Science Centre (NCN) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). The chapter describes the joint application process which has to take into account the diverging regulations of the two institutions. Further, we report on the cooperation between the two teams and the challenges we encountered during the project work, especially during the pandemic.

1 Einführung

Das Ziel unseres Forschungsprojekts besteht in einer integrativen Beschreibung des deutsch-polnischen Bilingualismus in Polen und Deutschland, die sowohl linguistische als auch soziolinguistische Aspekte umfasst. Im Untersuchungsfokus stehen die vom deutsch-polnischen Sprachkontakt verursachten morphosyntaktischen Abweichungen in der Sprache der Bilingualen zweier Generationen und deren Klassifizierung. Dabei wird auch auf die tschechisch-deutsche Zweisprachigkeit eingegangen, um zusätzliche unabhängige Evidenz für ausgewählte Aspekte der Forschungsfragen zu liefern. Das finale Ergebnis des Projekts sind digitale, linguistisch annotierte Korpora, die aktuelle Sprachdaten bilingualer Sprecherinnen und Sprecher zur Verfügung stellen. Die Umsetzung eines so umfassend ausgelegten Forschungsvorhabens setzt eine intensive internationale Zusammenarbeit voraus, an der sich polnische, tschechische und deutsche Linguistinnen und Linguisten beteiligen. Die Durchführung der oben genannten Forschungsziele kann folglich durch eine präzise und strukturierte Arbeitsteilung erfolgen, die sich aus der Erhebung des empirischen Forschungsmaterials und der ausführlichen sprachvergleichenden Analyse der erhobenen Sprachdaten zusammensetzt. Im vorliegenden Band möchten wir künftigen Projektleiterinnen und -leitern und internationalen Kooperationspartnerinnen und -partnern unseren Weg zur Verwirklichung der oben genannten Forschungsziele darstellen, indem wir darin auf die einzelnen Arbeitsschritte, die dabei angetroffenen Hindernisse und die von uns gefundenen Lösungen eingehen und alle diese Aspekte vor dem Hintergrund einer internationalen Zusammenarbeit beleuchten. Der vorliegende Band setzt sich aus zehn Kapiteln zusammen, in denen sich die Projektteilnehmerinnen

und -teilnehmer auf die ihrer Ansicht nach wichtigsten Fragen der Anfangs-, Mittel- und Endphase der gemeinsamen Projektarbeit beziehen und so einen Einblick in den Verlauf der einzelnen Arbeitsetappen gewähren. In den im vorliegenden Band enthaltenen Artikeln kommt zudem der projektinterne Arbeitsstil sehr gut zum Ausdruck, da fast alle Beiträge mehrstimmig und von Co-Autoren aus beiden Teams verfasst worden sind. Dadurch werden alle hier präsentierten Themen aus einer differenzierten Betrachtungsweise heraus dargestellt, die kennzeichnend für das Zusammenwirken in einem multinationalen Forschungsprojekt ist. Wir erhoffen uns, dass die vorliegende Arbeit zum einen als ein nützlicher und praktischer Wegweiser für die nachfolgenden internationalen Forschungsgruppen und zum anderen als eine detaillierte und informative Dokumentation der dreijährigen inhaltlich und organisatorisch komplexen wissenschaftlichen Kooperation und Teamarbeit fungieren wird.

2 Förderprogramm BEETHOVEN 2 und Bewerbungsprozess

Die Erfüllung der im vorigen Abschnitt geschilderten Forschungsziele wäre ohne ein geeignetes Förderungsprogramm kaum möglich gewesen. Das Forschungsvorhaben ist in den institutionellen Rahmen der Förderinitiative BEETHOVEN 2 eingebettet, die von der Deutschen Forschungsgemeinschaft (DFG) und dem polnischen Nationalen Wissenschaftszentrum (NCN, Narodowe Centrum Nauki) gemeinsam gegründet wurde. Die Partnerschaftsförderung BEETHOVEN wurde 2014 ins Leben gerufen, um integrierte deutsch-polnische Forschungsprojekte von herausragender wissenschaftlicher Qualität zu fördern.¹ Zunächst wurde der bilaterale Wettbewerb für die Geistes- und Sozialwissenschaften und Kunst ausgeschrieben, in der zweiten Ausschreibung von BEETHOVEN kamen auch die naturwissenschaftlichen Disziplinen hinzu. Das Programm richtet sich vorwiegend an wissenschaftliche Einrichtungen in beiden Ländern und schließt eine Kooperation mit dem industriellen Sektor grundsätzlich aus. Die Teilnahmebedingungen für das Programm sind klar definiert. Für die Bewerberinnen und Bewerber gelten in einigen Bereichen jedoch differenzierte Regeln, die von der jeweiligen nationalen Förderinstitution näher definiert werden. Die institutionell geprägten Unterschiede betreffen hauptsächlich die geringfügig variierenden Vorgaben im Hinblick auf die Bedingungen für die Kooperation mit außer-universitären Forschungseinrichtungen, die Möglichkeit zur Einreichung des Projektantrages bei mehr als einem Förderprogramm und die Finanzierung. Die restlichen formellen und inhaltlichen Kriterien der Antragstellung hingegen sind für die polnischen und die deutschen Bewerberinnen und Bewerber gleich. Dementsprechend muss das gemeinsame Forschungsprojekt aus einem polnischen und einem deutschen Forschungsteam bestehen und jeweils von einem Projektleiter koordiniert werden. Der Durchführungszeitraum ist ebenfalls vorgeschrieben und soll entweder 24 oder 36 Monate betragen. Die Antragstellung für BEETHOVEN erfolgt zweigleisig, indem der deutsch-polnische Forschungsantrag von Leiterinnen und Leitern der beiden Forschungsgruppen bei beiden Förderinstitutionen über zwei IT-Portalsysteme eingereicht wird: OSF (NCN) und ELAN (DFG). Der gemeinsame

¹ Die Information kommt von der Seite: https://www.dfg.de/foerderung/info_wissenschaft/2018/info_wissenschaft_18_60/index.html, letzter Zugriff am 17.03.2021.

Forschungsantrag wird auf Englisch verfasst, das polnische Forschungsteam ist zudem verpflichtet, einen zusätzlichen in PLN kalkulierten Finanzierungsantrag an das NCN in polnischer Sprache zu stellen. Was die Evaluierung des Forschungsvorhabens betrifft, führen die Partnerinstitutionen ein koordiniertes Peer-Review-Verfahren und einen gemeinsamen Auswahlprozess durch. Der Forschungsantrag durchläuft dementsprechend einen zweistufigen inhaltlichen Begutachtungsprozess. Zunächst wird das gemeinsame Forschungsvorhaben sowohl vom NCN als auch von der DFG durch jeweils mindestens zwei externe Gutachter überprüft. In der zweiten Begutachtungsstufe wird der Antrag von einem gemeinsam von der DFG und vom NCN zusammengestellten Expertengremium bewertet. Die Gremien setzen sich aus anerkannten Wissenschaftlerinnen und Wissenschaftlern der relevanten Disziplinen zusammen, die mit den Entscheidungsverfahren der beiden Institutionen vertraut sind. Alle Projektvorschläge werden anhand von drei Kategorien von Kriterien bewertet: 1) wissenschaftliche Qualifikation der Mitglieder der beiden Forschungsteams, 2) Signifikanz des wissenschaftlichen Beitrages und Innovationsgrad des Forschungsprojekts und 3) Plausibilität der geplanten Kosten und Umsetzbarkeit des Projekts. Zudem müssen die Projektpläne gut strukturierte Arbeitsprogramme beinhalten, die den Mehrwert der internationalen Zusammenarbeit klar aufzeigen. Dabei wird erwartet, dass jeder Partner einen wesentlichen Beitrag zum gemeinsamen Projekt leistet, was sich u. a. auch in der Übernahme organisatorischer Verantwortung widerspiegeln soll. Die aufgeführten Voraussetzungen haben ihre Entsprechung in den Expertengutachten gefunden, die der Bewilligung unseres Projektantrages zugrunde liegen. Insgesamt haben wir drei solche Stellungnahmen erhalten, die insbesondere den innovativen Charakter des konzipierten Forschungsvorhabens und der ausgewählten Untersuchungsmethoden im Bereich des deutsch-polnischen Sprachkontakts hervorgehoben haben. Weitere Pluspunkte haben wir für einen realistisch entwickelten Arbeitszeitplan und die interdisziplinäre Aufstellung des internationalen Forschungsteams erhalten. In der Gesamtbegründung des Antrages seitens des NCN wurden zudem die geplante Erstellung und Zurverfügungstellung des digitalen Datenkorpus positiv bewertet. Damit haben wir das von beiden Partnerschaftsinstitutionen hoch bewertete Allgemeinkriterium der Aufbewahrung und Nutzbarkeit von Forschungsdaten erfolgreich erfüllt. Der Weg zur Förderung war jedoch im Falle unseres Vorhabens länger als auf den ersten Blick ersichtlich. Unser Projektantrag wurde in der ersten Ausschreibung des Programms BEETHOVEN, das 2014 seinen Anfang hatte, abgelehnt. Das im Rahmen des vorliegenden Bandes beschriebene Projekt stellt somit ein bei der zweiten Ausschreibungsrunde bewilligtes überarbeitetes Forschungsunterfangen dar.

3 Getrennte Förderinstitutionen – eine Initiative

Wie im vorigen Abschnitt angedeutet, gibt es bei der gemeinsamen Antragstellung und dem Bewerbungsverfahren einige zu beachtende Abweichungen, die auf die systemischen Unterschiede zwischen den beiden Förderinstitutionen, dem NCN und der DFG, zurückzuführen sind. Die genannten Unterschiede beeinflussen nicht nur den Prozess der Beantragung der Förderung, sondern wirken sich zudem direkt auf die Arbeit der beiden Forschungsteams im laufenden Projekt aus. Zu den bedeutendsten Faktoren, deren Einfluss sich auf den Arbeitsverlauf im Projekt deutlich bemerkbar macht, zählen die separate Finanzierung und die in beiden Ländern bestehenden verschiedenen Bewertungssysteme

der projektrelevanten Publikationen. Was die Finanzierung des Forschungsprojekts anbelangt, unterliegt die Vergabe von finanziellen Mitteln länderspezifischen Regeln, die jeweils von der DFG und dem NCN festgelegt werden. Die dem Projekt zugrundeliegenden Budgets werden in der Phase der Antragstellung von beiden Teamleitungen separat ausgearbeitet und bei den Förderinstitutionen entsprechend eingereicht. Die separate Vergabe von Fördermitteln bringt eine klare Aufteilung der beiden Projektbudgets mit sich, was einerseits eine übersichtliche interne (innerhalb der Forschungsgruppen) und externe (gemeinsame Aufgaben der beiden Forschungsgruppen) Koordinationsmöglichkeit der Arbeitsteilung im Projekt sicherstellt. Gleichzeitig bewirkt das aber auch, dass bei der Realisierung der verschiedenen Arbeitsaufträge die Arbeitsorganisation in den Teams um einiges erschwert wird. Wie bereits erwähnt, muss das polnische Team im Laufe des Bewerbungsverfahrens eine zusätzliche Kosteneinschätzung in PLN vorlegen, die sehr detaillierte Angaben beinhaltet. Das Budget für den polnischen Teil des Projekts muss mindestens 150 000 PLN betragen und darf nach dem Beginn der Förderung nicht mehr angepasst werden. Die Finanzmittel müssen folglich auf der polnischen Projektseite im Voraus sehr präzise und vorausschauend für jedes Jahr der Laufzeit des Projekts eingeschätzt und geplant werden, was wenig Raum für eventuelle Abweichungen bietet und eine strenge bürokratische Handhabung zur Folge hat. Bei jeder Änderung der geplanten Ausgaben, sei es der Kauf eines anderen als im angegebenen Budget aufgelisteten Geräts oder sich infolge von externen Faktoren ändernde Reisekosten, muss die polnische Forschungsgruppe jeweils um eine entsprechende Erlaubnis bei den dafür zuständigen Mitarbeitern in ihrer Forschungseinrichtung ersuchen. Zudem fordert das NCN eine jährliche Berichterstattung, die sowohl die im abschließenden Jahr aufgewendeten Kosten als auch Informationen zu dem im gegebenen Zeitabschnitt realisierten Arbeitsplan umfasst. Die Komplexität der aufgestellten Kosten und die damit verbundene stark ausgebaute administrative Formalität bedeuten für die Leiterin des polnischen Teams weniger Freiheit in der Planung der Arbeitsabläufe auf der polnischen Seite und schaffen dabei einen zusätzlichen Zeitaufwand bei der Koordination der Projektdurchführung. Vorteile sind hingegen eine strenge Kontrolle der Ausgaben und gute Kenntnisse der Arbeiten im Projekt, was auch bei der Planung von abweichenden Vorhaben behilflich sein kann. Die Zusammenarbeit mit der DFG verläuft in dieser Hinsicht weniger restriktiv. Der Leiter des deutschen Teams hat in Fragen, welche die Finanzmittel betreffen, deutlich mehr Handlungsfreiheit und unterliegt nicht einem dermaßen komplexen administrativen Aufwand, wodurch sich die Projektkoordination auf der deutschen Seite wesentlich reibungsloser gestalten lässt.

Im Laufe der Projektarbeiten hat sich zudem gezeigt, dass es bei der Planung der einzelnen Arbeitsschritte relevant ist, die Fragen der jeweiligen Verwaltungsstruktur der beiden nationalen Forschungsinstitutionen bereits bei der Phase der Antragstellung einzubeziehen und vorab zu klären. Die Einstellungsmodalitäten der Forschungseinrichtungen, die dem Projekt im jeweiligen Land den institutionellen Rahmen geben, sind im Hinblick auf seine Umsetzbarkeit nicht zu unterschätzen. So ermöglichte das den polnischen Teil des Projekts überdachende Institut der Slavistik der Polnischen Akademie der Wissenschaften das Outsourcing der für die Korpuserstellung unentbehrlichen Transkriptionsarbeiten, indem es den Abschluss von sog. Werkverträgen mit freiberuflichen Transkribentinnen und Transkribenten zuließ. Der Abschluss von Arbeitsverträgen dieser Art mit mehreren Transkribentinnen und Transkribenten für unterschiedlich große Lieferungspakete garantierte eine unterbrechungsfreie Vorbereitung der Transkripte, die von

der Fluktuation der transkribierenden Personen unberührt blieb. Im Falle einer Aufgabe der weiteren Zusammenarbeit seitens der beauftragten Freiberufler war man in der Lage, relativ kurzfristig einen neuen Werkvertrag mit einem anderen Dienstleister zu schließen oder die Zusammenarbeit mit der vorübergehend ausgestiegenen Person zu einem späteren Zeitpunkt wieder aufzunehmen. Die Universität Regensburg hingegen, der Arbeitsort des deutschen Forschungsteams, bot in dieser Hinsicht weniger Spielraum an, da hier eine solche Vergabe von kleineren Aufträgen an externe Dienstleister kaum möglich war. An der deutschen Universität gab es bei der Verteilung derartiger Aufgaben eine andere Möglichkeit der Arbeitsvergabe, und zwar die WHK-Stellen, die aber wegen ihrer zeitlichen Festlegung keine vergleichbare Flexibilität in Bezug auf die Einstellungsmöglichkeiten garantierten.

Ein weiterer systembezogener Unterschied mit direkter Einflussnahme auf die gemeinsame Zusammenarbeit im Projekt betrifft das Bewertungssystem der Publikationen, das von beiden Förderinstitutionen angewandt wird. Während die DFG bei der Auswahl des Verlages für die Herausgabe des projektgewidmeten Bandes den Forscherinnen und Forschern freie Hand gibt, richtet sich dabei das NCN nach der Rangliste der Zeitschriften und Verlage, die einen hohen sog. *Impact Factor* aufweisen (siehe die Suchmaschine des polnischen Ministeriums für Bildung und Wissenschaft für Publikationsorgane <https://punktoza.pl/>). Die Bewertung der Publikationen wird in Polen anhand des Journals, in welchem sie veröffentlicht wurden, aufgestellt. Auf dieser Grundlage werden den Wissenschaftlerinnen und Wissenschaftlern und den Forschungsinstituten und Universitäten vom polnischen Ministerium für Wissenschaft und Hochschulbildung bestimmte Punktzahlen zugewiesen, die für den weiteren beruflichen Werdegang der Wissenschaftlerinnen und Wissenschaftler sowie auch für die Vergabe von Forschungsressourcen an die wissenschaftlichen Einrichtungen ausschlaggebend sind. In Deutschland wird die Forschungsleistung einzelner Wissenschaftlerinnen und Wissenschaftler ebenso anhand der Anzahl von Publikationen in Top-Journals gemessen, wobei diese Praxis dort hauptsächlich auf die Natur- und Sozialwissenschaften beschränkt ist und kein allgemeingültiges Bewertungsinstrument auf der staatlichen Ebene darstellt. Die Suche nach einem geeigneten Verlag für unseren Band wurde also von den formalen Vorgaben geprägt, die manche in der Germanistik und Slavistik bewährten und anerkannten Publikationsreihen ausschließen und dagegen einen Beitrag in einem hochrangigen, meist englischsprachigen Verlag bevorzugen und empfehlen. Dabei ist anzumerken, dass die Publikation in einem renommierten Verlag nicht nur ein zeitaufwendiges Unterfangen ist, dem viele Vorverhandlungen mit mehreren Verlagen vorangehen, sondern auch mit zusätzlichen Kosten einhergeht, die im Projektbudget vorab berücksichtigt werden sollten.

4 Zwei Forschungsgruppen – ein Team

Unterschiede, die systemischen Ursprungs sind, lassen sich in einem internationalen Forschungsprojekt schwer vermeiden. Umso wichtiger für eine erfolgreiche Umsetzung des Projekts wird also eine gute Zusammenarbeit zwischen beiden Forschungsteams. Dabei geht es nicht nur um eine effiziente und gut durchdachte Aufteilung des umfangreichen Arbeitspensums unter den Projektteilnehmerinnen und -teilnehmern, sondern vielmehr um eine qualitative und vertrauensbasierte Kooperation. Alle am LangGener-Projekt

mitwirkenden Personen sind Wissenschaftlerinnen und Wissenschaftler mit einschlägiger Expertise und internationaler Kooperationserfahrung. So ging auch dem Projektantrag LangGener eine langjährige Kooperation der beiden beteiligten Institute für Slavistik voraus. Die beiden Projektleiter, Prof. Dr. Björn Hansen und Prof. Dr. Anna Zielińska, sind einander bereits 1998 an der Universität in Hamburg begegnet, wo sie an dem dortigen Institut für Slavistik wissenschaftlich tätig waren. Seitdem hat sich die gemeinsame Kooperation zwischen den Sprachwissenschaftlern verfestigt und in der gegenseitigen Beteiligung an den zahlreichen sowohl vom Institut für Slavistik der Polnischen Akademie der Wissenschaften als auch vom Regensburger Forschungszentrum Deutsch in Mittel-, Ost- und Südosteuropa (DiMOS) veranstalteten Konferenzen und Tagungen geäußert. Der erste Entwurf des Projektes LangGener ist auch während einer der slavischen Sprachwissenschaft gewidmeten Konferenzen in der Polnischen Akademie der Wissenschaften 2014 entstanden, wo den beiden Wissenschaftlern die Idee kam, ihre Fachbereiche – Soziolinguistik und Morphosyntax – im Rahmen eines Projekts zum deutsch-slavischen Sprachkontakt zu vereinigen. Wie schon im 2. Abschnitt des Artikels erwähnt, wurde diese Idee bei der zweiten Edition des Programms BEETHOVEN (BEETHOVEN 2) positiv beschieden und erhielt im Jahre 2018 die Finanzierung. Aufbauend auf ihren Netzwerken und bisherigen Kooperationen haben die Projektleiter ihre Teams zusammengestellt, die sie sowohl mit erfahrenen und fachkundigen Sprachwissenschaftlerinnen und -wissenschaftlern als auch mit den Dokortitel anstrebenden Nachwuchswissenschaftlerinnen besetzt haben.

Am LangGener-Projekt sind also insgesamt zwei Forschungsgruppen beteiligt, die sich aus drei Teams zusammensetzen – dem polnischen, dem deutschen und dem deutsch-tschechischen Team. Das deutsch-tschechische Team wird von Prof. Dr. Marek Nekula (Professor für Bohemistik und Westslavistik und Leiter des Bohemicum – Center for Czech Studies) geleitet, der im Rahmen des Projekts LangGener als mitverantwortlicher Antragsteller die tschechisch-deutsche Zweisprachigkeit im böhmischen und bayerischen Kontext gemeinsam mit M.A. Aneta Bučková (Doktorandin, wiss. Mitarbeiterin) erforscht. Was das polnische Team angeht, so kommen neben Prof. Dr. Zielińska (Institut für Slavistik der Polnischen Akademie der Wissenschaften, Warschau) noch drei Germanistinnen aus Polen hinzu, und zwar Dr. habil. Felicja Księżyk (Universität Opatów), Dr. Barbara Jańczak (Universität Posen), Dr. Anna Jorroch (Universität Warschau) und die Polonistin Dr. Irena Prawdź (Institut für Slavistik der Polnischen Akademie der Wissenschaften Warschau, Postdoc). Dr. Barbara Jańczak, Dr. habil. Felicja Księżyk und Dr. Anna Jorroch haben in ihren bisherigen Arbeiten u. a. die deutschen Minderheiten in Schlesien (Księżyk) und Masurien (Jorroch) sowie die Sprachkontakte in der deutsch-polnischen Grenzregion (Jańczak, Zielińska) untersucht und so eine in Hinblick auf die Gewinnung und das Interviewen der Informantinnen und Informanten wertvolle Feldforschungserfahrung ins Projekt eingebracht. Neben Prof. Dr. Hansen (Inhaber des Lehrstuhls für slavistische Linguistik, Universität Regensburg) besteht das deutsche Team aus Dr. Michał Woźniak (Institut für Polnische Sprache der Polnischen Akademie der Wissenschaften, Krakau), dem für die technische Erstellung des Korpus verantwortlichen Korpuslinguisten, und zwei wiss. Mitarbeiterinnen: M.A. Carolin Centner (Universität Regensburg, Doktorandin) und M.A. Iga Kościółek (Universität Regensburg). Darüber hinaus haben auch externe Konsultantinnen und Konsultanten am Projekt mitgewirkt, deren Erfahrung in der Gestaltung von digitalen Datenbanken (Prof. Dr. Sandra Birzer,

Corpus of Modern Yiddish, Universität Bamberg; M.A. Roman Fisun, Universität Regensburg) und Expertise im Bereich des deutsch-polnischen Bilingualismus (Prof. Dr. Bernhard Brehmer, Universität Greifswald) für die Planungsphase des Workflows von großer Bedeutung war.

Was die Festlegung der Arbeitsteilung im Projekt betrifft, erfolgte die Zuweisung von Hauptaufgaben unter Berücksichtigung der Kompetenzen und Fachbereiche der beiden Forschungsgruppen. So übernahm das polnische Team die Verantwortung für die Organisation der Feldforschung sowie die Entwicklung und Durchführung der soziolinguistischen Annotation, wohingegen das deutsche Team für die Erstellung des Korpus und die morphosyntaktische Annotation zuständig war. Das tschechische Team hat einen großen Beitrag bei der Erarbeitung der soziolinguistischen Annotation und Bearbeitung der Metadaten geleistet. An allen zum Projekt-Workflow gehörenden Aufgaben waren jedoch alle Mitglieder des Projekts in gleichem Maße beteiligt, indem sie sämtliche im Projekt vorhandene Sprachdaten erhoben, technisch bearbeitet und entweder Musterreplikationen (PAT: pattern replications) oder soziolinguistische Angaben (SOZ) annotiert haben.

Die Zusammenarbeit zwischen beiden Forschungsgruppen stützte sich auf interne Kommunikation und die im Projekt angewendeten Software-Lösungen. Der Austausch der Projektdaten und der Zugriff darauf waren durch die Nutzung von speziellen Speicherplattformen wie MyFiles (verwaltet von der Universität Regensburg) und Git (erstellt und verwaltet von Dr. Michał Woźniak) möglich, die jedem der Projektmitglieder über die gesamte Laufzeit des Projekts zur Verfügung standen. Die interne Kommunikation verlief vorwiegend durch E-Mails, es fanden aber auch mehrere Skype- und Zoom-Treffen statt, was in der Zeit der Pandemie auch in unserem Projekt zum gängigen Kommunikationsmedium wurde. In der Zeit vor COVID-19, d. h. in den ersten zwei Jahren des Projekts, sind die beiden Forschungsgruppen mehrmals sowohl in Regensburg als auch in Warschau zu mehrtägigen Teambesprechungen zusammengetroffen. Insgesamt gab es in dieser Zeitperiode zwei Teamtreffen im Institut der Slavistik in Warschau, vier Treffen in Regensburg und ein Treffen in Krakau, die insbesondere für die Planung der weiteren Schritte im Projekt, aber auch für die Entstehung des gegenseitigen Vertrauens und die Teamentwicklung sehr bedeutungsvoll waren. Sämtliche Teambesprechungen wurden in Form von Protokollen dokumentiert und auf die Plattform Trello hochgeladen, ein internes Archivierungsmedium, wo auch alle in den Besprechungsrunden vorgestellten Power-Point-Präsentationen sowie Fotos von gemeinsamen Treffen festgehalten wurden. Parallel zur internen Kommunikation und Dokumentation der Projektarbeit wurden zusätzlich Schritte zur Vermittlung der Projektinhalte und der Projektergebnisse an die Öffentlichkeit ergriffen. Diesbezüglich haben die Projektteilnehmerinnen und -teilnehmer an mehreren fachlichen Konferenzen, Tagungen und Workshops teilgenommen, hierunter in Polen (Wrocław, Łubowice, Gdańsk, Łódź, Warszawa, Opole, Kędzierzyn-Koźle, Gorzów Wielkopolski), Deutschland (Frankfurt/Oder, Jena, Potsdam, Trier), Tschechien (Prag), Italien (Palermo), England (York, Sheffield) und USA (Boston). Wegen der epidemischen Lage und den damit verbundenen Beschränkungen konnten einige für 2020 geplanten Konferenzen entweder ausschließlich im Online-Modus stattfinden (zum Beispiel in Washington), andere wurden verlegt oder gänzlich gestrichen (zum Beispiel Internationale Konferenz in Namibia), was natürlich die Vermittlung der im Projekt erbrachten Leistungen wesentlich erschwert hat. Darüber hinaus wurden die projektrelevanten Erkenntnisse in fachlich anerkannten internationalen Jahrbüchern und Zeitschriften, wie zum Beispiel

in *Multilingua* (Zielińska & Księżyk 2021), *Language & History* (Nekula 2021a), *Wiener Slawistisches Jahrbuch* (Nekula 2021b), *Brücken – Zeitschrift für Sprach-, Literatur- und Kulturwissenschaft* (Bučková 2021a; Nekula 2021c) oder in *Naše řeč* (Bučková 2021b, Nekula 2021d), sowie in Sammel- und Tagungsbänden präsentiert, die bei Peter Lang Verlag (Jorroch 2020) oder Nomos Verlag (Nekula 2021) erschienen sind. Es sind noch weitere Veröffentlichungen geplant, wie zum Beispiel der 2021 im *Journal of Language Contact* erscheinende gemeinsame Beitrag „Und dann hat er sich verheiratet“ – German-Polish bilingualism on the example of reflexive pattern replications by the German minority from Pomerania- and Warmia-Masuria Provinces“ von C. Centner und B. Jańczak sowie der im Harrasowitz Verlag zu veröffentlichende Beitrag „Rekonstruktion von Identität in oberschlesischen Sprachbiographien“ von F. Księżyk. Bisher kann das Publikationskonto des Projekts elf Veröffentlichungen, die im Erscheinungszeitraum 2020–2021 herausgegeben wurden bzw. zu erwarten sind, vorweisen.

5 Abschlusswort

Die gemeinsame deutsch-polnische Zusammenarbeit im Projekt LangGener wurde innerhalb der dreijährigen Laufzeit des Projekts von unterschiedlichen, zum Teil auch unerwarteten internen als auch externen Faktoren geprägt, die die Forschungsarbeit auf verschiedene Art und Weise mitgestaltet haben. Die Suche nach und die Einstellung von geeigneten Transkribentinnen und Transkribenten und der Ausbruch der Pandemie gehören zu den größten Zäsuren bei der Ausführung der im Projekt festgesetzten Ziele. Trotz der anfänglichen Schwierigkeiten mit der Bewältigung dieser Hindernisse war es uns dennoch möglich, die vereinbarten Forschungsaufgaben erfolgreich durchzuführen und den Zeitplan relativ verzögerungslos einzuhalten. Ohne ein gut aufgestelltes zusammengeschweißtes Forschungsteam wäre jedoch ein solches Ergebnis mit Sicherheit nicht möglich gewesen. In den insgesamt vier Jahren der intensiven Projektarbeit hat nicht nur ein reger Austausch auf professioneller Ebene stattgefunden, vielmehr sind gegenseitiges Vertrauen und eine Partnerschaft entstanden, die allen Projektteilnehmerinnen und -teilnehmern gleichermaßen zuteilwurden. Die gemeinsame Arbeitsweise zeichnete sich durch eine kollegiale, tolerante Arbeitsstimmung aus, die immer, ungeachtet der akademischen Grade der Projektmitglieder, auf Augenhöhe verlaufen ist. Jedes Projektmitglied hatte das gleiche Recht, seine Meinung zum jeweiligen Thema zu äußern, Kritik zu üben und Verbesserungen vorzuschlagen. Diesem Ergebnis liegen zahlreiche, mehrstündige und -tägige persönliche Treffen zugrunde, die sich als essenziell für die Erarbeitung der partnerschaftlichen Kommunikationsgrundlage erwiesen haben. Obwohl die später erfolgte permanente Umstellung auf digitale Besprechungsräume (Skype, Zoom) von uns allen als eine Erschwernis wahrgenommen wurde, war die Aufrechterhaltung einer guten und inhaltlichen Kommunikation dank der vorherigen, intensiven Treffen in Person möglich. Gute Kommunikation, gegenseitiges Verständnis, Raum für gemeinsame Diskussionen und Feedback-Runden stellen unserer Erfahrung nach die wichtigsten Bestandteile eines gelungenen und erfolgreich durchgeführten Forschungsprojekts dar, und es sind eben jene Elemente der Zusammenarbeit, die wir nachfolgenden Forschungsgruppen ans Herz legen möchten.

Literatur

- Bučková, Aneta (2021a): Jazykové ideologie a jazykový management československých emigrantek a emigrantů v Bavorsku. In: *Naše řeč* 104. 5, 376–392.
- Bučková, Aneta (2021b): Syntaktische Musterreplikationen bei deutsch-tschechischen Bilingualen: Ein gebrauchsbasierter Ansatz. In: *Brücken – Zeitschrift für Sprach-, Literatur- und Kulturwissenschaft* N.F. 28. 2, 83–109.
- Centner, Carolin/Jańczak, Barbara (im Druck): Und dann hat er sich verheiratet – German-Polish bilingualism on the example of reflexive pattern replications by the German minority from Pomerania- and Warmia-Masuria Provinces. I.V.
- Deutsche Forschungsgemeinschaft (DFG): <https://www.dfg.de/>, letzter Zugriff am 07.12.2021.
- Jorroch, Anna (2020): Asymmetrische Wortäquivalenz in der Rede der Zweisprachigen. In: Maria Biskup, Anna Just (Hgg.), *Vielfalt der Linguistik: Bausteine zur diachronen und synchronen Linguistik*. Berlin: Peter Lang, 66–77.
- Księżyk, Felicja (2021): Rekonstruktion von Identität in oberschlesischen Sprachbiographien. In: *Die Welt der Slaven* 66. 1, 1–18.
- Narodowe Centrum Nauki (NCN): <https://ncn.gov.pl/ogloszenia/konkursy/beethoven2>, letzter Zugriff am 07.12.2021.
- Nekula, Marek (2021): Grenzen und Sprachgrenzen in der Sprachwissenschaft. In: Maria Klessmann, Dominik Gerst, Hannes Krämer (eds), *Grenzforschung: Handbuch für Wissenschaft und Studium*. Baden-Baden: Nomos Verlag, 400–418.
- Nekula, Marek (2021a): Language loyalty and language shift in Bohemia in the long 19th century based on the language biography of Friedrich (Bedřich) Smetana. In: *Language & History* 64. 2, 92–112
- Nekula, Marek (2021b): Varieties of German in the Czech Republic. In: *Wiener Slavistisches Jahrbuch* N.F. 9, 69–103.
- Nekula, Marek (2021c): Sprachideologien und Sprachmanagement in sprachbiographischen Interviews: Sprecher mit tschechischem Migrationshintergrund in Bayern. In: *Brücken – Zeitschrift für Sprach-, Literatur- und Kulturwissenschaft* N.F. 28. 2, 63–82.
- Nekula, Marek (2021d): Čeština v Německu – jazyk v kontaktu a izolaci v „generaci exil“. In: *Naše řeč* 104. 5, 359–375.
- Punktoza: <https://punktoza.pl/>, letzter Zugriff am 07.12.2021.
- Zielińska, Anna/Księżyk, Felicja (2021): Language shifts in the language biographies of immigrants from Upper Silesia residing in Germany. In: *Multilingua* 40. 5, 675–706.

Konzeptuelle Überlegungen

Der Generationsbegriff

Carolin Centner

Abstract:

The notion of *generation*

In linguistics the term *generation* is often used somewhat ambiguously in different contexts and with different meanings. The aim of this chapter is to shed some light into the use of the term *generation* in linguistics by dividing it into three dimensions where it is used. From a practical perspective, it is shown how the term is used in publications on German-Slavonic language contact. As none of these definitions is suitable for the Lang-Gener project, a new approach how to define *generations* will be presented in the end of the chapter.

1 Der Generationsbegriff in der Linguistik

Generation ist ein Begriff, der in verschiedenen Kontexten, auch in der öffentlichen Debatte, häufig gebraucht wird, beispielsweise *babyboomer* Generation oder *millennials*. Dabei wird auf Konflikte zwischen Generationen hingewiesen, denen bestimmte Eigenschaften zugeschrieben werden. Auch im wissenschaftlichen Diskurs ist dieser Begriff, der in der Soziologie von Karl Mannheim (1928) geprägt wurde, weit verbreitet. Doch was genau kennzeichnet eine Generation? Wie Neuland treffend beobachtet, ist es „bemerkenswert, dass der Generationsbegriff zwar ‚multidisziplinäre Perspektiven‘ aufweist und in den Bildungs-, Sozial- und Kulturwissenschaften eine bedeutende Rolle spielt, nicht aber in der Sprachwissenschaft, in deren Handbüchern und Grundlagenwerken er keinerlei Beachtung findet“ (Neuland 2015: 11). Darüber hinaus werden Generationsbegriffe in verschiedenen Kontexten auf unterschiedliche Weise verwendet, sodass es nicht möglich ist, einheitliche Kriterien oder gar eine Definition des Begriffs ausfindig zu machen. Da das Projekt jedoch in seinem deutsch-polnischen Teil zwei Generationen miteinander vergleicht, gilt es, zunächst den Begriff der Generation genauer zu beleuchten. Weitere Ausführungen bezüglich der Definition der Generationen im Projekt sind ebenfalls in Centner (in Vorbereitung) zu finden.

1.1 Unterschiedliche Dimensionen des Generationsbegriffs

Die Unschärfen in der linguistischen Definition des Generationsbegriffs thematisiert Gerstenberg (2015). Sie untersucht die Differenzen der Verwendungen und macht drei

Varianten aus, in welchen der Generationsbegriff im Vordergrund steht und unterschiedlich gebraucht wird. Der absolute *Generationsbegriff* wird vor allem im soziolinguistischen Kontext benutzt. Der zentrale Ausgangspunkt dabei ist „die Beobachtung und Deutung“ sprachlicher Variation, welche mit Generationen verbunden ist (Gerstenberg 2015: 51). Generation wird hier als Altersschicht festgelegt, wobei die Kindheit und Jugend der soziolinguistischen Argumentation von Weinreich et al. (1968: 108) folgend als prägende Phasen gesehen werden. Es wird zwischen *generational change*, welcher vorrangig die lautliche und morphologische Ebene betrifft, und *communal change*, welcher Sprachwandel innerhalb einer Sprachgemeinschaft im Bereich der Lexik und Syntax meint, unterschieden (Gerstenberg 2015: 52ff.).

Für den *historisch-soziologischen Generationsbegriff* stehen Begriffe wie *Kontinuität* vs. *Bruch* im Vordergrund. Der Begriff kommt immer dann zum Tragen, wenn ein historisches Ereignis eine Spaltung zwischen einer jüngeren und einer älteren Generation hervorruft. Als Beispiel hierfür nennt Gerstenberg (2015: 56) die Industrialisierung, infolgeder die jüngere Generation gegen die ältere rebellierte, oder die 1968er Bewegung, die als Auflehnung gegen die ältere, ‚verstaubte‘ Generation gesehen werden kann. Wie auch beim *absoluten Generationsbegriff*, werden hier eher Altersschichten als genau definierte Generationen betrachtet. Es fehlt an einer Legitimierung des Generationsbegriffs durch eine genaue Abgrenzung außerhalb historischer Ereignisse.

Eine weitere Möglichkeit, Generationen zu definieren, sieht Gerstenberg (2015: 54f.) in der *relativen Verwendung des Generationsbegriffs*, welche vor allem in der Migrationslinguistik zum Tragen kommt. Laut Gerstenberg wurde dieser Begriff vordergründig durch die Debatte um die Sozialisation und Integration der ‚zweiten Generation‘ der Gastarbeiterfamilien in den 1970er Jahren in Deutschland geprägt. Im Hinblick auf den Erwerb der deutschen Sprache wurde diese Generation allerdings als defizitär dargestellt, sodass der Begriff zudem negativ konnotiert ist (ebd.).

Da der Migrations-Generationsbegriff eng mit der Geschichte der Migration im untersuchten Land verknüpft ist, gilt es, sich hier an konkreten historischen Kontexten zu orientieren. Für den deutschen Kontext nennt Gerstenberg als Bezugspunkt die Studie von Schrader et al. (1976), welche sich an der ‚zweiten Generation‘ orientiert. In dieser Studie wird die erste Generation in Deutschland durch die Einreise charakterisiert; diese findet in den 1950er Jahren statt. Die Nachfahren dieser Generation, die zweite Generation geht in den 1970ern zur Schule, während wiederum deren Nachfahren, also die dritte Generation ab den 1990ern Form annimmt (Gerstenberg 2015: 55). Dieses Modell kennzeichnet Generationen in Wellen.

Für den Migrationsdiskurs ist diese Lösung ausreichend und wird oft angewandt. Es stellt sich allerdings die Frage, ob in dem vorliegenden Forschungsvorhaben, das besondere Spezifika aufweist, dieses Modell anwendbar ist. Eine Besonderheit ist, dass das Forschungsvorhaben zwei Generationen, die aber in unterschiedlichen Ländern angesiedelt sind, vergleicht. Eine Generation wird in Polen untersucht, die andere in Deutschland. Beide Generationen haben den Fokus auf deutsch-polnisch Bilinguale gemeinsam, weswegen im Folgenden geprüft wird, wie Generationen im deutsch-polnisch bilingualen Migrationskontext definiert werden.

1.2 Der Generationsbegriff im Kontext deutsch-slavisch Bilingualer

Ein Grundlagenwerk hierzu ist das *Handbuch der deutschen Sprachminderheiten in Mittel- und Osteuropa*. Hier werden, wie in einem Teil des Forschungsvorhabens auch, deutsche Minderheiten in Mittel- und Osteuropa beschrieben. Polen wird hier von Lasatowicz & Weger (2008: 145ff.) behandelt, wobei man sich nach Informationen über Geographie und Geschichte vor allem auf das Opperner Schlesische konzentriert. Dabei nimmt der sprachliche Aspekt einen kleinen Teil ein: Ein Kapitel wird dem generationsspezifischen Sprachgebrauch gewidmet. Die Daten hierfür stammen aus dem Opperner Spracharchiv der Universität Oppern und wurden mittels Tonbandaufnahmen gesammelt, wobei der Fokus auf „Sprechern der ältesten Generation“ liegt (Lasatowicz & Weger 2008: 159). Im Rahmen des Projekts wurden die Probandinnen und Probanden in drei Altersgruppen eingeteilt, die im weiteren Verlauf ebenfalls Generationen genannt werden (vgl. zum Beispiel ebd.: 164). Dabei werden die Altersgruppen ausschließlich anhand des Geburtsjahres eingeteilt. So gehören Sprecherinnen und Sprecher, welche vor 1940 geboren wurden, der ältesten, also in diesem Fall der ersten Generation an. Die zweite Generation beinhaltet Sprecherinnen und Sprecher, welche zwischen 1940 und 1970 geboren wurden, die dritte wiederum Sprecherinnen und Sprecher, die nach 1970 geboren wurden (ebd.: 159). Eine Erklärung für die scheinbar arbiträre Einteilung, zum Beispiel in Form von einer historischen Argumentation, gibt es nicht. Lediglich die Spanne von 30 Jahren entspricht in etwa dem natürlichen Abstand zwischen Generationen, wie diese innerhalb einer Familie auftreten. Dieses Modell ist damit nicht für das Forschungsvorhaben anwendbar, in welchem linguistische Faktoren zentral sind, um Sprachkontakt innerhalb zweier unterschiedlicher Generationen zu untersuchen.

Im gleichen Handbuch werden zur Untersuchung der deutsch-tschechischen Sprachkontaktsituation zwar ähnliche Kriterien wie im Text zu Polen gewählt, jedoch vor einem anderen Hintergrund. Hierfür präsentiert Tišerová (2008: 203) folgende Tabelle:

	Generation I vor 1938 geboren	Generation II zwischen 1939 und 1945 geboren	Generation III zwischen 1946 und 1980 geboren	Generation IV seit 1981 geboren
L1	Deutscher Dialekt mit Transferer- erscheinungen	Deutscher Dialekt mit Attritions- erscheinungen; Tschechisch	Tschechisch; Deutscher Dialekt nur rudimentär	Tschechisch
L1'	Standarddeutsch nur rudimentär			
L2	Tschechisch teilweise als Interlanguage	Standarddeutsch als Interlanguage	Standarddeutsch als Interlanguage, meist mit großer Entfernung zur Zielsprache	Standarddeutsch als Interlanguage, auf unterschiedli- chem Niveau

Tabelle 1: Sprechertypen nach Alter (Tišerová 2008: 203)

Im Gegensatz zum Kapitel über die Minderheit in Polen, kann man hier eine geschichtliche Orientierung bei der Wahl der Jahresgrenzen der Generationen erkennen. Auffallend aber ist besonders der unterschiedliche Fokus. Bei dem tschechischen Beispiel stehen die Sprachkompetenzen im Vordergrund. Anhand des linguistischen Kriteriums lässt sich durch die Einteilung in Generationen ein Erkenntnisgewinn generieren. So lässt sich anhand der Tabelle beobachten, dass es innerhalb von vier Generationen zu einem Sprachwechsel (*language shift*) kommt. Während in der Generation I der deutsche Dialekt L1 ist, ist bei Generation IV Tschechisch L1.

Auch für das LangGener-Projekt ist *language shift* ein wichtiger Begriff. Diesen definiert Pauwels (2016: 18) folgendermaßen:

„LS [language shift; Anm. d. Verf.] involves the gradual replacement of one’s main language or languages, often labelled L1, by another language, usually referred to as L2, in all spheres of usage“.

Wichtig ist hierbei anzumerken, dass sich L1 nicht auf die zuerst erlernte Sprache, sondern auf die dominante Sprache bezieht. Zudem wird hier betont, dass bei einem vollständigen *language shift* alle Bereiche der Sprachen von einem Dominanzwechsel betroffen sein müssen.

Riehl (2014: 68) bezieht sich auf Clyne (2003: 20) und nennt drei Aspekte des Sprecherwechsels, wobei der erste hier besonders relevant erscheint: „Änderung des Sprachverhaltens in einer ganzen Sprachgemeinschaft, in einer Teilgruppe der Sprachgemeinschaft oder beim Individuum“. Dabei vollzieht sich dieser Sprachwechsel in traditionellen Sprachgemeinschaften allmählich. Bei allochthonen Sprachminderheiten verläuft dieser Wechsel jedoch deutlich schneller; gerade bei städtischen Migrantinnen und Migranten kann es durch die fehlende Weitergabe der Sprache innerhalb einer Generation zu einem *language shift* kommen (ebd.: 69).

Die oben angeführte Tabelle eignet sich dazu, den *shift* innerhalb in Isolation geratener Sprechergruppen darzustellen (zur kritischen Diskussion vgl. Nekula 2021). Für die Zwecke dieser Arbeit, vor allem in Hinblick auf die Generationen in zwei unterschiedlichen Ländern, ist diese Definition der Generationen jedoch nicht ausreichend.

Eine andere Perspektive der Generationsfrage bietet Kaluza (2002), der sich mit polnischen Immigrantinnen und Immigranten in Deutschland aus einer historischen Perspektive auseinandersetzt. Im Gegensatz zu Lasatowicz & Weger (2008) wählt er keine Zeitspanne, welche aus Familienkonstellationen entsteht, sondern charakterisiert Auswanderungswellen. Die erste Migrationswelle von Polen nach Deutschland beschreibt er als Ruhr-Polinnen und -Polen, welche um 1900 aus den zu Preußen gehörenden Teilen Polens vorwiegend in das Gebiet an Rhein und Ruhr auswanderten. Die zweite Welle stellt die Gruppe der nach Deutschland verschleppten Zwangsarbeiterinnen und Zwangsarbeiter und KZ-Häftlinge, die nach 1945 in Deutschland geblieben sind, dar. Die dritte Gruppe besteht aus sogenannten Spätaussiedlerinnen und Spätaussiedlern, die nach 1945 zwar in Polen geblieben sind, ab den 1950ern aber wieder nach Deutschland abgewandert sind. Als letzte Migrationswelle werden diejenigen Auswanderinnen und Auswanderer genannt, welche nach den 1980ern auf der Suche nach Arbeit nach Deutschland gekommen sind (Kaluza 2002: 699ff.). Auch Kaluza wählt historische Ereignisse als Kriterium, um die Wellen voneinander abzugrenzen. Er verwendet den Begriff der Generation nicht.

Trotzdem erscheint die dritte Gruppe für die vorliegende Untersuchung insofern relevant, als dass sie Spätaussiedlerinnen und Spätaussiedler umfasst, welche ebenfalls im Rahmen des Projekts untersucht werden. Allerdings nennt er keine linguistischen Kriterien, die diese Gruppe einengen würden.

Zusammenfassend lässt sich sagen, dass die bisherigen Definitionen des Generationsbegriffs für das vorliegende Forschungsvorhaben nicht zielführend sind. Die Spezifik des Projekts, welches Generationen in zwei Ländern untersucht, erlaubt es nicht, rein historische Ereignisse als Eckpfeiler für die Definition zu verwenden. Zudem ergibt sich aus den bisherigen Definitionen eine Gruppe von Probandinnen und Probanden, welche unter linguistischen Gesichtspunkten sehr heterogen erscheint. Dies mag besonders für die Soziolinguistinnen und Soziolinguisten von Interesse sein. Da aber das Hauptziel des Projektes im Vergleich beider Generationen hinsichtlich Sprachkontaktphänomenen besteht, muss diese Gruppe von Probandinnen und Probanden durch linguistische Kriterien enger definiert werden. Es kann also Gerstenberg nur zugestimmt werden, wenn sie sagt: „wenn die Gleichsetzung von Altersgruppe mit Generation aufgegeben wird, so die zugrundeliegende Auffassung, können die Phänomene sprachlichen Wandels nicht nur beschrieben, sondern auch besser verstanden werden“ (Gerstenberg 2015: 51).

1.3 Definition der Generationen im LangGener-Projekt

In diesem Sinne definieren wir die Generationen in unserem Projekt bzw. in seinem deutsch-polnischen Teil anhand einer Auswahl von linguistischen Kriterien. Hierzu muss aber zunächst die Benennung der Generationen angesprochen werden. Wie bereits zuvor deutlich geworden ist, werden Generationen oft durchnummeriert. Davon ausgehend wäre die Generation, welche in Polen lebt und im Projekt die ältere Generation darstellt, die erste Generation, während die zweite Generation diejenigen Immigrantinnen und Immigranten in Deutschland meinen würde, welche vor der Auswanderung der Gruppe der autochthonen Deutschen in Polen zugehörten. Allerdings ist das in Hinblick auf die Definition der Generationen aus einem migrationslinguistischen Blickwinkel heraus irreführend. Dies ist auf die Spezifik, dass die Generationen in unterschiedlichen Ländern untersucht werden, zurückzuführen. Die vermeintlich erste Generation, also die Deutschsprachigen in Polen, ist aus einem polnischen Blickwinkel nicht die erste Generation, sondern vielmehr eine der letzten, deren Angehörigen noch ihr Deutsch behalten, denn nach ihr kommt es weitestgehend zu einem *language shift*. Auch ist diese Generation nicht nach Polen eingewandert, sondern nach dem Krieg geblieben, lediglich das Land hat sich hier verändert. Auch die vermeintlich zweite Generation stellt im migrationslinguistischen Diskurs vielmehr die erste Generation dar, da diese Probandinnen und Probanden aus Polen nach Deutschland migrieren. Um diese Verwirrung zu vermeiden und die gewählten Termini einzuhalten, sollen im Projekt die Generationen nicht mithilfe von Ordinalia klassifiziert werden. Die ältere Generation, also die deutsche Minderheit in Polen, wird im Folgenden nach dem Land, in dem sie wohnt, „Generation Polen“ genannt. Im Gegensatz dazu wird die Generation der Auswanderinnen und Auswanderer nach Deutschland analog dazu „Generation Deutschland“ genannt.

„Generation Polen“ (GP) umfasst Probandinnen und Probanden, die auf dem Gebiet des Deutschen Reichs geboren wurden, welches infolge des Zweiten Weltkriegs an Polen fiel. Definitorische Merkmale sind daher einerseits die Zweisprachigkeit der Sprecherinnen und Sprecher und andererseits die Herkunftsregion. Die zwischen 1933 und 1945 geborenen Probandinnen und Probanden wachsen auch nach 1945 zunächst in einer deutschsprachigen familiären Umgebung auf, ihre L1 ist Deutsch. Durch den politischen Wechsel um 1945 und die veränderten Rahmenbedingungen, die die Sprachenpolitik und Sprachplanung auf der Makroebene verändern, erlernen sie das Polnische noch vor dem Abschluss der sogenannten kritischen Periode, welche auf das 12. Lebensjahr festgesetzt wird (Riehl 2014: 91). Sie gelten damit als Sprecherinnen und Sprecher mit sukzessivem kindlichen (und schulischen) Zweitspracherwerb (cL2 Sprecher). Sie haben in beiden Sprachen (Deutsch und Polnisch) kommunikative Kompetenzen, wobei sich diese in verschiedenen Domänen unterscheiden können.

„Generation Deutschland“ (GD) umfasst Probandinnen und Probanden, die bereits nach dem Krieg in Polen geboren sind. Diese Personen sind Nachfahren der „Generation Polen“ insofern, als dass mindestens ein Elternteil dieser Personen deutsch-polnisch bilingual ist und aus den Gebieten stammt, welche seit 1945 zu Polen gehören. Eine Verwandtschaft zur im Projekt untersuchten „Generation Polen“ muss nicht gegeben sein. Im Gegensatz zur „Generation Polen“ ist die L1 der Probandinnen und Probanden aus der „Generation Deutschland“ Polnisch. Obwohl die Eltern Deutsch können, erwerben die Probandinnen und Probanden der „Generation Deutschland“ das Deutsche erst nach der Auswanderung nach Deutschland, die zumeist zwischen 1960 und 1990 erfolgt. Zum Zeitpunkt des Zweitspracherwerbs sind die „Probandinnen und Probanden“ bereits erwachsen, sodass sie als Sprecherinnen und Sprecher mit Zweitspracherwerb im Erwachsenenalter (aL2 Sprecher) gelten. Die Auswanderung aus Polen liegt zum Zeitpunkt der Untersuchung mindestens 10 Jahre zurück, sodass die Probandinnen und Probanden kommunikative Kompetenzen in beiden Sprachen aufweisen können. Diese Generation lebt über das gesamte Gebiet der Bundesrepublik verstreut, weswegen keine besonderen Gebiete für die Datenerhebung ausgemacht werden können, wie bei der „Generation Polen“. Durch die Abstammung von der „Generation Polen“, die territorial bestimmten Gebieten angehört, ergibt sich hier eine Parallele zwischen den Generationen. Dank der deutschen Abstammung durften Probandinnen und Probanden der „Generation Deutschland“ die deutsche Staatsbürgerschaft erlangen und weisen damit auch Parallelen zu den Spätaussiedlerinnen und Spätaussiedlern im deutsch-tschechischen Teilprojekt auf.

Der Spracherwerb und die Sprachkompetenz, die Dominanz sowie die Distribution von Sprachen auf verschiedene Funktionsbereiche und Domänen oder die Intensität des Sprachkontakts sind dann Parameter bei der Herausarbeitung von Sprachbiographien. Die Validität der erhobenen soziolinguistischen Metadaten in Bezug auf die Sprachbiographien und ihre Erhärtung zu prototypischen Sprachbiographien der „Generation Polen“ und der „Generation Deutschland“ wird im Kapitel zur soziolinguistischen Annotation (vgl. Jańczak, Nekula & Zielińska in diesem Band) diskutiert.

Dabei orientieren sich die Auswahl der Probandinnen und Probanden sowie die Rekonstruktion ihrer Sprachregime an historischen Eckpfeilern. Auch daher ergeben sich zwischen den Generationen nicht viele Schnittpunkte, was eine Vergleichbarkeit beider Generationen erschwert. Gemeinsamer Eckpfeiler hingegen ist die Zweisprachigkeit so-

wie eine territoriale Zugehörigkeit zu den Gebieten, die einst deutsch waren und nach 1945 an Polen fielen. So haben beide Generationen gemeinsam, dass es sich um deutsch-polnisch bilinguale Probandinnen und Probanden handelt, die kommunikative Kompetenzen in beiden Sprachen aufweisen. Die Erwerbsreihenfolge und das Erwerbsalter unterscheiden sich jedoch grundlegend.

Aufgrund historischer Gegebenheiten ist es nicht möglich, zwei unterschiedliche Generationen mit gleichen linguistischen Parametern zu finden. Ziel des Projekts ist jedoch nicht eine genaue vergleichende Studie, sondern die Beantwortung der Frage, welche Unterschiede und Gemeinsamkeiten es hinsichtlich des Sprachgebrauchs und -kontakts zwischen zwei historisch geprägten Gruppen gibt. Während also die Unterschiede bei der Interpretation berücksichtigt werden müssen, können zumindest Hypothesen in Bezug auf verschiedene linguistische Parameter und deren Einfluss auf Sprachkontakt generiert werden.

Literatur

- Centner, Carolin (in Vorb.): *Lexikalische Replikation bei deutsch-polnisch Bilingualen in zwei Generationen*. Inauguraldissertation zur Erlangung des Doktorgrades Universität Regensburg.
- Clyne, Michael (2003): *Dynamics of a Language Contact. English and Immigrant Languages*. Cambridge: Cambridge University Press.
- Gerstenberg, Annette (2015): Absolute, relationale und historische Generationsbegriffe in der Sprachwissenschaft: Perspektiven ihrer Verwendung. In: Eva Neuland (Hg.), *Sprache der Generationen*. Frankfurt/Main: Peter Lang, 49–64.
- Kaluza, Andrzej (2002): Zuwanderer aus Polen in Deutschland. *Utopie kreativ*, 699–709.
- Lasatowicz, Maria/Weger, Tobias (2008): Polen. In: Ludwig Eichinger, Albrecht Plewnia, Claudia M. Riehl (Hgg.), *Handbuch der deutschen Sprachminderheiten in Mittel- und Osteuropa*. Tübingen: Narr, 145–170.
- Mannheim, Karl (1928): Das Problem der Generationen. *Kölner Vierteljahrshefte für Soziologie* 7, 157–185, 309–330.
- Nekula, Marek (2021): Varieties of German in the Czech Republic. In: *Wiener Slavistisches Jahrbuch* N.F. 9, 69–103.
- Neuland, Eva (Hg.) (2015): *Sprache der Generationen*. Frankfurt/Main: Peter Lang.
- Pauwels, Anne (2016): *Language Maintenance and Shift*. Cambridge: Cambridge University Press.
- Riehl, Claudia Maria (2014): *Mehrsprachigkeit*. Darmstadt: WBG.
- Schrader, Achim/Nikles, Bruno/Griese, Hartmut M. (1976): *Die Zweite Generation. Sozialisation und Akkulturation ausländischer Kinder in der Bundesrepublik*. Kronberg: Athenäum.
- Tišerová, Pavla (2008): Tschechien. In: Ludwig Eichinger, Albrecht Plewnia, Claudia M. Riehl (Hgg.), *Handbuch der deutschen Sprachminderheiten in Mittel- und Osteuropa*. Tübingen: Narr, 171–242.
- Weinreich, Uriel/Labov, William/Herzog, Marvin (1968): Empirical foundations for a theory of language change. In: Winfried P. Lehmann, Malkiel Yakov (Hgg.), *Directions for Historical Linguistics: A Symposium*. Austin: University of Texas Press, 95–195.

Sprachbiographische Annotation im deutsch-polnischen Korpus

Barbara Alicja Jańczak – Marek Nekula – Anna Zielińska

Abstract:

Annotation of language biographies in the German-Polish corpus

This chapter focuses on the sociolinguistic dimension of our research. Discussed will be the categorization and annotation process of sociolinguistically relevant metadata, which are an integral part of the digital corpus of German-Polish bilinguals. The aim of the corpus is on the one hand the description of dialectal features of the German dialects in the “Generation Poland” and, on the other hand, the correlation between the two language regimes (in Poland dominated by Polish and in Germany dominated by German) with the social variables involved and the morphosyntactic changes initiated by language contact. The annotation of the linguistic biographical interviews enables the elaboration of the linguistic biographical profiles that might serve to generalize. In this chapter the authors discuss the methodology and cognitive goals of linguistic biographical research. Further, a systematic categorization of sociolinguistic annotation labels will be presented. Finally, some problems with the sociolinguistic annotation in the German-Polish corpus are discussed and the first results on the way to the generalization of prototypical linguistic biographies in “Generation Poland” and “Generation Germany” are introduced.

1 Einführung

Soziolinguistisch relevante Metadaten sind ein wertvoller Teil digitaler Korpora im Allgemeinen. Da die digital aufbereiteten Texte unterschiedlicher (regionaler) Provenienz und unterschiedlicher Genres in den allgemeinen Korpora in einem ausgewogenen Verhältnis stehen sollen, werden ihre Metadaten entsprechend dokumentiert und bei der Auswertung von sprachlichen Daten genutzt. So lässt sich etwa die regional oder genrespezifisch bedingte Präferenz in Bezug auf Variation von sprachlichen Mitteln aufzeigen. An die allgemeinen oral und skriptural basierten Korpora werden – wie dies etwa im Tschechischen Nationalkorpus der Fall ist – spezialisierte Subkorpora angeschlossen. Dazu gehören u. a. auch das Akquisitions- oder das Briefkorpus, die den Sprachgebrauch in einem spezifischen Kontext oder Genre für die Forschung verfügbar machen.

Das Korpus von Sprachbiographien deutsch-polnischer Bilingualer gehört zu solchen spezialisierten Sprachkorpora. Es ist im Rahmen des Projektes „Language across Generations: Contact Induced Change in Morphosyntax in German-Polish Bilingual Speech“ entstanden und enthält digital aufbereitete Transkripte von Aufnahmen sprachbiographischer

Interviews mit deutsch-polnischen Bilingualen, die zwei Generationen zugeordnet werden können. Die Bezeichnung „Generation“ wird in dem Projekt gebraucht, um die Differenz einerseits in Bezug auf den Zeitraum der Geburt und andererseits in Bezug auf die Sprachregime, in denen beide Gruppen von Interviewten sozialisiert wurden und leben, deutlich zu machen (zum Generationsbegriff siehe Centner in diesem Band).¹ In diesem Sinne werden sie als „Generation Polen“ und „Generation Deutschland“ bezeichnet. Während also die deutsch-polnischen Bilingualen in der „Generation Polen“ Deutsch in ehemals dominant deutschsprachigen Regionen als ihre Erstsprache und Polnisch nach dem Zweiten Weltkrieg als ihre Zweitsprache erworben haben, war es in der „Generation Deutschland“ gerade umgekehrt. Diese Gruppe stammt zwar aus Familien mit deutscher Herkunft, die Interviewten sind jedoch in ein dominant polnisches Sprachregime hineingeboren und erst nach der Emigration nach Deutschland haben sie Deutsch erworben bzw. ihre Vorkenntnisse ausgebaut. Das Sample wird im Allgemeinen in Centner sowie in Hansen & Nekula und Bučková & Nekula, die Aufnahmen in Jorroch & Prawdzic bzw. Bučková & Nekula, die technische Bearbeitung, Transkription und Speicherung von Aufnahmen und ihren Transkripten in Bučková & Centner, die formorientierte Annotation in Bučková, Centner, Książek & Prawdzic und die finale technische Aufstellung des Korpus in Woźniak jeweils in diesem Band detailliert dargestellt.

Das Ziel des deutsch-polnischen Korpus ist einerseits die Dokumentation der deutschen Dialekte in der „Generation Polen“, andererseits geht es um die Korrelation zwischen den beiden Sprachregimen mit den darin wirksamen sozialen Variablen und dem durch Sprachkontakt initiierten Sprachwandel im Bereich der Morphosyntax. Dabei stellt sich die Frage, ob und wie die Sprachkontaktphänomene und ihre Typen sich unterscheiden, die im Polnischen in der „Generation Polen“ und in der „Generation Deutschland“ bzw. im Deutschen in der „Generation Deutschland“ und in der „Generation Polen“ festzustellen sind. Diesem Ziel dient dann die Annotation der sprachbiographischen Interviews, auf deren Grundlage die sprachbiographischen Profile und deren Generalisierung erarbeitet werden. Über das Projekt hinaus dient die sprachbiographische Annotation der Spracherwerbs- und Sprachkontaktforschung im Allgemeinen sowie der auf die untersuchten Regionen bezogenen historischen Soziolinguistik im Besonderen.

Daher gehen wir hier auf die sprachbiographische Annotation näher ein. Um diese angemessen darzustellen, werden die in einem spezialisierten Korpus verfügbaren Transkripte von sprachbiographischen Interviews zunächst in die sprachbiographische Forschung eingeordnet und ihre methodischen Ausgangspunkte und Erkenntnisziele kurz skizziert. Im zweiten Schritt werden die Kategorien der sprachbiographischen Annotation systematisch dargestellt, die auch in anderen Kontexten genutzt werden bzw. genutzt werden können. Anschließend werden Probleme bei der sprachbiographischen Annotation im deutsch-polnischen Korpus diskutiert und die ersten Ergebnisse auf dem Weg zur Generalisierung von prototypischen Sprachbiographien in der „Generation Polen“ und in der „Generation Deutschland“ präsentiert.

¹ Im Sinne von Kroskrity (2000) oder Coulmas (2005) verstehen wir unter Sprachregime eine durch Sprachverordnungen und Sprachideologien etablierte Sprachsituation, in der die Sprachen im Sinne einer sozialen Hierarchie (vertikal) und/oder einer territorialen Distribution (horizontal) festgelegt sind.

2 Sprachbiographie(n) – ein theoretischer Rahmen

Eine Sprachbiographie, d. h. die Verflechtung von Lebensgeschichte und Sprache(n), kann man auf unterschiedliche Art und Weise greifbar machen: durch Tagebücher, die den kindlichen oder eigenen Spracherwerb und -gebrauch begleiten, sowie durch sprach(en)-fokussierte biographische Narrationen in literarischen Autobiographien und Erinnerungen oder in Tagebüchern und Briefen (vgl. u. a. Nekula & Rychnovská 2016). Auch werden etwa Sprachbiographien von Kindern, deren Mehrsprachigkeit durch Testverfahren erfasst wird, durch die (halb)strukturierte Befragung ihrer Eltern gewonnen (Anstatt 2009; Brehmer & Mehlhorn 2015). Busch (2017) fokussiert bei Kindern und klinischen Patienten die Methode der Sprachenportraits, an die dann die Narrationen anschließen. Dabei versteht sie das Konzept der Sprachbiographie offen, d. h. mit Blick auf den Erwerb und die Verwendung von Varietäten einer Sprache und/oder unterschiedlicher Sprachen, wobei die äußere Mehrsprachigkeit nicht selten die innere Mehrsprachigkeit einschränkt. Von Głuszkowski (2011) werden Sprachbiographien auf die sprachenfokussierte Lebensgeschichte mehrsprachiger Personen beschränkt.

Bei den Sprachbiographien, auf die wir im Rahmen unseres Projekts zurückgreifen, orientieren wir uns an der biographischen Forschung, die seit den 1990er Jahren im Zuge des „narrative turn“ (Scheffel 2012) auch in die Sprachwissenschaft vordringt. Diese sprachbiographische Forschung versteht unter „Sprachbiographien“ sowohl eine *qualitative Methode*, die sich der freien narrativen Interviews bedient, als auch eine *Datenquelle* (Nekvapil 2004a). In diesen Interviews werden bei der lebensgeschichtlichen Erzählung Sprachen und ihre Varietäten fokussiert. Die dabei gewonnenen *metasprachlichen Daten* über Spracherwerbsstrategien oder Gebrauchskontexte von Varietäten und Sprachen seitens der interviewten Sprachbenutzerinnen und -benutzer werden dann im Rahmen der Spracherwerbsforschung oder bei der soziolinguistischen Rekonstruktion der Sprachsituation und deren Wandel genutzt. Eine besondere Bedeutung hat diese Methode bei der Rekonstruktion von Sprachsituationen und deren Wandel in „randständigen Bereichen“ (Franceschini 2001), in denen vermittels der quantitativen Methoden wie Zensus kaum zuverlässige Daten über Spracherwerb, -gebrauch, -erhalt und -verlust sowie die subjektive Einschätzung dieser Daten durch die Akteure zu erhalten sind. Zu solchen randständigen Bereichen gehört eben auch die sprachlich prekäre Situation autochthoner und allochthoner Minderheiten (Werlen 2002; Franceschini & Miecznikowski 2004; Treichel 2004 u. a.), die durch solche Interviews in ihrem Wandel im Laufe von bis zu drei Generationen nachvollziehbar gemacht werden kann.

Die Nutzung der dabei gewonnenen *sprachlichen Daten* spielte dagegen in dieser Ausprägung der sprachbiographischen Forschung interessanterweise bisher eher eine kleine Rolle. Davon setzt sich unser Projekt ausdrücklich ab, indem die narrativen sprachbiographischen Interviews mit erwachsenen Respondentinnen und Respondenten nicht nur als Quelle der metasprachlichen Daten, sondern auch als Quelle der sprachlichen Daten dienen. Damit schließen wir an die Mehrsprachigkeitsforschung an, in der die *mündliche Narration* als Textgenre genutzt wird, in dem die kognitive Kontrolle der Sprachbenutzerinnen und -benutzer gemindert ist und die Sprachkontaktphänomene demnach verstärkt auftreten (vgl. Polinsky 2006; Montrul 2008; Benmamoun, Montrul & Polinsky 2013). Die in unseren Aufnahmen auftretenden Sprachkontaktphänomene werden zudem in unserem digital aufbereiteten Korpus formorientiert annotiert. Darüber hinaus nutzen wir

aber auch die inhaltliche Ausrichtung der narrativen Interviews auf die Sprachbiographie. So werden im deutsch-polnischen Korpus neben den sprachlichen Daten auch die metasprachlichen Daten annotiert. Beide Datentypen werden jeweils unabhängig voneinander ausgewertet, analysiert und anschließend miteinander in Beziehung gesetzt.

Durch die Überführung der Interviews in ein Korpus sowie durch ihre Annotation, die quantitative Auswertung und die Korrelation der Ergebnisse dieser Analysen steigt auch die Relevanz und Objektivität der qualitativ gewonnenen Daten. Die Frage der *Validität* der qualitativen Forschung wird dabei immer wieder aufgeworfen. Bei den sprachlichen Daten aus der freien Narration ist sie – anders als beim Experiment oder Testen – per se gegeben. Ihre Authentizität argumentieren etwa Polinsky (2006) oder Montrul (2008) damit, dass die kognitive Kontrolle über die sprachliche Form der Narration im Falle eines Themas, das die Interviewten persönlich einnimmt, wie es bei einem sprachbiographischen Interview der Fall ist, nur begrenzt ist. Daher kann man davon ausgehen, dass die Sprache der sprachbiographischen Interviews der Sprache in der Alltagsinteraktion nahekommt. Die Relevanz von metasprachlichen Daten wurde oben angesprochen. Die Tatsache, dass die Sprachbiographien annotiert wurden, erlaubt die Anwendung gemischter (qualitativ-quantitativer) Methoden.

In Bezug auf die erzählte Welt ist ihre Validität u. a. auch dadurch gegeben, dass man nicht prominente, sondern Durchschnittsprecherinnen und -sprecher auswählt und somit ihr Erleben der Sprachsituation und den darin verankerten Spracherwerb, -gebrauch, -erhalt und -verlust fokussiert. Die Validität ihrer Aussagen ergibt sich zudem auch daraus, dass die Durchschnittsprecherinnen und -sprecher untereinander vernetzt waren und sind und ihre Narrationen auch teilen, sodass sie das eigene Erleben mit dem der Anderen vergleichen können und u. a. auch vermittels der „extreme case formulations“ (Pomerantz 1986) wie „immer“, „stets“, „alle“ u. a. m., generalisieren:

(1) meine Mutter war Deutsche der Vater auch / meine Mutter hat *immer* Deutsch gesprochen / so wie ich schon habe gesagt / *immer* mit die Kinder sie sprach nur Deutsch meine Mutter (EU_KET_I_GP_DE)²

Daran kann auch die Forschung anschließen, indem sie Einzelbiographien vergleicht und diese in einer „prototypischen Sprachbiographie“ (Hašová 2004) verdichtet, wie dies im dritten Abschnitt aufgezeigt wird. Die Zuverlässigkeit der in den narrativen Interviews erhobenen metasprachlichen Daten stellt Nekvapil (2004a) dadurch unter Beweis, dass er – ungeachtet der wechselnden Ethnizität der Interviewerinnen und Interviewer, der wechselnden Forschungsfragen sowie des zeitlichen Abstands zwischen den Erhebungen – ihre Stabilität nachweist. Auch in den im deutsch-polnischen Teilprojekt durchgeführten Interviews, die durch ein deutsch-polnisches Tandem geführt wurden, schien es in Bezug auf die Ethnizität der Fragenden keine Differenz in den Inhalten zu geben.

In Bezug auf die gewonnenen metasprachlichen Daten muss man sich allerdings stets dessen bewusst werden, dass die in den sprachbiographischen Interviews enthaltenen metasprachlichen Daten nicht „die“ Realität sind, sondern dass man diese lediglich als

² Die Segmentierung des Transkripts in ELAN ist im Text mit einer vertikalen Linie markiert worden.

ihre narrative Repräsentation verstehen und diese in Lebens-, Subjekt- und Textrealität ausdifferenzieren muss (Nekvapil 2004a; Pavlenko 2007). Die Aussagen über die „fakto-graphische“ *Lebensrealität* des Spracherwerbs, -gebrauchs, -erhalts, -wechsels und -verlusts sind dabei nicht naiv als gegeben zu nehmen, sondern durch den Vergleich mit metasprachlichen Daten aus anderen sprachbiographischen Interviews, wie die Angaben zum Verbot des Deutschen in der Öffentlichkeit in (2), (3) und (4), zu validieren oder durch andere Methoden, wie die Auswertung von Dokumenten (Sprachverordnungen, Zeugnissen u. a.) oder Notizen auf der Basis der teilnehmenden Beobachtung während der Feldforschung (Meng 2004) zu triangulieren:

(2) po niemiecku babcia z nami nie rozmawiała/ ponieważ .. @ po* z tego co zawsze mówiła po wojnie było zabronione płaciła kary/ więc później już się tak nauczyła po śląsku/ że rozmawiała z nami po śląsku/ babcia znała dobrze niemiecki (KN_LAM_GD_PL)

Die Großmutter sprach nicht Deutsch mit uns / weil.. @ po* da sie immer sagte nach dem Krieg war es verboten, man musste Strafe zahlen / also lernte sie schon später Schlesisch / so dass sie mit uns Schlesisch sprach / Großmutter konnte gut Deutsch

(3) (alles) mit mit mit elf war hier ich im Gericht auch / das sollten wir unterschreiben/ meine Mutter als hätte deutsche Sprache nicht nutzen werden/

Interv.: (wann war das?)

/drei drei Tage drei Nächte/ und ich n * habe wir haben nich * ich habe nicht unterschrieben meine Mutter auch nicht (CS_GIZ_GP_DE)

(4) ja siebenundvierzig Jahre konnten wir hier nicht ausüben die Sprache nicht/

Interv.: (in der Schule auch nicht?)

/ in der Schule nein es gab nicht in der Schule/

Interv.: (die Kinder zwischen so untereinander durften auch kein Deutsch)

/ <@> nein sondern waren immer welche/ die ver * verpetzt hatten haben dann verpetzt/ und denn denn hat man Konsequenzen gehabt (CS_GIZ_GP_DE)

Bei der Rekonstruktion der sprachlichen Entwicklung von zweisprachigen Personen auf der Basis von autobiographischen Interviews muss man allerdings dieselben Beschränkungen und Hindernisse wie bei allen Untersuchungen zum autobiographischen Gedächtnis berücksichtigen. Die im Interview vermittelte Lebensrealität kann vor allem angesichts des fortgeschrittenen Alters der Informantinnen und Informanten der „Generation Polen“ durch Gedächtnislücken, durch den Wunsch sich im positiven Licht darzustellen sowie durch die Auswahl der erzählten Episoden und andere Faktoren beeinträchtigt werden (Maruszewski 2005).

Die *Subjektrealität* offenbart sich in den Interviews einerseits in der Versprachlichung von „Vorstellungen über die Varietäten und Sprachen“ (Silverstein 1979; Woolard 1998) und deren Erwerb, Gebrauch, Erhalt und Verlust, andererseits in der Versprachlichung des subjektiven Erlebens von Varietäten und Sprachen und ihres Erwerbs, Gebrauchs, Erhalts, Verbots oder Verlusts. Beides, die *Sprachideologien* in Bezug auf das sprachliche Repertoire und das *Spracherleben* des sprachlichen Repertoires (Busch 2017: 18–19), ist dabei durch die Wertung eng miteinander verbunden.

Das von John Gumperz (1964: 137–138) geprägte statische Konzept des *sprachlichen Repertoires*, wonach die Sprecherinnen und Sprecher aus dem raumbezogen verstandenen kollektiven Repertoire „weapons of everyday communication“ auswählen, reicht hier allerdings nicht aus. Daher orientieren wir uns an dem individuell und zeitlich verstandenen Konzept des Repertoires, wie es Blommaert (2008: 17) formuliert: „The fact is, however, that someone’s linguistic repertoire reflects a *life*, and not just birth, and it is a life that is lived in real sociocultural, historical and political space.“ Das individuell verstandene sprachliche Repertoire verändert sich demnach im Laufe des Lebens: Die Erstsprache kann zur sekundären und die Zweitsprache zur primären Sprache werden.

Das Repertoire wird jedoch in der Interaktion nicht nur automatisiert eingesetzt, sondern auch – in seinem Wandel – bewusst reflektiert und geplant (um)gestaltet. In diesem Sinne kann man die mehrsprachigen Sprachbiographien durch das „Verhalten gegenüber der Sprache“ (Neustupný 1978; Jernudd & Neustupný 1987) als Ergebnis des individuell sowie in den Familien *organisierten Sprachmanagements* verstehen, das auf der *Mikroebene* (mit Blick auf künftige sprachliche Interaktionen der Akteure in den Umgebungssprache(n) in einem weiteren sozialen Rahmen) auf das organisierte Sprachmanagement – den gesteuerten Umgang mit den Sprachen – auf der (staatlichen) *Makroebene* reagiert (Nekvapil 2009; Nekvapil & Sherman 2015). Das Sprachmanagement auf beiden Ebenen wird dabei jeweils durch eine *Abweichung von einer sprachlichen Norm* ausgelöst, wobei zu solchen Normen neben *sprachlichen* Normen im engeren Sinne auch die *kommunikativen* und *soziokulturellen* Normen zählen. So hatte der offizielle Wechsel vom deutschen zum polnischen Sprachregime, zu dem es in den 1945 an Polen angeschlossenen Gebieten in der Nachkriegszeit auf der Makroebene kam, zur Folge, dass viele der dort verbliebenen Bürger des Dritten Reiches, zu denen auch die Eltern der Respondentinnen und Respondenten der „Generation Polen“ zählen, auf die neuen kommunikativen Normen im öffentlichen Raum reagiert und selbst sowie bei der Erziehung der Kinder graduell zum Polnischen gewechselt hatten, wie die Eltern einer Informantin aus Piła, die in den Kursen zuerst Polnisch lernten und es dann zu Hause benutzten:

(5) Interv.: (also, und die Eltern haben Deutsch gesprochen?)

/ja sie haben erst @ fünfundvierzig ein @ so Kursus gemacht/ wo sie polnisch sie konnten nicht polnisch sprechen/und da haben sie denn jelernt nicht/no so perfekt konnten sie nie na polnisch (NH_PIL_GP_DE)

(6) Interv.: (Haben Sie mit Ihren Eltern deutsch gesprochen?)

/@ ja zu Anfang deutsch und nachher gemischt deutsch polnisch alles alles zusammen (NH_PIL_GP_DE)

Andernfalls würden sie durch die Absenz des Polnischen oder das „schlechte“ Polnisch kommunikativ auffallen und sich so auch ihre soziale Akzeptanz und ihren sozialen Aufstieg verbauen. Selbst wenn sie der Überzeugung waren, dass Deutsch ihre „Muttersprache“ ist, die ihre Identität definieren soll, und daher zu erhalten und an die nächste Generation weiterzugeben wäre, hatten sich viele von ihnen die Sprachideologien der (polnischen) Majorität zu eigen gemacht, dass Deutsch angesichts des polnischen Sprachregimes überholt und unbedeutend sei und daher nicht erworben werden müsse, während

Polnisch, das im öffentlichen Raum aktiv und richtig zu nutzen war, beim Erwerb und Gebrauch – selbst in der Familie – zu präferieren sei:

(7) no w każdym bądź razie .. ta nauka języka polskiego @ szła bardzo szybko/jako dzieci @ na pewno się szybciej ucza/ poza tym .. myśmy się wstydzieli być Niemcami/bo nas wytykano @/i i i wyzywano/i my jako dzieci chcieliśmy koniecznie się nauczyć szybko tego języka/i @ i nam się chyba to udało/gorzej było z naszymi rodzicami/więc w domu mówiło się jeszcze po niemiecku/chociaż w międzyczasie to rodzice też się nauczyli języka polskiego ale dużo dużo później/ (IT_ELB_GP_PL)

na, jedenfalls .. der Polnischunterricht @ ging sehr schnell / da die Kinder @ sicher schneller lernen / außerdem .. schämten wir uns Deutsche zu sein / denn man zeigte mit dem Finger auf uns / und und und beschimpfte uns / und wir als Kinder wollten natürlich schnell diese Sprache lernen / und @ und uns ist es wohl gelungen / schlimmer war es mit unseren Eltern / also zu Hause wurde noch Deutsch gesprochen / obwohl die Eltern inzwischen auch Polnisch gelernt hatten, aber viel später/

Die Sprachideologien, die an der Schnittstelle der sozialen und sprachlichen Normen zu verorten und Ausdruck von Gruppeninteressen sind (Woolard 1998; Kroskirty 2000), begründeten somit rational – mit dem Argument des Nutzens oder der Identität – das jeweilige Sprachmanagement, das bei den Minderheiten mit dem Argument des Nutzens, der sozialen Akzeptanz und des sozialen Aufstiegs auf den Erwerb der dominanten *Umgebungssprache* abzielt. Das organisierte Sprachmanagement, das – wie gezeigt werden konnte – auf Sprachideologien basiert (vgl. auch Nekula 2021) bzw. diskursiv, durch einen sprachideologischen Diskurs initiiert wird, wird in den sprachbiographischen Interviews seitens der Interviewten retrospektiv (sowie teils auch prospektiv) durch „language management summaries“ (Nekvapil 2004b) mit den Forscherinnen und Forschern geteilt, durch die man Einblick in die Strategien des Spracherwerbs der Umgebungssprache und des Spracherhalts der Minderheitensprachen bekommt.

Im folgenden Abschnitt wird das deutsch-polnische Korpus kurz vorgestellt. Auf metasprachliche Daten ausgewählter Interviews gestützt, werden anschließend prototypische Sprachbiographien der „Generation Polen“ und der „Generation Deutschland“ stichwortartig umrissen, und zwar im Hinblick darauf, was für die sprachbiographische Annotation der Transkripte sprachbiographischer Aufnahmen und für die Rekonstruktion von Sprachbiographien relevant ist.

3 Prototypische Sprachbiographien im deutsch-polnischen Korpus

In das Korpus wurden 30 Respondentinnen und Respondenten der „Generation Polen“ und 29 der „Generation Deutschland“ aufgenommen.³ Informantinnen und Informanten der „Generation Polen“ wurden zwischen 1932 und 1944 in den zu der Zeit deutschen

³ Obgleich das aufgenommene sprachbiographische Material viel umfassender ist, wurden Aufnahmen vieler Informantinnen und Informanten in das finale Korpus, üblicherweise wegen Nichterfüllung formaler Anforderungen für die Einbeziehung einer Person in eine bestimmte Generation, nicht übernommen.

Gebieten geboren. Sie stammen vor allem aus Ostpreußen (9 Personen) und Schlesien (8 Personen, die meisten davon aus Oberschlesien), einzelne Respondentinnen und Respondenten stammen aus den Provinzen Pommern (6 Personen), dem Regierungsbezirk Posen-Westpreußen oder Westpreußen. Der Altersunterschied innerhalb der „Generation Deutschland“ ist beträchtlich größer. Diese Gruppe von Informantinnen und Informanten wurde zwischen 1951 und 1998 geboren. Fast 70% der Respondentinnen und Respondenten stammen ursprünglich aus Oberschlesien, die Herkunftsgebiete der übrigen 30% verteilen sich gleichmäßig zwischen den Woiwodschaften Pommern, Niederschlesien, Lebuser Land und Großpolen.

Aufgrund der Aussagen der Respondentinnen und Respondenten ist es möglich, für die „Generation Polen“ und die „Generation Deutschland“ jeweils eine prototypische Sprachbiographie (bei der „Generation Polen“ mit der Variante des Spracherhalts und Sprachverlusts) zu entwerfen.

Wie bereits erwähnt, sind Personen, die als „Generation Polen“ eingestuft wurden, auf dem Gebiet des Deutschen Reichs geboren, bevor diese Gebiete nach dem Zweiten Weltkrieg an Polen fielen. Diese Respondentinnen und Respondenten haben Deutsch als L1, das Polnische dann vor dem 12. Lebensjahr, also wahrscheinlich noch vor Abschluss der sensitiven Periode erworben.

Ab der Jahreswende 1944/45 wurden die Deutschen und ihre Sprache auf diesem Gebiet Opfer der repressiven Politik des polnischen Staates, die den Erwartungen der Polinnen und Polen entsprach. Sie waren erfüllt vom Wunsch nach Rache und Wiedergutmachung für das im Zweiten Weltkrieg erlittene Leid. Man wollte die Deutschen internieren und aussiedeln, die deutsche Sprache sollte verschwinden (Kacprzak 2010: 215). In der Tat wurden in den ersten Nachkriegsjahren drei Millionen Deutsche aus Polen ausgesiedelt. Bleiben durften nur besonders qualifizierte Spezialistinnen und Spezialisten, sobald sie für die Industrie unentbehrlich waren. Auch Personen, die ihre polnische Abstammung belegen konnten, durften bleiben. Sie wurden allerdings der sprachlichen und kulturellen Polonisierung unterworfen (Madajczyk 1998: 66–67). In Polen betrachtete man die deutsche Sprache im Allgemeinen als etwas Feindliches. Daher wurde jeder, der Deutsch oder Polnisch mit deutschen Interferenzen sprach, als Feind betrachtet. Nicht nur die Behörden verfolgten, die Anordnungen von oben ausführend, die Deutschsprechenden, sondern auch die Leute in der Nachbarschaft, die Lehrerinnen und Lehrer sowie die Schulkinder taten das mit großem Eifer. Auch in den diskutierten Interviews, die sich zu einer prototypischen Sprachbiographie zusammenfügen lassen, ist von physischer Gewalt, Verspottung und Verfolgung deutschsprechender Personen die Rede, wie dies in den Aussagen (2), (3) und (4) sichtbar wird.

In (5) und (6) wird deutlich, dass der abrupte Sprachwechsel im öffentlichen Raum und der Erwerb der polnischen Sprache im Allgemeinen herausfordernd waren, auch wenn sie im Einzelnen in den verschiedenen Gebieten und Familien unterschiedlich verliefen. Mit dem Einstieg ins Berufsleben und ökonomischen Verselbstständigung der Respondentinnen und Respondenten stieg die Relevanz des Polnischen weiter. Im nächsten Lebensabschnitt – nach der Gründung eigener Familien – erfolgte der Sprachwechsel (*language shift*) zum Polnischen bei unseren Informantinnen und Informanten auch im Privaten. Dies trifft selbst dann zu, wenn die Ehepartnerin oder der Ehepartner aus einer deutschsprachigen Familie stammte. So wurde das Deutsche in der Regel nicht an die Kinder weitergegeben. Ausnahmen bilden die homogamen Familien aus Schlesien, in

denen zu Hause Deutsch gesprochen und an die Kinder weitergegeben wurde. Bei den Interviewten ist die Kompetenz in beiden Sprachen kommunikativ, obwohl die Interaktion in Deutsch auf wenige Domänen beschränkt ist.

Wie im Abschnitt 1 erwähnt wurde, ist die „Generation Deutschland“ in Polen geboren. Die Interviewten aus dieser Generation haben im Elternhaus kein (oder nur sehr wenig) Deutsch gesprochen. Ihre L1 ist Polnisch, das Deutsche wurde erst in Deutschland erworben. Die Emigration und der Erwerb des Deutschen erfolgten erst nach der sensitiven Periode, sodass Polnisch als Erstsprache auch angemessen stabilisiert war. Sowohl zu Hause als auch in ihrem sozialen Umfeld wird Deutsch und Polnisch gesprochen, wobei sich die Domänen der Verwendung unterscheiden können. Die Aufenthaltsdauer in Deutschland beträgt mindestens 10 Jahre.

Damit handelt es sich hier um Personen aus den 1945 an Polen angeschlossenen Gebieten, die im Deutschen als Aussiedler bzw. Spätaussiedler gelten:

„The term *Aussiedler* was coined in the early 1950s as the influx of German minorities from east-central and southeastern Europe and the Soviet Union continued after the flight and expulsion during the immediate postwar years had come to an end. The Constitution of the Federal Republic of Germany (Article 116, paragraph 1) guarantees the admission of these persons as German citizens: ‚Unless otherwise provided by a law, a German within the meaning of this Constitution is a person who possesses German citizenship or who has been admitted to the territory of the German Reich within the boundaries of December 31, 1937 as a refugee or expellee of German ethnic origin or as the spouse or descendant of such person.‘ Decisive for admission as *Aussiedler* are the ‚German ethnicity‘ of the immigrants and the ‚expulsion pressure‘ in their count[r]y of origin. These criteria were formulated shortly after World War II, when millions of Germans fled or were expelled from territories east of the Oder and Neisse rivers.“ (Dietz 2011: 245)

Die Aussiedlerinnen und Aussiedler unterschieden sich von den anderen Immigrantinnen und Immigranten durch die Gründe ihrer Ausreise: Neben politischen und ökonomischen Gründen spielte in vielen Fällen die deutsche Selbstidentifikation eine Rolle. Nach der Aussiedlung erhielten sie in Deutschland die deutsche Staatsbürgerschaft, umfassende soziale Unterstützung vom Staat sowie kostenlose Sprachkurse. Ihr Wunsch war es, als „Germans among Germans“ zu leben (Dietz 2011: 248). Die Weitergabe der deutschen Sprache wurde allerdings in den meisten Familien so nachhaltig unterbrochen, dass die Angehörigen der „Generation Deutschland“ in der neuen Heimat die deutsche Sprache erst von Grund auf in Sprachkursen erwerben mussten und manchmal keine hohe Sprachkompetenz mehr erreichen konnten. Das trifft vor allem für die Personen zu, welche außer dem Sprachkurs keine weitere Ausbildung in Deutschland anstrebten, und für diejenigen, die sich Ehepartnerinnen und Ehepartner in der Gruppe der Spätaussiedlerinnen und -aussiedler suchten und zu Hause hauptsächlich Polnisch oder Schlesisch sprachen. Mangelnde Sprachkenntnisse waren anfangs der Grund für viele Probleme und Frustrationen in Deutschland. In den Ehen, wo beide Partner aus Polen kamen, geschah die Wahl der Sprache spontan: Man sprach in unterschiedlichen Situationen Polnisch und Deutsch, aber für die Erziehung der Kinder wählte man bewusst Deutsch. Diese Wahl wurde unabhängig von der Sprachkompetenz der Eltern getroffen. Selbst wo die Kenntnisse des Deutschen schlecht waren, wurde es dennoch mit den Kindern gesprochen. Eine Ausnahme bilden Familien, die sich für eine zweisprachige Erziehung entschieden haben.

4 Einordnung von Sprachbiographien durch Annotation

Eine individuelle „Sprachbiographie“, die von einem narrativen sprachbiographischen Interview abgeleitet ist, ist bekanntlich ein wissenschaftlich basiertes Konstrukt. Beim Prozess der Verdichtung von metasprachlichen Daten in der transkribierten Aufnahme kommt es zur Auslassung von dafür irrelevanten Interviewabschnitten oder sich wiederholenden metasprachlichen Daten, zur Selektion der in Bezug auf Lebensabschnitte und Sprachdomänen markanten Daten sowie zu deren Generalisierung und Integration in einem chronologisch geordneten „Lebenslauf“ (vgl. auch Nekvapil 2004a). So geordnete „Lebensläufe“ kann man untereinander vergleichen, um die gewonnenen metasprachlichen Daten zu korrelieren und zu validieren, aber auch um diese „Lebensläufe“ für die „Generation Polen“ bzw. für die „Generation Deutschland“ in einer prototypischen Sprachbiographie zusammenzuführen (zur mitteleuropäischen Sprachbiographie der in der Tschechoslowakei verbliebenen Deutschen vgl. Hašová 2004).

Um den Prozess der Verdichtung von sprachbiographisch relevanten metasprachlichen Daten in einer individuellen Sprachbiographie nachvollziehbar zu machen und die Generalisierung der darin integrierten metasprachlichen Daten zu objektivieren sowie durch schnelle Suche nach soziolinguistischen Daten den effizienten Vergleich von Sprachbiographien und ihren Aspekten nach Generationen und Regionen zu ermöglichen, schien die sprachbiographische Annotation der Transkripte von sprachbiographischen Interviews im deutsch-polnischen Korpus geboten zu sein und wurde entsprechend umgesetzt. Die sprachbiographisch relevante Stelle wird im jeweiligen Transkript im digitalen Korpus als Block markiert und in einer speziellen Spur mithilfe der weiter unten angegebenen Kürzel annotiert. Beim Vergleich mit Ergebnissen, die auf der Annotation von semantischen und syntaktischen Musterentlehnungen basieren, scheint es möglich zu sein, den Auswirkungen von soziolinguistischen Faktoren auf den Sprachgebrauch und damit auch auf die Frequenz und Typen von Musterentlehnungen näher zu kommen.

Bei der soziolinguistischen Rekonstruktion der individuellen Sprachbiographie in Bezug auf die sog. *Lebensrealität* sowie beim Vergleich von Sprachbiographien, der mit Blick auf die Lebensrealität auch mit externen Daten wie Sprachverordnungen trianguliert werden kann, ist die Annotation von Bedeutung, die die lebensgeschichtlichen Entwicklungsphasen mit Spracherwerb und -erhalt und die Domänen mit dem Sprachgebrauch verknüpft. Diese Verknüpfung ist für die Einordnung des Spracherwerbs und des Typs der Mehrsprachigkeit, die Einschätzung des Spracherhalts oder des Sprachwechsels (language shift) sowie die Einordnung der Sprachkontaktsituation, bei der die Erstsprache zur sekundären Sprache und die Zweitsprache zur primären Sprache wird (zur Terminologie vgl. u. a. Riehl 2014), bis hin zum individuellen Sprachverlust bzw. kollektiven Sprachtod entscheidend.

Bei der Annotation der *Lebensphasen* orientieren wir uns an Meng (2004) und unterscheiden Kleinkindalter (early childhood = EC), Vorschulalter (childhood = C), Schulalter (school age = SA), Jugendalter (adolescence = ADO), Erwachsenenalter (adult age = AA) und spätes Erwachsenenalter (late adult age = LAA). Die *Domäne* verstehen wir mit Joshua Fishman (1964) als einen institutionell oder anders stabilisierten „sozialen Kontext der Interaktion“, der durch ein Bündel von Parametern (Adressat, Raum/Setting und Thema) sowie durch damit verbundene Rollen und Verhaltensregeln charakterisiert ist bzw. charakterisiert werden kann. Der soziale Raum, der zu den sozialen Parametern

zählt, ist durch die Wahl der Varietät oder Sprache mit konstituiert. Angesichts der „räumlichen“ Verortung der jeweiligen Domäne, d. h. der Familie im Haus, der Religion in der Kirche u. ä., wird allerdings die Domäne als ein zu statisches Konstrukt kritisiert. So wird etwa aufgezeigt, dass das Codeswitching im Sinne von John Gumperz (vgl. Auer 2013: 164–174) zum Kontextualisierungshinweis werden kann, der adressaten- oder themengebunden den entsprechenden sozialen Kontext der Interaktion jenseits eines institutionalisierten Raums herstellen kann. Die Kritik wird in diesem Zusammenhang auch an dem Konzept der räumlich verstandenen Institutionalisierung geübt und der Terminus „Domäne“ durch den des „Funktionsbereichs“ ersetzt (zur kritischen Diskussion des Begriffs der Domäne vgl. Werlen 2004).

Wir bleiben allerdings bei dem funktional erweiterten Konzept der Domäne und annotieren folgende Domänen, die sich in einem Sprachkontaktraum mit unterschiedlichen Sprachen verbinden: Familie/Haus (family = F), Nachbarschaft/Wohnort (neighbourhood = N), Religion/Kirche (religion = R), Freunde (friends = FR), Bildung/Schule (education = E), Arbeit (work = W), Verwaltung/Büro (administration = A), Medien/Fernsehen, Presse, Internet (media = M) u. a. Über die klassische Bestimmung der Domänen hinaus werden spezifisch auch die Domänen bzw. Funktionsbereiche der Minderheitenvereine (minority associations = MA) und des Reisens (journey = J) im Korpus annotiert. Angesichts der Relevanz der Erhebung des Sprachgebrauchs in den Ego-Dokumenten (Tagebücher, Briefe) wurden diese abgespalten und spezifisch (ego documents = ED) annotiert.

In diesem Sinne kann man (8) mit Blick auf die Lebensphase und die Domäne mit SA und N und (9) mit SA und E annotieren:

(8) wir haben doch unter uns Deutsch gesprochen/ wenn wir gespielt haben irgendwo auf der Straße/ man hat früher auf der Straße gespielt oder irgendwo bei auf dem Hof aber meistens auf der an der Straße nicht und so/ denn die wenn die mit uns spielen wollten/ müssten sie/ die ganze Kindheit die ganze Kindheit war so @ (EM_UKT_II_GP_DE)

(9) na aber nachher nach nach dem siebten Schuljahr nicht? ging ich nach Mrągowo ins Lyzeum/ ne und da na da hat man nur Polnisch gesprochen nicht (EM_UKT_II_GP_DE)

Die Annotation, die die sog. *Subjektrealität* reflektiert, bei der es um die Konzeptualisierung von Sprachen geht, erfasst das *Spracherleben* (lived experience of language = LEL) und die *Sprachideologien* (language ideologies = LI). Das Spracherleben ist, wie dies in (10) deutlich wird, mit Emotionen wie Angst, Scham oder Ärger verbunden. Die Sprachideologien als Vorstellungen über die Sprachen begründen und rationalisieren, wie dies in (11) deutlich wird, den Sprach(en)erwerb, -gebrauch, -erhalt, -wechsel oder -verlust:

(10) doch doch doch doch to też było na początku/ ten najgorsza najgorszy problem to była ta szpracha/ to było to był dla mnie to było katastrofa/ @ na początku myśmy to uczy* jak się uczyłam na kursach językowych tej Sprache/ aber @ jak poszłam do urzędu coś załatwiać @ bo niektóre wyrazy mają Doppelbedeutung/ i @ i jak jak jakbym coś źle zrozumiała to mi było so peinlich (NF_PAD_I_GD_PL)

doch, doch, doch, doch das war auch am Anfang / der schlimmste, das schlimmste Problem war die Sprache / Das war, das war für mich, das war eine Katastrophe / @ am Anfang lernten wir * wie ich auf den Sprachkursen die Sprache lernte / aber @ wenn ich zur Behörde ging etwas zu

erledigen @ denn manche Ausdrücke hatten eine Doppelbedeutung / und @ und wie, wie ich etwas falsch verstand war es mir so peinlich

(11) ja zumal in der in der in damalige Zeit herrschte @ nicht überwiegend aber schon die Meinung dass man unter Umstände Kinder in den Edukationsweise verwirren könnte durch diese Bilingualität / das sich überhaupt nicht so herausgestellt hat / interessanterweise muss ich sagen aus der Erfahrung dass wenn die am Anfang zweisprachig aufwachsen dann irgendwie haben sie so bisschen in der Tat so eine Verzögerung (NL_REG_GD_DE_)

Die Annotation des mit diesen Konzeptualisierungen verbundenen und des dadurch begründeten *Sprachmanagements* (language management = LM), das sich in den Interviews in „language management summaries“ von den Interviewten verdichtet, berücksichtigt lediglich das organisierte Sprachmanagement in unterschiedlichen Stadien des Sprachmanagementprozesses.⁴ Neben dem Bemerkten und Bewerten einer Abweichung von der sprachlichen, kommunikativen oder soziokulturellen Norm bzw. Idealnorn gehören dazu auch der Entwurf und die Implementierung einer Korrektur (vgl. Nekvapil 2009). Auch weil das organisierte Sprachmanagement etwa auf der Mikroebene der Familie und das organisierte Sprachmanagement auf der Makroebene staatlicher Sprachverordnungen für den öffentlichen Sprachgebrauch in der Verwaltung und der Schule in Wechselwirkung stehen und dadurch z. T. schwierig zu trennen sind, werden diese bei der Annotation nicht weiter differenziert.

Die hier skizzierte scheinbar einfache und eindeutige Annotation wird im folgenden Abschnitt kritisch diskutiert.

5 Sprachbiographische Annotation im Korpus deutsch-polnischer Bilingualer

Die soziolinguistische Annotation und die Interpretation des aufgenommenen Materials wurden von drei Forscherinnen vorgenommen. Im Laufe des Annotationsprozesses kamen mehrere Probleme zum Vorschein, die sowohl inhaltlicher als auch technischer Natur waren.

Auf der analytischen Ebene kam es zur Uneindeutigkeit bei der Kategorisierung jeweiliger Konzeptualisierungen.⁵ In der Konzeptualisierungsspur (genannt Konzeptualisierung – siehe Typ 1) wurden drei Kategorien markiert: LEL, LI, LM. Nicht immer war eine klare Trennung der Kategorien möglich und die Meinungen der Annotatorinnen

⁴ In der Sprachmanagementtheorie wird – etwa in Nekvapil (2009) – einfaches und organisiertes Sprachmanagement unterschieden. Das erstere geschieht ad hoc in der Interaktion, die in der gegebenen Situation abläuft, während das andere Sprachmanagement transsituationell ist und auf die optimale Ausgestaltung von Interaktionen eines bestimmten Typus im Allgemeinen abzielt.

⁵ Solch ein Problem kam bei der Annotation der weiteren Kategorien (Domäne und Lebensphase) nicht vor.

bezüglich der Zuschreibung gingen auseinander. Am häufigsten schien die Grenze zwischen LM und LEL verwischt zu sein.⁶ Dem gehen wir im folgenden Abschnitt nach.



Abbildung 1: Elan Typ 1 Unklarheiten bei der Konzeptualisierung

Im obigen Bild, das einen Auszug einer ELAN-Annotation zeigt, kann man sehen, dass der Satz: „die haben uns immer jenannt albo Szwabys albo hitlerowcy|“ zusammen mit der folgenden Aussage als Sprachmanagement (LM) klassifiziert wurde. Die Erzählung von AL_PAR geht weiter:

(12) ... jak co @ was war denn ach ty hitlerowiec | na nachdem haben da bei Polenzeit haben sie so ver * ver * verboten | dass so kann man nicht so * | dass nicht soll Deutsch sprechen aber wir .. wo wie wir uns Deutsch haben getr * haben wir immer Deutsch gesprochen | und hier waren viel Deutsche noch (AL_PAR_GP_DE)

Die oben präsentierte Klassifizierung der Aussage als LM war von der Korrektorin anders interpretiert worden. Die Aussage:

(13) die haben uns immer jenannt albo Szwabys albo hitlerowcy | jak co @ was war denn ach ty hitlerowiec | na nachdem haben da bei Polenzeit haben sie so ver * ver * verboten

wurde als LEL eingestuft, da sie auf schwierige Erlebnisse des Informanten, die in Folge des Krieges passiert sind, hinweist. Aus der Äußerung geht hervor, dass der Informant als Kind direkt nach dem Krieg schikaniert und beschimpft wurde. Zugleich wurde aber der Textabschnitt:

⁶ Viel einfacher waren die Sprachideologien zu unterscheiden. Wie schon früher angedeutet wurde, bezeichnet Busch Sprachideologien als Konstrukte, die den Sprachen oder Varietäten aufgrund von ethnischer, nationaler bzw. sozialer Zugehörigkeiten einen gewissen Wert zuschreiben und mit Macht verbinden und den Gebrauch bzw. die Verweigerung bestimmter Sprachen legitimieren (vgl. Busch 2015b: 57).

(14) na nachdem haben da bei Polenzeit haben sie so ver * ver * verboten | dass nicht soll Deutsch sprechen aber wir .. wo wie wir uns Deutsch haben getr * haben wir immer Deutsch gesprochen | und hier waren viel Deutsche noch

als LM markiert.

Die Schwierigkeiten bezüglich der Kategorisierung ergaben sich vor allem aus dem vagen LEL-Begriff. LEL wurde im Sinne von Busch (2015c: 3) „a first person perspective focussing on how they live language as a subjective experience“ verstanden, d. h. als subjektive Erfahrung des Erlebens einer Sprache durch die Sprecherin oder den Sprecher. Busch selbst geht von der Komplementarität des Spracherlebens und der Sprachideologie aus, die sich, ihrer Meinung nach, aus der Perspektive einer zweiten bzw. distanzierten Person darauf konzentriert, wie die Sprecherin oder der Sprecher durch Sprache als sprechendes Subjekt konstituiert wurde (vgl. ebd.). Die Unklarheiten bezüglich der Zuordnung zeigten sich insbesondere beim individuellen Sprachmanagement⁷ und Spracherleben. Bei der Skizzierung des individuellen Sprachmanagements wird der Umgang mit sprachlichen Problemen thematisiert, zugleich ist aber das Spracherleben als Reaktion auf bestimmte Sprachprobleme im Spiel. Es ist an bestimmte Ereignisse gebunden, die Ergebnis der jeweiligen Sprachsituation sind. Dies legt die Frage nahe, wann die Erfahrung, die von den Interviewten angesprochen wird, als Spracherleben und wann sie als Wertung als ein Teil des Sprachmanagements bzw. beides verstanden werden soll.

Nicht nur die Abgrenzung zum Sprachmanagement, sondern eben auch die Unschärfe des Spracherlebenkonzeptes bereiteten Schwierigkeiten bei der Annotation. Unklar war die Tatsache, ob das Spracherleben explizit ausgedrückt werden muss, um entsprechende Textstellen als Spracherleben zu kategorisieren. Textabschnitte, in denen Emotionen der Interviewten im Zusammenhang mit dem Sprachgebrauch explizit benannt wurden, wie die Aussage von JS_STR, waren bei der Annotation des Spracherlebens unproblematisch:

(15) no wie pani ja mówię w języku polskim | oczywiście się strasznie strasznie się zawsze wstydzę | jak mi wyleci jakieś takie słowo | bo ja jeszcze ten mój mąż obok stoi | ja nie lubię nie nie lubię tego | że w Polsce jestem i po niemiecku coś mówię (JS_STR_GD_PL)

Na wissen Sie, ich spreche Polnisch, natürlich schäme ich mich immer furchtbar, wenn mir ein Wort herausrutscht | weil wenn dazu mein Mann daneben steht | ich mag es nicht ich mag es nicht, dass ich in Polen bin und Deutsch spreche.

Das angeführte Beispiel benennt ohne Zweifel Gefühle und Erlebnisse der Informantin. Sie „schämt sich furchtbar“, wenn ihr ein Wort auf Polnisch fehlt. Sie „mag es nicht“,

⁷ Wie im 1. Abschnitt ausgeführt, verstehen wir Sprachmanagement als Oberbegriff für alle Handlungen, die unternommen werden, um „sprachliche Probleme“ zu lösen: „Management refers to a wide range of acts of attention to ‘language problems’“ (Neustupný & Nekvapil 2003: 185). Auf der Makroebene umfasst es die ‚Sprachplanung‘, die in der ‚Sprachpolitik‘ verankert sein kann, auf der Mikroebene – die individuellen Maßnahmen zur Behebung sprachlicher Probleme (vgl. Neustupný & Nekvapil 2003). Gerade diese individuelle Perspektive, welche auch die ‚Bottom-up-Sprachplanung‘ berücksichtigt (Nekvapil & Sherman 2015) scheint sich in der Phase der (emotionalen) Bewertung mit der LEL-Kategorie zu überlappen.

wenn sie in Polen etwas auf Deutsch sagt. Sowohl die Geschehnisse als auch die mit ihnen zusammenhängenden Emotionen werden explizit präsentiert. Die Einstufung solcher Aussagen bereitete den Annotatorinnen daher keine Probleme.

Komplizierter war die Bestimmung des Spracherlebens bei der Annotation derjenigen Interviewausschnitte, in denen die von Busch (2015b: 62f.) erwähnten „Schlüsselworte“ nicht vorkamen, welche das Erleben explizit zum Ausdruck bringen (zum Beispiel Angst, Scham, Zorn auf dem einen Ende oder Stolz auf dem anderen Ende der emotionalen Skala), und nur implizit auf die Erlebnisse hinweisen, die die Sprachbiographie der Sprecherin oder des Sprechers beeinflusst haben könnten. Einige Annotatorinnen annotierten nur den expliziten Ausdruck von Emotionen, die anderen berücksichtigten auch implizite Hinweise auf gewonnene Erfahrungen. Busch (2015c: 9) bemerkt, dass das Spracherleben eine emotionale oder leibliche Erfahrung dramatischer Begegnungen mit den Anderen sei, die „part of the repertoire, in the form of explicit and implicit linguistic attitudes and habitualized patterns of language practice“ werden. Diese phänomenologische Herangehensweise geht über das wortwörtliche Zitieren hinaus und lässt Raum für Interpretation.⁸ Es ist jedoch nicht zu vermeiden, dass viele der Interviewausschnitte, die von den Annotatorinnen als LM klassifiziert wurden, weil darin auf Sprachmanagement hingewiesen wird, zugleich wichtige emotionale Erlebnisse und Erfahrungen der Interviewten thematisieren (oft jedoch nicht explizit und in einem längeren Erzählabschnitt ausgedrückt), wie dies im Ausschnitt aus dem XL_PIL-Interview deutlich wird. XL_PIL berichtet, wie ihm als kleinem Kind deutsche Bücher vorgelesen wurden:

(16) ich sage auf meine meine Sprache, das ist eine Märchensprache | sonst gleich nach dem Kriege waren hier doch keine polnische Bücher nur deutsche | die waren überall | no und meine Mutter @ da wo eine Märchen @ buch ein Märchenbuch war hat uns abends z* zu einschlafen hat sie uns @ gelesen die die Märchen | ja darum sage ich | dass meine Sprache das ist eine Märchensprache (XL_PIL_GP_DE)

Der Abschnitt wurde als LM annotiert, da hier der Informant ohne Zweifel davon berichtet, wie ihm Deutsch beigebracht wurde. Im zitierten Ausschnitt ließe sich jedoch auch ein Hinweis auf Spracherleben finden. Implizit verweist der Informant auch auf seine emotionale Verfassung, indem er sein Deutsch als eine „Märchensprache“ bezeichnet. Die Märchensprache im Beispiel (16) ist aber eine Sprache, die einen alternativen Fluchtraum eröffnet und in die mysteriöse und wunderbare Welt von Feen und seltsamen Wesen führt. Der Textabschnitt weist auf eine positive emotionale Ladung der Erinnerungen und auf ein positives Spracherleben des Deutschen in Abgrenzung zur Umgebung hin.

⁸ Auch Buschs methodologische Überlegungen zur Untersuchung von Mehrsprachigkeit (Busch 2018) scheinen die breitere Auffassung von der LEL-Definition zu bestätigen. So schlägt Busch (2018: 6) vor, Sprache anhand von Spracheinstellungen oder körperlichen und emotionalen Aspekten von Spracherleben zu fokussieren. Das sprachliche Repertoire soll u. a. anhand von sog. Sprachenportraits („language portrait“) untersucht werden. Der multimodale Charakter ermöglicht das Repertoire auf eine bildbezogene und eine sprachlich-diskursive Art und Weise zu repräsentieren. Diese Vorgehensweise rückt eine interpretative Analyse in den Vordergrund (vgl. ebd.). Zur Interdependenz von Sprachgebrauch und der emotionalen Beziehung dazu siehe u. a. Kramsch (2009) oder Pavlenko (2012).

Bei der Standardisierung der Kategorisierung war daher die Rolle der Korrektorinnen der soziolinguistischen Annotation von großer Bedeutung. Als Spracherleben wurden auch Interviewabschnitte eingestuft, die implizit auf Erlebnisse und Erfahrungen der Interviewten in Bezug auf Sprache hingewiesen haben.

Wie schon früher erwähnt wurde, können manche Aussagen gleichzeitig beiden Konzeptualisierungsmustern zugeschrieben werden, wie dies im nächsten Abschnitt ausgeführt wird:

	00:03:42.000	00:03:44.000	00:03:46.000	00:03:48.000	00:03:50.000	00:03:52.000
Sp_Orth_Klausen [122]	meine Mutter hatte sehr sie sprach kein Polnisch als sie als auch noch als sie zur Schule kam rj					
Int1 [11]						
Int2 [01]						
Konzeptualisierung [2]	S: LM, S: LEL					

Abbildung 2: Elan Typ 2 Zusammenfallen von LM und LE

In dem angeführten Beispiel werden gleichzeitig LM und LEL zum Ausdruck gebracht. Die Interviewte NA_HAM erzählt die Geschichte der Einschulung ihrer Mutter nach dem Zweiten Weltkrieg:

(17) meine Mutter hatte sehr sie sprach kein Polnisch als sie als auch noch als sie zur Schule kam mit sechs Jahren | wurde ist auch in Polen auch ganz schön @ von den Kindern ja @ gehänselt worden | dies diesbezüglich die hatte überall in halt * Hakenkreuze | wenn sie von der Schule kam auf dem auf dem das war ja direkt nach dem Zweiten Weltkrieg | also war es halt nicht einfach für sie (NA_HAM_GD_DE)

Mit Blick auf die Abweichung von der neuen kommunikativen Norm durch die Absenz des Polnischen, wird LM schon am Anfang der Aussage thematisiert: „meine Mutter hatte sehr/ sie sprach kein Polnisch“. Implizit ist LM in der Erzählung auch weiter vorhanden. Der Hörer wird informiert, wie deutschstämmige Kinder in der Schule behandelt wurden. Zugleich kommt aber in dieser Geschichte LEL wertend zum Ausdruck: „wurde [...] von den Kindern ja ja @ gehänselt worden“, „[die hatte überall [...] Hakenkreuze], also war es halt nicht einfach für sie“. Während die LEL-Annotation an der Stelle endet, wird LM weiter fortgesetzt.

Wie oben bemerkt wurde, ergibt sich eine weitere Schwierigkeit aus der Vielfalt und unterschiedlichen Länge von metasprachlichen Daten, die in einem Interviewabschnitt zu annotieren waren. Die metasprachlichen Daten im Interview etwa zu Entwicklungsphasen, Sprachdomänen und Konzeptualisierung traten teils synchron, teils versetzt auf, was zu häufigen Annotationsfehlern (wie zum Beispiel fehlenden Abschlussmarkierungen des jeweiligen Blocks) führte. Außerdem haben sich Konzeptualisierungen oft auch partiell überlappt (eine fing an, bevor die andere endete), was die Annotationsaufgabe weiter erschwert hat.

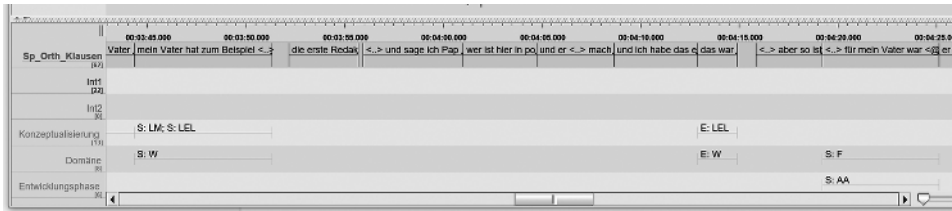


Abbildung 3: Elan Typ 3 Vielfalt und asynchrones Vorkommen der Annotationskategorien

Wie dies das Beispiel oben visualisiert, fängt die Annotation von LM, LEL und W gleichzeitig an. Während LEL und W abgeschlossen werden, wird LM fortgesetzt und sein Abschluss ist in dem hier visualisierten Ausschnitt nicht sichtbar. Zugleich fangen aber bereits auch F und AA an. Dieses Beispiel zeigt, dass die Kompliziertheit soziolinguistischer Annotation in der gleichzeitigen Interpretation des ganzen Interviews und seiner einzelnen Teile besteht. Probleme bereitet die richtige Abgrenzung der Textteile für die Annotation. Darüber hinaus werden einige Aussagen durch andere Themen unterbrochen, andere bleiben unvollendet, wie das für gesprochene Texte charakteristisch ist. Bei einem größeren Untersuchungsteam erhöht sich ferner die Wahrscheinlichkeit einer unterschiedlichen Interpretation vergleichbarer Beispiele.

Einige Probleme ergeben sich auch aus technischen Einschränkungen des Korpusaufbaus. Wegen der Anforderungen bezüglich der Aufteilung der Korpora in „polnische“ und „deutsche“ Daten, wurden nur einsprachige Audioabschnitte, sog. „clean Data“, berücksichtigt. Alle längeren Audioabschnitte, in denen zum Beispiel Codeswitching oder Gesprächsstörungen vorkamen, mussten aus dem Korpus entfernt werden.⁹ Dieses Vorgehen führte an vielen Stellen zur Entstehung inhaltlicher Lücken in der annotierten Transkription der Erzählung im Korpus. Die genaue Bestimmung des Themas und somit bestimmter Kategorien erwies sich für die Annotatorinnen aufgrund der fehlenden Referenz an manchen Stellen als schwierig oder sogar unmöglich und musste von den Interviewerinnen nachbereitet werden.

Zusätzlich wurden die Aufnahmen immer beim Kodewechsel geschnitten. Im Falle der Personen, bei denen die Alternation des sprachlichen Codes häufig vorkam, entstanden viele kurze Audiodateien (zum Beispiel 50 Sek. lang), bei denen die Annotation manchmal unmöglich war oder für die Annotatorinnen keinen Sinn ergab: Sie wussten etwa, dass es im folgenden Abschnitt beispielsweise um Sprachmanagement geht. Diese Information wurde jedoch nur aufgrund der Kenntnis der vorherigen Datei gewonnen. Somit erwies sich die Markierung der Kategorie in der jeweiligen Datei als sinnlos, da diese Zuschreibung anhand des vorhandenen Materials nicht sichtbar war.¹⁰

⁹ Weggelassen wurden auch Aufnahmestellen mit Telefonaten, privaten Gesprächen oder Abschnitte, die aufgrund der zu detaillierten Informationen die Identität der Interviewten offenbaren könnten.

¹⁰ In den Dateien, die nur kurze Interviewausschnitte enthalten, konnten wegen des fehlenden Textes und der fehlenden Informationen die einzelnen Annotationskategorien nicht markiert

6 Fazit

Im Rahmen des Projektes „Language across Generations: Contact Induced Change in Morphosyntax in German-Polish Bilingual Speech“ entstand ein spezialisiertes und innovatives Korpus von Sprachbiographien deutsch-polnischer Bilingualer. Es enthält digital aufbereitete und annotierte Transkripte sprachbiographischer Interviews von zwei Generationen deutsch-polnischer Bilingualer.

Die Besonderheit unseres Projektes liegt darin, dass die narrativen sprachbiographischen Interviews zugleich eine Quelle von sprachlichen und von metasprachlichen Daten sind. Somit wird versucht, die Interdependenz sprachkontaktbezogener formorientierter und soziolinguistischer Forschungsergebnisse zu untersuchen.

Die Tatsache, dass die gewonnenen Daten sowohl qualitativ als auch quantitativ ausgewertet werden konnten, erlaubte den Forscherinnen und Forschern prototypische Sprachbiographien für die jeweilige Generation herauszuarbeiten. Es ließen sich Muster bezüglich des Spracherwerbs, -gebrauchs und auch -verlusts bestimmen, die sich aufgrund der vertretenen Sprachideologien und des darauf zurückzuführenden Sprachmanagements zwischen der „Generation Polen“ und der „Generation Deutschland“ stark voneinander unterscheiden.

Die Annotation und Analyse metasprachlicher Daten erfolgte in Bezug auf drei Kategorien: „Lebensphase“, „Domäne“ und „Konzeptualisierung der Sprache“. Die Annotation bereitete gewisse Schwierigkeiten, die sowohl technischer als auch konzeptueller Natur waren. Eine Herausforderung für den Korpusaufbau wird auch das Abrufen der metasprachlichen Daten sein, die anders als sprachliche Daten, das Abrufen längerer Textpassagen mit vielen Nebeninformationen (zum Beispiel Äußerungen der Interviewer) ermöglichen sollten.

Die soziolinguistische Annotation im polnisch-deutschen Korpus ist eine Pionierarbeit, die innovativ und herausfordernd war. Das Team hat die Hoffnung, dass dieser Ansatz Interesse weckt und die Diskussion über die Verbindung von Soziolinguistik und Korpuslinguistik anregt.

Literatur

- Anstatt, Tanja (2009): Der Erwerb der Familiensprache: Zur Entwicklung des Russischen bei bilingualen Kindern in Deutschland. In: Ingrid Gogolin, Ursula Neumann (Hgg.), *Streitfall Zweisprachigkeit – The Bilingualism Controversy*. Wiesbaden: VS Verlag für Sozialwissenschaften, 111–131.
- Auer, Peter (2013): *Sprachliche Interaktion*. Tübingen: Niemeyer.
- Benmamoun, Ellabas/Montrul, Silvina/Polinsky, Maria (2013): Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics* 39.3–4, 129–181.
- Blommaert, Jan (2008): Language, asylum, and the national order. *Working Papers in Urban Language & Literacies* 50, 2–21.

werden. Bei der Suchoption im Korpus wären diese Textabschnitte außerhalb des Kontexts angezeigt worden.

- Brehmer, Bernhard/Mehlhorn, Grit (2015): Russisch als Herkunftssprache in Deutschland. Ein holistischer Ansatz zur Erforschung des Potenzials von Herkunftssprachen. *Zeitschrift für Fremdsprachenforschung* 26.1, 85–123.
- Busch, Brigitta (2015a): Expanding the notion of the linguistic repertoire: on the concept of Spracherleben – the lived experience of language. *Applied Linguistics* 38.3, 340–358.
- Busch, Brigitta (2015b): Zwischen Fremd- und Selbstwahrnehmung. Zum Konzept des Spracherlebens. In: Anna Schnitzer, Rebecca Mörgen (Hgg.), *Mehrsprachigkeit und (Un-)Gesagtes. Sprache als soziale Praxis im Kontext von Heterogenität, Differenz und Ungleichheit*. Weinheim, Basel: Beltz Juventa, 49–66.
- Busch, Brigitta (2015c): Linguistic repertoire and *Spracherleben*, the lived experience of language. *Working Papers in Urban Language & Literacies* 148, 2–16.
- Busch, Brigitta (2017): *Mehrsprachigkeit*. 2. Aufl. Wien: facultas.
- Busch, Brigitta (2018): The language portrait in multilingualism research: Theoretical and methodological considerations. *Working Papers in Urban Language & Literacies* 236, 2–13.
- Coulmas, Florian (2005): Changing language regimes in global environments. *International Journal of the Sociology of Language* 175/176, 3–15.
- Dietz, Barbara (2011): Aussiedler/Spätaussiedler in Germany since 1950. In: Klaus J. Bade, Pieter C. Emmer, Leo Lucassen, Jochen Oltmer (Hgg.), *The Encyclopedia of Migration and Minorities in Europe: From the Seventeenth Century to the Present*. Cambridge: Cambridge University Press, 245–250.
- Fishman, Joshua A. (1964): Language maintenance and language shift as a field of inquiry: A definition of the field and suggestions for its further development. *Linguistics* 2.9, 32–70.
- Franceschini, Rita (2001): Sprachbiographie randständiger Sprecher. In: Rita Franceschini (Hg.), *Biographie und Interkulturalität: Diskurs und Lebenspraxis*. Tübingen: Stauffenburg, 111–125.
- Franceschini, Rita/Miecznikowski, Johana (Hgg.) (2004): *Leben mit mehreren Sprachen. Sprachbiographien*. Bern, Berlin: Peter Lang.
- Głuszkowski, Michał (2011): *Socjologiczne i psychologiczne uwarunkowania dwujęzyczności staroobrzedowców regionu suwalsko-augustowskiego*. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.
- Gumperz, John (1964): Linguistic and social interaction in two communities. *American Anthropologist* 66, 137–152.
- Hašová, Lucie (2004): Eine mitteleuropäische Sprachbiographie. In: Rita Franceschini, Johana Miecznikowski (Hgg.), *Leben mit mehreren Sprachen. Sprachbiographien / Vivre avec plusieurs langues. Biographies langagières*. Bern, Berlin: Peter Lang, 173–186.
- Jernudd, Bjorn H./Neustupný, Jiří V. (1987): Language planning: For whom? In: Lorne Laforge (Hg.), *Proceedings of the International Colloquium on Language Planning*. Québec: Presses de l'Université Laval, 69–84.
- Kacprzak, Paweł (2010): Polityka władz polskich wobec ludności niemieckiej w okresie funkcjonowania Ministerstwa Ziemi Odzyskanych. *Czasopismo Prawno-Historyczne* 62.2, 215–235.
- Kramsch, Claire (2009): *The Multilingual Subject*. Oxford: Oxford University Press.
- Kroskrity, Paul V. (Hg.) (2000): *Regimes of Language: Ideologies, Politics, and Identities*. Santa Fe (NM): School of American Research Press; Oxford: James Currey.
- Madajczyk, Piotr (Hg.) (1998): *Mniejszości narodowe w Polsce. Państwo i społeczeństwo polskie a mniejszości narodowe w okresach przełomów politycznych (1944–1989)*. Warszawa: Instytut Studiów Politycznych Polskiej Akademii Nauk.
- Maruszewski, Tomasz (2005): *Pamięć autobiograficzna*. Gdańsk: Gdańskie Towarzystwo Psychologiczne.
- Meng, Katharina (2004): Russlanddeutsche Sprachbiographien – Rückblick auf ein Projekt. In: Rita Franceschini, Johana Miecznikowski (Hgg.), *Leben mit mehreren Sprachen. Sprachbiographien / Vivre avec plusieurs langues. Biographies langagières*. Bern, Berlin: Peter Lang, 97–117.

- Montrul, Silvina A. (2008): *Incomplete Acquisition in Bilingualism: Re-Examining the Age Factor*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Nekula, Marek (2021): Sprachideologien und Sprachmanagement in sprachbiographischen Interviews: Sprecher mit tschechischem Migrationshintergrund in Bayern. *Brücken – Zeitschrift für Sprach-, Literatur- und Kulturwissenschaft* 28.2, 63–82.
- Nekula, Marek/Rychnovská, Lucie (2016): Jazyk Smetanových dopisů v dobovém kontextu / The Language of Smetana's Letters and Diaries in the Context of His Time. / Die Sprache von Smetanas Briefen im Kontext seiner Zeit. In: Olga Mojžišová, Milan Pospíšil (Hgg.), *Bedřich Smetana, Korespondence I (1840–1862). Kritická edice*. Praha: Národní muzeum – Koniasch Latin Press, 28–73, 123–181, 235–291, 429–444.
- Nekvapil, Jiří (2003): Language biographies and the analysis of language situations: On the life of the German community in the Czech Republic. *International Journal of the Sociology of Language* 162, 63–83.
- Nekvapil, Jiří (2004a): Sprachbiographien und Analyse der Sprachsituation: Zur Situation der Deutschen in der Tschechischen Republik. In: Rita Franceschini, Johanna Miecznikowski (Hgg.), *Leben mit mehreren Sprachen. Sprachbiographien*. Bern: Peter Lang, 147–172.
- Nekvapil, Jiří (2004b): Language biographies and management summaries. *Language Management in Contact Situations* 3, 9–33.
- Nekvapil, Jiří (2009): The integrative potential of language management theory. In: Jiří Nekvapil, Tamah Sherman (Hgg.), *Language Management in Contact Situations: Perspectives from Three Continents*. Berlin: Lang, 1–11.
- Nekvapil, Jiří (2012): From language planning to language management: J. V. Neustupný's heritage. *Media and Communication Studies* 63, 5–21.
- Nekvapil, Jiří/Sherman, Tamah (2015): An introduction: Language management theory in language policy and planning. *International Journal of the Sociology of Language* 232, 1–12.
- Neustupný, Jiří V. (1978): *Post-structural Approaches to Language*. Tokyo: University of Tokyo Press.
- Neustupný, Jiří V./Nekvapil, Jiří (2003): Language management in the Czech Republic. *Current Issues in Language Planning* 4, 181–366.
- Pavlenko, Aneta (2007): Autobiographic narratives as data in applied linguistics. *Applied Linguistics* 28.2, 163–188.
- Pavlenko, Aneta (2012): Multilingualism and emotions. In: Marilyn Martin-Jones, Adrian Blackledge, Angela Creese (Hgg.), *The Routledge Handbook of Multilingualism*. London: Routledge, 454–469.
- Polinsky, Maria (2006): Incomplete acquisition: American Russian. *Journal of Slavic Linguistics* 14, 191–262.
- Pomerantz, Anita (1986): Extreme case formulations: A way of legitimizing claims. *Human Studies* 9, 219–229.
- Riehl, Claudia Maria (2014): *Sprachkontaktforschung: Eine Einführung*. 3., überarbeitete Auflage. Tübingen: Narr.
- Scheffel, Michael (2012): Nach dem ‚narrative turn‘: Handbücher und Lexika des 21. Jahrhunderts. *DIEGESIS. Interdisziplinäres E-Journal für Erzählforschung / Interdisciplinary E-Journal for Narrative Research* 1.1, 43–55.
- Silverstein, Michael (1979): Language structure and linguistic ideology. In: Paul R. Cline et al. (Hgg.), *The Elements: A Parasession on Linguistic Units and Levels*. Chicago: Chicago Linguistic Society, 193–247.
- Treichel, Barbara (2004): *Identitätsarbeit, Sprachenbiographien und Mehrsprachigkeit. Autobiographisch-narrative Interviews mit Walisern zur sprachlichen Figuration von Identität und Gesellschaft*. Frankfurt/Main, Berlin: Peter Lang.
- Werlen, Iwar (2002): Sprachbiographien – Wie italienische Migrantinnen und Migranten der zweiten Generation in der deutschen Schweiz ihr Sprachleben sehen. In: Kirsten Adamzik, Eva Ross

- (Hgg.), *Biografie linguistique. Biographies langagières – Biografias linguisticas – Sprachbiographien*. In: *Bulletin vals-alsa* (Neuchâtel) 76, 57–79.
- Werlen, Iwar (2004): Domäne / Domain. In: Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier, Peter Trudgill (Hgg.), *Sociolinguistics: An International Handbook of the Science of Language and Society / Soziolinguistik: Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft*. 2., vollständig neu bearbeitete und erweiterte Auflage. Berlin, New York: Walter de Gruyter, 335–340.
- Woolard, Kathryn A. (1998): Introduction: Language ideology as a field of inquiry. In: Bambi B. Schieffelin, Kathryn Woolard et al. (Hgg.), *Language Ideologies: Practice and Theory*. Oxford, New York: Oxford University Press, 3–47.

Sprachstrukturelle Annotation der LangGener-Korpora: Typologie und Abgrenzungsprobleme

Aneta Bučková – Carolin Centner – Felicja Księżyk – Irena Prawdzc

Abstract:

Annotation of linguistic structures in the LangGener corpora: Typology and delimitation problems

The paper presents the theoretical background as well as the practical implementation of the analysis of language contact phenomena in the LangGener corpus. Our analysis is based on the distinction between borrowing of concrete phonological material and the replication of abstract linguistic patterns. These two types of language-contact-related deviations from the baseline of comparison (homeland baseline) are complemented by further annotational categories – code-switching, word order, other deviations and self-correction. The text discusses these phenomena on the basis of the material collected in the project and addresses problems concerning the annotation, for each of which a uniform solution is proposed.

1 Inhalt und Aufbau des Kapitels

Basierend auf der Typologie Matras' (2009) und seiner Unterscheidung zwischen materiellen Replikationen (MAT) und Musterreplikationen (PAT), die in Abschnitt 2 des vorliegenden Kapitels erklärt wird, sowie den psycholinguistischen Erkenntnissen über die Mechanismen der Replikation, soll im Folgenden die Umsetzung einer Klassifizierung der Ergebnisse des Sprachkontakts im LangGener-Projekt gezeigt werden. Neben den Kategorien PAT und MAT erweisen sich weitere Kategorien außerhalb der genannten Dichotomie als notwendig, um die Ergebnisse zu klassifizieren. So enthält das Korpus darüber hinaus auch die Kategorien AA (andere Abweichung), WO (word order), SC (self-correction) und CS (Codeswitching). Bei der Analyse der Abweichungen wird vom Vergleich mit einer Baseline bzw. Heimatland-Baseline ausgegangen. Diese stellt nicht die standardsprachliche Varietät der untersuchten Sprachen dar, sondern jeweils die im Heimatland gesprochene Umgangssprache (siehe Hansen & Nekula in diesem Band; zur Bestimmung der Heimatland-Baseline im deutsch-tschechischen Teilprojekt siehe auch Bučková & Nekula in diesem Band; zu Abweichungen von der Baseline, die eine Kontinuität der schlesischen und niederdeutschen Dialekte des Deutschen bei der „Generation Polen“ bilden, siehe Jorroch, Księżyk & Zielińska in diesem Band).

In Abschnitt 3 werden die angewandten Kategorien genauer vorgestellt und Beispiele dafür gezeigt. Dabei wird zunächst auf die Bestimmung des Phrasentyps bei der

Annotation eingegangen (Abschnitt 3.1). In den folgenden Abschnitten werden die einzelnen kontaktlinguistischen Annotationskategorien genauer vorgestellt und mit Beispielen aus dem Korpus illustriert. Die Abschnitte über PAT (Abschnitt 3.2), AA (Abschnitt 3.4) und SC (Abschnitt 3.7) folgen dabei aufgrund der vielen Beispiele der Aufteilung nach Phrasentypen. Bei den Abschnitten über MAT (Abschnitt 3.3) und CS (Abschnitt 3.5) erfolgt die Aufteilung nach typologischen Kriterien der jeweils verwendeten Frameworks. Im Abschnitt zur Wortfolge (Abschnitt 3.6) wird zwischen Abweichungen im Deutschen und Polnischen unterschieden.

Bei der Darstellung der einzelnen Annotationskategorien wird zunächst auf prototypische Fälle eingegangen, woraufhin jedoch auch Grenzfälle diskutiert werden, bei deren Einstufung sich zusätzliche Kriterien als erforderlich erwiesen haben.

2 Forschungsüberblick zum deutsch-polnischen bzw. deutsch-tschechischen Sprachkontakt und Typologien der Ergebnisse des Sprachkontakts

Die meisten Arbeiten, die bisher zum deutsch-polnischen Sprachkontakt erschienen sind, sind diachron angelegt und untersuchen überwiegend Lehnwörter aus dem Deutschen ins Polnische (Karszniewicz-Mazur 1988; Nowowiejski 1996) bzw. die Sprache der *Polonia* in Deutschland mit Fokus auf Lehnwörtern (Michalewska 1991; Wójtowiczowa 1995). Erst in der neueren Forschung gibt es Arbeiten, die sich auch auf durch Sprachkontakt veränderte Strukturen im Sprachgebrauch bilingualer Personen konzentrieren (Brehmer & Czachór 2012 zum analytischen Futur; Brehmer & Rothweiler 2012 zum Erwerb der Genusmarkierung; Błaszczuk 2018 zum Irrealis-Marker *by*; sowie zentrale Arbeiten von Weiss 1982, 1987).

Auch im Bereich des deutsch-tschechischen Sprachkontakts haben einige Arbeiten eine diachrone Perspektive eingenommen (Berger 2008, 2014 und die darin diskutierte Literatur) und die Lexik (Nekula 1996a; Půda 2010; Newerkla 2013; Trost & Jelínek 2000) oder Anredemuster (Betsch 2000) fokussiert. Einen synchron orientierten Blick bieten Nekula, Šichová & Valdřová (2013) oder Feldmeier (2014). Die gegenseitige Beeinflussung des Tschechischen und des Deutschen in einer bilingualen Sprachsituation wird vornehmlich am Beispiel der tschechischen Sprachgemeinschaft in Wien (Fischer 1993; Balhar 1995; Zeman 2011; Newerkla 2017 u. a.) oder der deutschen Varietät in Prag (Trost 1962; Skála 1966; Bachmann 2013) untersucht, auch der Sprachkontakt in böhmischen und mährischen Dialekten wird beleuchtet (vgl. Literatur in Nekula & Večerka 2017; Nekula 2021). Andere Autorinnen und Autoren nähern sich den deutsch-tschechischen Sprachbeziehungen wiederum anhand von Idiolekten, zum Beispiel von Lenka Reinerová (Bučková 2016), Franz Kafka (Trost 1964; Nekula 2003; Blahak 2015) oder Bedřich Smetana (Nekula & Rychnovská 2013, 2016).

Eine Typologie der Resultate des deutsch-polnischen Sprachkontakts, die sich auf die Unterscheidung zwischen materiellem und nicht-materiellem Transfer stützt, wie sie im Projekt angestrebt wird, ist dabei nur vereinzelt zu finden. Für den deutsch-polnischen Sprachkontakt stellt Pelka (2006), angelehnt an die Typologien Weinreichs (1977) und Clynes (2003), eine Typologie für die Resultate des deutsch-polnischen bzw. deutsch-schlesischen Sprachkontakts auf. Dabei untersucht sie anhand eines selbst erstellten Korpus aus verdeckten Aufnahmen und Wortprotokollen den Einfluss des Polnischen bzw.

des schlesischen Ethnolekts auf das Deutsche der deutschen Minderheit in der Gegend von Oberglogau in Oberschlesien. Den genannten Typologien folgend unterteilt sie die Ergebnisse des Sprachkontakts in phonetisch-phonologische, lexikalisch-semantische und grammatische (morphosyntaktische) Transferenzen.

Neben der Dreiteilung der Typologie, die sich an den Ebenen des Sprachsystems orientiert, gibt es weitere Ansätze, wie sich die Ergebnisse des Sprachkontakts einteilen lassen. Eine Typologie, die in der neueren Forschung weit rezipiert wird, stammt von Matras (2009) und Sakel (2008). Neben Aspekten des gesellschaftlichen und des diachronen Sprachwandels geht Matras (2009) auch auf die Unterscheidung zwischen *replication of linguistic matter* (MAT) und *pattern replication* (PAT) ein. Matras & Sakel (2007: 829f.) definieren MAT und PAT wie folgt:

„Contact-induced language change can lead to direct replication of morphemes and phonological shapes from a source language; we shall refer to this in the following as replication of linguistic matter, abbreviated MAT. Language contact can also lead to re-shaping of language-internal structures. In the latter process, the formal substance or matter is not imported but is taken from the inherited stock of forms of the recipient or replica language (i. e. the language that is undergoing change). Rather, it is the patterns of distribution, of grammatical and semantic meaning, and of formal-syntactic arrangement at various levels (discourse, clause, phrase, or word) that are modelled on an external source. We call this pattern replication, abbreviated PAT.“

Durch die Unterscheidung zwischen PAT und MAT wird die Dreiteilung nach den Systemebenen aufgehoben, da PAT auf verschiedenen Ebenen stattfinden kann. So nennt Matras (2009: 245) als Beispiel für die Kategorie *lexical semantics*, die in etwa der *Lehnübersetzung* bei Weinreich (1977) entspricht, das Wort *Fernsehen*, das nach dem Vorbild *television* aus dem Englischen gebildet wurde. Hier wurde also nicht *television* als Lehnwort in das Deutsche übernommen, sondern die einzelnen englischen Morpheme mit deutschen Morphemen wiedergegeben. Auch die Gründe für die Übernahme und die Mechanismen dahinter werden bei Matras thematisiert. So geht er davon aus, dass beiden Konstruktionen, die in Kontakt treten, ein gemeinsames *Pivot*, also ein gemeinsamer Angelpunkt, zugrunde liegt. Dieses *Pivot* wird mit einem Muster in der anderen Sprache assoziiert (Matras 2009: 240f.).

Die folgende Darstellung der Annotation sowohl hinsichtlich der sprachformbezogenen Annotation als auch der Bestimmung des Phrasentyps ist als eine Anwendung der Theorie Matras' auf unser Korpusmaterial zu sehen. Sie stellt daher keine theoretische Auseinandersetzung mit den bisherigen Ansätzen dar, sondern einen Versuch, diese auf Daten der gesprochenen Sprache im deutsch-polnischen bzw. deutsch-tschechischen Sprachkontakt anzuwenden.

3 Typologie und Annotation der Ergebnisse des Sprachkontakts im LangGener-Projekt

Die im Rahmen der Feldforschung aufgenommenen Interviews wurden technisch weiterbearbeitet und transkribiert (siehe Bučková & Centner in diesem Band). Anschließend wurden die Transkripte sprachformbezogen und soziolinguistisch annotiert. Auf die soziolinguistische Annotation gehen Jańczak, Nekula & Zielińska (in diesem Band) ein.

Wir befassen uns im Folgenden mit der sprachformbezogenen Annotation, die in Bezug auf Klausen erfolgt ist (zur Bestimmung von Klausengrenzen siehe Bučková & Prawdzcic in diesem Band).

Die sprachstrukturelle Annotation wurde auf zwei Ebenen vorgenommen. Auf der ersten Ebene wurde im Abschnitt der betreffenden Klausen die syntaktische Phrase eingetragen, in der eine Abweichung von der Heimatland-Baseline identifiziert wurde. Dazu wurde die Spur *Sp_Phrasentyp* im Programm ELAN genutzt. Auf der zweiten Ebene wurde die Art der auf den Sprachkontakt zurückgeführten Abweichung spezifiziert. Dazu wurde die Spur *Sp_PAT* genutzt.

3.1 Phrasentypen

In diesem Abschnitt gehen wir auf die Annotation von Phrasen ein, in deren Rahmen wir die Abweichung von der Heimatland-Baseline identifiziert und annotiert haben. Tabelle 1 bietet eine Übersicht über die verwendeten syntaktischen Phrasentypen. Phrasentypen werden möglichst theorieneutral durch den jeweiligen Kopf bestimmt, der die grammatischen Eigenschaften der Wortgruppe bestimmt. Die betreffende syntaktische Phrase wird in der Tabelle für eine bessere Übersichtlichkeit durch Fettdruck hervorgehoben, die Variante der Heimatland-Baseline wird in eckigen Klammern ergänzt:

Phrase	Label	Beispiel	Quelle
Nominalphrase	NP	in Zirndorf war der [das] Lager ¹ für Ausländer oder Asylbewerber oder	EQ_HER_MI_DE
Präpositionalphrase	PP	und so den [der] ganzen Tschechoslowakei in dieses Land gekommen aus allen Ecken der Tschechoslowakei	KC_ISM_MI_DE
Verbalphrase	VP	und dann wenn wir ihm [ihn] geschimpft haben	KN_MUN_MI_DE
Adjektivphrase	AP	ja und dann große [großen] Hoffnungen und alles mögliche und so weiter	WG_MUN_MI_DE
Adverbialphrase	AdvP	hat sehr schöne [schön] ausgesehen	EQ_HER_MI_DE
Satz	S	und # dann hat jemand gemerkt dass mein Schwiegervater war eigentlich ein Prager Deutscher [dass mein Schwiegervater eigentlich ein Prager Deutscher war]	KG_MUN_MI_DE

Tabelle 1: Phrasentypen

¹ Für die Bedeutung der bei der Transkription verwendeten Sonderzeichen siehe Bučková & Prawdzcic in diesem Band.

Um im LangGener-Projekt eine einheitliche Annotation zu gewährleisten, hat sich das Annotationsteam bei Zweifelsfällen auf einheitliche Lösungen festgelegt. Eine allgemeine Festlegung unsererseits lautete, dass der vorhandene und nicht der fehlende Phrasentyp annotiert wird. So wurde bei *wir haben [mit] ihm Polnisch geredet zu Hause* (BK_AUG_GD_DE) VP notiert und nicht etwa PP, obwohl in der Klausel eine Präposition fehlt. Des Weiteren wurde bei abweichender Adjektivendung immer AP annotiert, auch wenn das (nicht weiter modifizierte) Adjektiv sich innerhalb einer NP befindet (zum Beispiel bei *das ist mein älteste[r] Bruder*, IL_PIL_GP_DE). Im Gegensatz dazu wurde bei abweichender Genuszuweisung, die sich ebenso auf das Adjektiv auswirkte, NP annotiert (zum Beispiel bei *# ich habe dann eben _ gerade die in # _ im Herbst ne im im _ # September habe ich die das [den] dritte[n] Jahrgang angefangen*, KC_ISM_MI_DE).

Bei Abweichung von der Heimatland-Baseline innerhalb einer Präpositionalphrase wurde immer PP annotiert (zum Beispiel bei *in der Nähe von Karlsbad war ein Riesengebiet ausgewiesen f * zu Übungen für die [das] Militär*, KC_ISM_MI_DE). Im Falle von Abweichungen bei Ergänzungen, in denen das Verb über den Kasus bzw. die Präposition entscheidet (zum Beispiel bei *und sie hat sich da beworben in [bei] einem so einem _ staatlichen Bauernhof*, KC_ISM_MI_DE), wurde dagegen VP annotiert.

Bei Abweichungen von der Heimatland-Baseline, die die Negation betrafen, wurde auf der Satzebene S annotiert (zum Beispiel *habe ich Deutsch nicht [kein Deutsch] gesprochen*, BK_HAM_GD_DE). Ebenso wurde S bei Abweichungen annotiert, die Partikeln und Konjunktionen betrafen. Des Weiteren wurden abweichende Wortstellungen, bei denen sich die Abweichung auf den ganzen Satz bezog, als S kategorisiert. Abweichungen in der Wortstellung innerhalb einzelner Phrasen, etwa in der AdvP, wurden jeweils in der Annotation berücksichtigt (zum Beispiel bei *wie sie # Deutsch gut spricht [wie gut sie # Deutsch spricht]*, BK_HAM_GD_DE).

3.2 Pattern replications

Wie bereits erwähnt, basiert die Unterscheidung zwischen materiellem Transfer (MAT) und Übernahme der Muster einer Sprache (PAT) auf der Theorie Matras'. Während von manchen Bilingualen direkte, materielle Übernahmen, die kognitiv kontrolliert werden können, vermieden werden, werden laut Matras Musterreplikationen, die sich der kognitiven Kontrolle entziehen, aus dem Repertoire der Bilingualen (*bilingual repertoire*) gebildet. Entscheidend ist, dass dadurch das kommunikative Ziel erreicht wird (Matras 2009: 234). Dabei sind v. a. jene Kategorien von *pattern replication* betroffen, in denen wenig materieller Transfer stattfindet, wie beispielsweise definite und indefinite Artikel (Matras 2009: 236).

Im Folgenden werden im Einzelnen Typen von Konstruktionen aufgezeigt, die mit dem Label PAT annotiert wurden. Zum Teil mussten dabei pragmatische Lösungen getroffen werden, um die gesprochensprachlichen Daten in eine Typologie einordnen zu können. Da nicht nur die Art der von uns identifizierten Abweichung von der Heimatland-Baseline, sondern auch die syntaktische Ebene innerhalb der Klausel annotiert wurde, auf der die Abweichung eingeordnet werden kann, werden die Beispiele für die PAT-Annotation nach Phrasentypen geordnet.

3.2.1 Satzebene

Auf der Satzebene werden Musterreplikationen eingeordnet, die Subjekt, Negation und Wortfolge betreffen. Wir ordnen hier auch Konjunktionen und Diskursmarker ein, weil sie syntaktisch auf der Satzebene wirken. In Wirklichkeit basiert allerdings die Abweichung von der Heimatland-Baseline auf den Unterschieden in der Semantik oder Form der betreffenden lexikalischen Einheiten im Deutschen und Polnischen/Tschechischen.

Eine der Abweichungen, die bereits auf Satzebene annotiert wird, tritt bei fehlendem pronominalem Subjekt im Deutschen auf (Beispiel (1)), wo man die Auswirkung des tschechischen bzw. polnischen pro-drop vermuten kann:

- (1) weil sonst kann [ich] ja zurückfahren (JS_STR_GD_DE)

Umgekehrt gibt es im Polnischen sowie im Tschechischen Fälle des nicht-kanonischen pronominalen Subjekts, die wir als PAT annotiert haben (Beispiele (2) und (3)).

- (2) a potom jsem # řekla že **já** chci do školy [že chci do školy] (TU_HAG_SP_CZ)

,und dann sagte ich dass **ich** auch in die Schule will‘

- (3) potym już go wzie * wzięli na stałe nie ? potym już go wzie * i **on** potem był ranny [i potem był ranny] (BT_KLO_GP_PL)

,dann haben sie ihn schon auf Dauer ge* genommen * und dann wurde **er** verwundet‘

Fälle von Musterreplikationen aus der Kontaktsprache liegen auch in den deutschen Äußerungen (4) und (5) der interviewten Bilingualen vor. In Anlehnung an das Polnische fehlt darin das formale Subjekt *es*, das im Polnischen keine Entsprechung hat, im Deutschen dagegen u. a. wie hier in Verbindung eines Kopulaverbs mit einem adjektivischen Prädikativ obligatorisch wäre (vgl. Engel 1996: 190; Engel 1994: 171; Helbig & Buscha 1996: 399).

- (4) de waren Derfer da **war** [es] **schlimmer** dann Derfer da **war** [es] **besser** nich (BT_KLO_GP_DE)

- (5) das heißt er besucht die Tochter _ öfter wie **wie** [es] **möglich ist** oder so was nicht ? (KL_BIE_GD_DE)

Während die Negation im deutschen Standard auf der Satzebene mithilfe nur eines Negationsworts ausgedrückt wird (Helbig & Buscha 2005: 544), gilt im Polnischen und Tschechischen *Negative Concord* (sog. doppelte Negation), weswegen man beim Auftreten von zwei Negatoren bei einfach negativer Lesart im Deutschen (Beispiel (6)) auf Satzebene einen Transfer aus dem Polnischen/Tschechischen voraussetzen kann:

- (6) wenn **keiner** das **nicht** wusste [wenn keiner das wusste] (BT_KLO_GP_DE)

Auch durch die Ausdifferenzierung der Negationswörter im Deutschen wie *kein* vs. *nicht*, die beide im Polnischen durch *nie* ‘nicht’ vor dem Prädikat wiedergegeben

werden, kommt es zu Musterreplikationen auf Satzebene, da im Polnischen die Negation sowohl vor dem Substantiv als auch vor dem Verb mit *nie* realisiert wird (vgl. Engel et al. 2018: 211). Aufgrund dessen kann es zu einer Musterreplikation bei der Negation kommen, wenn die Vielfalt der Negationswörter im Deutschen nach polnischem Vorbild durch *nicht* (nach dem Muster *nie*) wiedergegeben wird, obwohl, wie in Beispiel (7), *kein* vor dem Substantiv stehen sollte. Bei diesem Beispiel handelt es sich nicht um eine pragmatische Funktion, die eine Betonung der negativen Partikel rechtfertigen würde. Die Verwendung dieser ist daher auf Sprachkontakt zurückzuführen:

- (7) wenn ein Kunde Deutsch **nicht** spricht [wenn ein Kunde kein Deutsch spricht]
(LC_BUR_GD_DE)

Ebenfalls auf Satzebene werden Beispiele annotiert, bei denen eine im gegebenen Kontext abweichende Konjunktion bei der Einleitung des Satzes – in Beispiel (8) *wenn* statt *als* – verwendet wird:

- (8) **wenn** [als] sie erstes Mal geheiratet hat mit siebzehn Jahren (SV_NAK_GP_DE)

Auch Diskursmarker als ein Typ der annotierten Abweichung werden als PAT-Replikationen auf Satzebene betrachtet, wie in Beispiel (9):

- (9) ich ich bin so eine kleine Lehrerin **sag mal so** (BK_AUG_GD_DE)

Das zitierte Beispiel (9) stellt eine Nachbildung der polnischen Phrase *powiem tak* dar, eines häufigen Diskursmarkers, der durch Bilinguale im Deutschen mit *sag mal so* wiedergegeben wird.

3.2.2 Nominalphrase

Auf der Ebene der Nominalphrase haben wir PAT-Replikationen eingeordnet, die Artikel sowie lexikalische Neologismen betreffen.

Wie bereits von Matras (2009: 235) beobachtet, finden im Bereich der definiten und indefiniten Artikel Musterreplikationen statt. Diese sind darauf zurückzuführen, dass im Deutschen Artikel als Begleiter der Substantive stehen, wofür das Polnische keine strukturellen Entsprechungen hat (Engel et al. 1999: 818). Im Polnischen und Tschechischen spielt beim Ausdruck der (Un-)Bestimmtheit vielmehr die Wortfolge eine zentrale Rolle (vgl. u. a. Nekula 1996b). Wann im Deutschen ein bestimmter, unbestimmter oder Nullartikel gebraucht wird, ist dabei von semantischen und syntaktischen sowie kommunikativen Faktoren abhängig (Helbig & Buscha 2005: 329). Beispiele, in denen es beim Artikel zu einer Abweichung von der Heimatland-Baseline kommt, wurden dennoch einheitlich als PAT NP, also als Musterreplikation im Bereich der Nominalphrase annotiert. Im Falle des fehlenden Artikels im Deutschen, wo man den Einfluss des Polnischen/Tschechischen vermuten kann, wird quasi das fehlende Muster repliziert, wie in Beispiel (10):

- (10) Wir haben auch [eine] **polnische Schule** (BL_AUG_GD_DE)

Aufgrund der indeterminierten Bedeutung ist hier im Deutschen ein unbestimmter Artikel notwendig (Helbig & Buscha 2005: 336). Da das Polnische aber keine Artikel hat, gibt es im Deutschen folglich nur Beispiele, in denen ein Artikel aufgrund des Musters im Polnischen fehlt. Auch bei deutsch-tschechischen Bilingualen (Beispiel (11)) kann es zu solchen Abweichungen von der Heimatland-Baseline kommen:

- (11) auch in einem Lager zuerst _ das war so [eine] **alte Fabrik** (BL_BLA_SP_DE)

Spiegelbildlich gilt dies auch umgekehrt, wenn also im Polnischen oder Tschechischen vor dem Substantiv ein überflüssiges Indefinit- oder Demonstrativpronomen steht. Weder im Polnischen noch im Tschechischen gibt es Artikel, wie wir sie aus dem Deutschen kennen, sondern lediglich Determinationsphrasen. Diese haben jedoch eine deiktische Funktion. Auch sind sie deutlich weniger frequent als im Deutschen. Durch das Vorbild des Deutschen ist in den sprachlichen Daten der Bilingualen ein häufiger Gebrauch des Indefinit- oder Demonstrativpronomens, ähnlich einem Artikel im Deutschen, zu verzeichnen (Beispiele (12)–(14)):

- (12) i **te** dzieci jak później się spotkali (UM_JAN_GP_PL)

,und die Kinder, wenn sie sich später getroffen haben‘

- (13) že jako _ # mi nemůžou nabídnout **nějaký** stálý místo (BJ_MUN_SP_CZ)

,dass quasi _ # sie mir keine feste Stelle anbieten können‘

- (14) ony _ ony **jedyn** czas w szpit * w szpitalu byli w Koźlu (BT_KLO_GP_PL)

,sie _ sie waren eine Zeit im Kranken* im Krankenhaus waren sie in Kosel‘

Auch wenn es im Bereich der Nominalphrase selten zu lexikalischen Neologismen kommt, sind diese ebenfalls zu verzeichnen, wie in Beispiel (15):

- (15) nám vykrádali několikrát nám vykradli # **schránku do* dopisovou [poštovní]** (WG_MUN_MI_CZ)

,uns wurde ein paar mal # der Briefkasten ausgeraubt‘

In (15) bildet der Sprecher nach dem deutschen Muster *Briefkasten* die Bezeichnung *dopisová schránka* ‚Brief.ADJ, Kasten‘, wobei in der Heimatland-Baseline statt *dopisová* ‚Brief.ADJ‘ das Adjektiv *poštovní* ‚Post.ADJ‘ zu gebrauchen wäre.

Als PATs in Form einer kontaktinduzierten Grammatikalisierung werden ebenso die nachfolgenden Belege ((16) und (17)) klassifiziert. Hier führt der Sprachkontakt zur Entstehung neuer grammatischer Konstruktionen in der nachbildenden Sprache, etwa zum Gebrauch eines einheitlichen reflexiven Possessivpronomens (vgl. Engel et al. 2018: 71; Bartnicka et al. 2004: 298) nach dem polnischen Muster:

- (16) von Neustadt sind sind dort # Herrn vor allem Herrn die Artikel schreiben von **seinem** [ihrem] Ort von **seiner** [ihrer] Gegend (RN_BRO_GP_DE)

- (17) wir sind ja hier wir haben hier **sein** [unser] DFK (LU_KLO_GP_DE)

3.2.3 Verbalphrase

Im Bereich der Verbalphrase gibt es zahlreiche Abweichungen von der Heimatland-Baseline. Diese können sowohl lexiko-semantischer (Beispiele (19)-(21)) als auch morphosyntaktischer (Beispiele (22)-(27)) Natur sein. Auch eine Kombination aus beidem ist möglich (Beispiel (18)). Im Deutschen wird eine Existenzkonstruktion mit dem Verb *geben* gebildet, welches in Verbindung mit dem Horizont-Pronomen *es* steht. Während in Beispiel (18) auch eine Existenzkonstruktion mit dem Verb *sein* möglich ist, ist das Horizontpronomen dennoch obligatorisch.

- (18) mal **war** eine Zeit wo sie das _ wo ganz wenige Schüler **waren**
[es gab mal eine Zeit wo es ganz wenige Schüler gab] (FL_HAM_GD_DE)
- (19) Dennis **jest** [ma] trzydzieści jeden [lat] (KL_BIE_GD_PL)
,Dennis ist einunddreißig‘
- (20) **zrobiłam** [ukończyłam] jeszcze dwa miesiące sprachkursu (BK_AUG_GD_PL)
,ich habe noch zwei Monate Sprachkurs gemacht‘
- (21) teď budeš zdravěj tak **z toho něco udělej** [tak toho využij] (KL_HEI_SP_DE)
,jetzt wirst du gesund also mach etwas draus‘

Diese Beispiele beruhen meist auf einer Polysemie in der Modellsprache die in der Replikasprache aber anders distribuiert ist. Diese Beispiele gilt es genauer zu untersuchen im Hinblick auf die Mechanismen, die hinter den Replikationen stecken, aber auch in Bezug auf die semantischen Felder häufig replizierter Konstruktionen.

Auch im Bereich der morphosyntaktischen Musterentlehnungen gibt es viele Beispiele ((22)–(27)) auf der Ebene der Verbalphrase:

- (22) **das** [daran] erinnere ich mich noch sehr sehr gut (FL_HAM_GD_DE)
- (23) aber wir wir Kinder waren da gezwungen **dessen** [uns das] zu hören [anzuhören]
(LC_BUR_GD_DE)
- (24) ale # **takie specjalne** jakieś **wyrażenia** [takich specjalnych jakichś wyrażeń] albo coś to
musiałeś się uczyć nie? (NT_LUB_GD_PL)
,aber irgendwelche solche speziellen Ausdrücke oder so musstest du lernen, nicht?‘
- (25) no a jsme se taky _ soukromě **mluvili** dobře **o o všechno** [o všem] (KL_HEI_SP_CZ)
,wir haben auch privat über alles gut geredet‘
- (26) da **sind wir** [wurden wir] schon auch als Nachbarsleute **ausgelacht** (LO_LAM_GD_DE)
- (27) aber Deutsch **war** [wurde] nicht **angeboten** (FL_HAM_GD_DE)

Während sich in den Beispielen (22) bis (25) die Rektion nach der Rektion des äquivalenten Verbs in der Modellsprache richtet, handelt es sich bei (26) um das Muster der polnischen Passivbildung, das im Deutschen repliziert wird. Des Weiteren wurden Beispiele, bei denen anstelle des Vorgangspassivs das Zustandspassiv gebildet wurde, ebenfalls als PAT VP annotiert (27).

Darüber hinaus gibt es bei der Verbalphrase auch Transfers im Bereich der Wortbildung, bei welchen beispielsweise Präfixe aus der Modellsprache in der Replikasprache repliziert wurden. So geht in (28) beispielsweise das Präfix *ab-* vor dem Verb *saugen* auf das Polnisch *od-* im Verb *odkurzyć* ‚(staub)saugen‘ zurück:

- (28) nicht am Sonntag # Auto waschen Haup* # Haus **absaugen** [saugen] _
(LO_LAM_GD_DE)

3.2.4 Präpositionalphrase

Viele Beispiele, die innerhalb einer Präpositionalphrase annotiert wurden (wie in Beispiel (29)), gehen auf den Artikelgebrauch im Deutschen zurück, analog zu den Beispielen in der Nominalphrase:

- (29) ober aber von [der] **andere**[n] **Seite** das # das einfach ist (BL_AUG_GD_DE)

Auch Beispiele für eine Replikation der Präposition ((30) und (31)), wie sie in der Modellsprache vorliegt, wurden in der Replikasprache als PAT PP annotiert. Dies ist besonders im Polnischen und Tschechischen, wie auch in vielen anderen slavischen Sprachen und Dialekten der Fall, in denen den präpositionalen Konstruktionen im Deutschen in der polnischen und tschechischen Heimatland-Baseline eine Nominalphrase im Instrumental entspricht, während die Bilingualen eine Präpositionalphrase verwenden:

- (30) taka pani się zatrzymała tam **z autem** [autem] (NT_LUB_GD_PL)
,eine Frau hat dort mit einem Auto angehalten‘

- (31) no tak _ jsem jel sám na pohřeb _ **s autem** [autem] (KL_HEI_SP_CZ)
,na so _ bin ich alleine zur Beerdigung gefahren _ mit dem Auto‘

Ähnliches trifft auch auf possessive Konstruktionen (Beispiel (32)) zu, die in der polnischen und tschechischen Heimatland-Baseline mithilfe des einfachen Genitivs ausgedrückt werden, während man dafür im Deutschen eine Präpositionalphrase mit *von* verwendet. Eine possessive Konstruktion mit Präposition wird auch von deutsch-polnischen und deutsch-tschechischen Bilingualen genutzt, sodass hier eine Replikation naheliegt:

- (32) jechała z nami **od mojej ci** * # no kuzynka moja **od mamy od starszej siostry**
(BQ_RAC_GP_PL)
,mit uns ist die Ta* # na meine Cousine von der Mutter von der älteren Schwester gefahren‘

3.2.5 Adverbialphrase

Auf der Ebene der Adverbialphrase kann es ebenfalls zu Transfers aus dem Polnischen und Tschechischen kommen:

- (33) aber nachdem [danach] haben sie [das]verboten das (AL_PAR_GP_DE)

Im Beispiel (33) verwendet der Sprecher das im Deutschen nicht in adverbialer Form gebräuchliche *nachdem*, dem das polnische Adverb ‚*potem*‘ als Muster zugrunde liegt: Die einzelnen Morpheme der polnischen Entsprechung *po+tem* werden dabei direkt ins Deutsche übersetzt.

In (34) greift der Sprecher im Tschechischen hingegen auf ein deutsches Muster zurück. Es handelt sich um eine Polysemie-Kopierung des Adverbs *dobře* ‚gut‘, das im Kontrast zu der Heimatland-Baseline auch in der prädikativen Position gebraucht wird:

- (34) a byl jsem **dobře** [dobrý] na škole _ na německé potom (II_ROD_SP_CZ)
 ‚und ich war gut in der Schule _ in der deutschen dann‘

Auch bei Lokal- oder Temporaladverbialen (Beispiele (35) und (36)) kann es aufgrund einer Polysemie in der Modellsprache zu Replikationen in der Replikasprache kommen:

- (35) die sind noch nicht gefahren **dort**[hin] (BT_KLO_GP_DE)
- (36) (takový co) jsem # **předem** [předtím] taky ještě nikdy neviděla (BJ_MUN_SP_CZ)
 ‚solche, die ich # vorher noch nie gesehen habe‘

3.2.6 Adjektivphrase

Bei Abweichungen in der Adjektivphrase handelt es sich zumeist um andere Abweichungen (siehe Abschnitt 3.4), die kein Muster in der Modellsprache haben und beispielsweise auf fehlende Kongruenz zwischen Adjektiv und Substantiv zurückzuführen sind.

Vereinzelt gibt es jedoch Beispiele, die durch die Polysemie der Adjektive in der Modellsprache zu erklären sind, wie im Beispiel (37).

- (37) tak levo levo levo levo levo _ a #_ to není už # **volné** [svobodné] tady (II_ROD_SP_CZ)
 ‚so links links links links _ und #_ es ist nicht mehr # frei hier‘

Die Verwendung von *volný* in der Bilingualen-Sprache in (37) ist auf die Polysemie des deutschen Adjektivs *frei* zurückzuführen, dessen Bedeutungen ‚unbesetzt‘ und ‚locker‘ in der tschechischen Heimatland-Baseline das Adjektiv *volný* abdeckt, während die Bedeutung ‚frei (von Freiheit)‘ durch das Adjektiv *svobodný* wiedergegeben werden kann. Diese Unterscheidung geht in der tschechischen Bilingualen-Sprache verloren.

3.3 Replikation des sprachlichen Materials

Von den als PAT bezeichneten Musterentlehnungen unterscheiden Matras & Sakel (2007: 829) die sog. *matter replication* (MAT), worunter sie „direct replication of morphemes and phonological shapes from a source language“ verstehen. Es geht um die Entlehnung konkreten phonologischen Materials aus der Modell- in die Replikasprache. In der englischsprachigen Literatur wird für diese Sprachkontakterscheinung der Begriff *borrowing* verwendet, der aufgrund seiner Metaphorizität nicht unproblematisch ist (vgl. Matras 2009: 146). Im Deutschen wird in dieser Bedeutung der Terminus *Transferenz* benutzt (vgl. Riehl 2014: 39).

Ausschlaggebend für die Einstufung einer Sprachkontakterscheinung als MAT war neben der phonologischen Form auch die morphologische Integration in die Replikasprache. Dieser Aspekt ist vor allem für die Abgrenzung zum Codeswitching von Bedeutung (siehe Abschnitt 3.5). In den deutschsprachigen Interviewteilen geschieht dies zum Beispiel durch die Voranstellung eines Artikels vor das transferierte Substantiv. So überträgt die Sprecherin in (38) die tschechische Bezeichnung *Šumava* in eine deutsche Klausel, obwohl das Deutsche über das Bedeutungsäquivalent *Böhmerwald* verfügt. Sie integriert das Wort mittels eines definiten Artikels (*die Šumava*). Den gleichen Vorgang kann man auch unter (39) bei einer materiellen Entlehnung aus dem Polnischen (*die Tajemnice*) beobachten.

- (38) es ist _ ich liebe die _ die die die Fahrt nach Budweis wenn ich über den Bayerischen Wald und über den _ also # **die die Šumava** [Böhmerwald] (KG_MUN_MI_DE)
- (39) und er hat die **Tajemnice** [Geheimnisse] (RN_BRO_GP_DE)

Im Tschechischen und im Polnischen werden die Entlehnungen unter anderem mittels Endungen der Nominaldeklinations integriert (vgl. Riehl 2014: 100). In (40) verwendet die Sprecherin in einer tschechischsprachigen Äußerung anstatt des tschechischen Lexems *škola* ‚Schule‘ das deutsche Lexem *Schule*, an das sie jedoch die Endung *-u* anhängt, die im Tschechischen den Akkusativ Singular bei Feminina markiert.

- (40) a já jsem neměla žádnou německou **\$schulu\$** (TU_HAG_SP_CZ)
 ‚aber ich hatte keine deutsche Schule‘

Im Beispiel (41) erfolgt die Integration ins Polnische durch das Anhängen eines polnischen Verbalsuffixes und einer Endung an den deutschen Verbstamm:

- (41) no ale potem myśmy też **flichtowali** [uciekali] (BQ_RAC_GP_PL)
 ‚na aber dann sind wir auch geflüchtet‘

Als MAT wurden ebenfalls die sog. Diskursmarker (*utterance modifiers*) annotiert. Darunter versteht Matras (1998: 282) „an extended grouping of various discourse-regulating elements, discourse markers, and focus particles“. Ihre Funktion ist laut Matras (ebd.: 324f.) das Herstellen und Aufrechterhalten der mentalen Kooperation zwischen den Kommunikationspartnern. Die Sprecherin oder der Sprecher verwendet sie, um die

Aufmerksamkeit der Hörerin oder des Hörers zu lenken und ihre bzw. seine Reaktionen zu antizipieren usw. Wie Matras (2009: 137) erklärt, kann es bei Bilingualen aufgrund des Fokussierens auf die mentale Kooperation mit dem Gesprächspartner dazu kommen, dass die oder der Bilinguale auf die Diskursmarker ihrer bzw. seiner pragmatisch dominanten Sprache zurückgreift. Das gewählte Set von Diskursmarkern wird dann laut Matras (2009: 145) unabhängig vom Interaktionskontext angewendet und kann daher letztendlich von einer ganzen Sprachgemeinschaft angenommen werden. Die Übernahme durch eine ganze Sprachgemeinschaft, wie sie beispielsweise im Tschechischen im Falle von *holt* ‚halt‘ stattfindet, ist jedoch in den von uns annotierten Klausen (noch) nicht gegeben. Vielmehr geht es um eine Verwendung, die sich auf bilinguale Sprecherinnen und Sprecher beschränkt. Dennoch kann man davon ausgehen, dass es sich in solchen Fällen nicht um Codeswitching, sondern um *matter replication* handelt, auch wenn es dafür keine äußeren Merkmale der Integration wie Flexion gibt.

In (42) wählt die Sprecherin beispielsweise die deutsche Partikel *eben*, um ihre Feststellung zu verstärken. In diesem sowie in analogen Fällen (43), (44) wurde MAT S annotiert. Über ähnliche Verwendungsweisen von Diskursmarkern berichten auch Blankenhorn (2003: 75ff.) in Bezug auf Russlanddeutsche und Hlavac (2006) im Zusammenhang mit Sprecherinnen und Sprechern des Kroatischen in Australien.

(42) a to bylo pro mě **(eben)** takový _ blbý (TU_HAG_SP_CZ)

‚und das war für mich **eben** so _ blöd‘

(43) **no** # es interessiert mich (KL_GDA_GP_DE)

(44) ale też nie aż takim żeby coś teraz po śląsku **vielleicht** mówić albo albo rozumieć (LC_BUR_GD_PL)

‚aber auch nicht so einer dass man jetzt auf schlesisch vielleicht sprechen oder oder verstehen‘

3.4 Andere Abweichungen²

Unter anderen Abweichungen (AA) verstehen wir in unserem Korpus diejenigen sprachlichen Erscheinungen, die gegen die Norm der Heimatland-Baseline verstoßen und für die kein Muster in der Modellsprache gefunden werden kann. Zum Teil können sie auf den indirekten Sprachkontakt, d. h. etwa auf die Attrition der Erstsprache infolge der Dominanz der Umgebungssprache, zurückgeführt werden. Zum Teil können sie dialektal begründet sein (siehe Jorroch, Księżyk & Zielińska in diesem Band).

² Der Begriff Abweichung wird in der vorliegenden Monographie wertneutral zur Bezeichnung von Differenzen verwendet, deren Ursachen uneindeutig sind. Damit wird keine Wertung impliziert – diese Kategorie wird ausschließlich deskriptiv angewandt.

3.4.1 Verbalphrase

Viele Abweichungen im verbalen Bereich betreffen die Rektion. Vor allem handelt es sich dabei um den Gebrauch des Objektes in der Akkusativ- oder Dativform, was aus dem Einfluss der Dialekte in Schlesien, West- und Ostpreußen resultiert (vgl. Jorroch, Księżyk & Zielińska in diesem Band). Beispielsweise kommen Personalpronomen, die im Akkusativ auftreten sollen, im Dativ vor oder umgekehrt, ohne dass dies auf ein Muster in der polnischen Sprache zurückgeführt werden kann. In den Beispielen (45) und (46) würden die Personalpronomen im Polnischen in demselben Kasus vorkommen, wie auch in der deutschen Baseline, nämlich in (45) in Akkusativ und in (46) in Dativ:

(45) aber **sie fanden ihm** [ihn] nicht (UM_JAN_GP_DE)

(46) na [mir] **mich will keiner glauben** (UM_JAN_GP_DE)

Abweichungen in der Verbalrektion, die nicht – wie im Falle von PAT – auf ein Muster in der Modellsprache zurückzuführen sind, kommen auch im Tschechischen vor. In (47) verwendet die Sprecherin das Verb *vyučit se* ‚erlernen‘ mit einem Objekt, dessen Form im Vokativ steht (*sestro*). Dies entspricht nicht der Baseline, in der entweder Instrumental oder eine Präpositionalphrase mit Akkusativ verlangt wird.

(47) a # potom jsem se právě učila # vyučila **zdravotní sestro** _ ano ? (IL_REG_SP_CZ)
 ,und # dann habe ich eben gelernt # Krankenschwester gelernt _ ja ?‘

Andere Abweichungen werden auch dann gekennzeichnet, wenn ein Verb untypisch konjugiert wurde. Das Paradigma kann dabei dialektal begründet werden wie in Beispiel (48) (siehe Jorroch, Księżyk & Zielińska in diesem Band), muss es aber nicht, wie in (49) gezeigt wird:

(48) dann **fruchen** [fragen] **sie** (UM_JAN_GP_DE)

(49) ich **nimm** [nehme] (UM_JAN_GP_DE)

Eine dialektale Erscheinung bildet auch die Nullendung in den Verbformen der ersten und zweiten Person im Singular und Plural des Präteritums und Konjunktivs im Polnischen. Dort, wo die Flexionsendungen *-m*, *-ś*, *-śmy*, *-ście* stehen sollten, taucht die Nullendung auf, so wie in (50). Die Sprecherin kommt aus Oberschlesien. Derartige analytische Formen der Verbflexion gelten nach Synowiec (1992: 61) als Kennzeichen des schlesisch-polnischen Regiolekt. In den Wohnort der Sprecherin sind allerdings nach dem Bevölkerungsaustausch nach 1945 Migrantinnen und Migranten aus den ehemaligen polnischen Ostgebieten und aus Zentralpolen gekommen, um sich dort niederzulassen (siehe Archiwum Państwowe w Katowicach [Staatsarchiv in Katowice]). Die Nullendung in den aufgezählten Verbformen ist ebenso ein dialektales Merkmal des nordöstlichen Polnischen (Joachimiak 1999: 28). Es zeigt sich somit, dass diese Erscheinung eine relativ großräumige Verbreitung in den polnischen Dialekten aufweist:

- (50) i my z Teresą nie **położyli** [położyłyśmy] się / tylko **siedzieli** [siedziałyśmy] nie ?
(BQ_RAC_GP_PL)

,und ich und Teresa haben uns nicht hingelegt sondern haben nur gesessen nicht?‘

Als eine indirekte Auswirkung des Sprachkontakts ist die Tendenz zur Bildung von Wörtern mit ungewöhnlichen Präfixen zu sehen (vgl. auch Riehl 2014: 115). In (51) wollte der Sprecher zunächst völlig auf ein Präfix verzichten. An dem Abbruch ist zu erkennen, dass es doch zu einer Präfigierung kam, welche allerdings nicht der Heimatland-Baseline entspricht, in der das präfigierte Verb *angefangen* heißt:

- (51) also # ich **habe** Berufsschule **ge* abgefangen** [angefangen] (SO_LAM_GD_DE)

Eine der Ursachen für die Entstehung der Abweichungen ist das Ökonomieprinzip. „In der Morphologie erzielt man Ökonomie meist durch eine Vereinfachung von Formen und Reduktion des Formenreichtums“ (Riehl 2014: 119). Im Polnischen steht das Genus des Subjekts mit dem Genus im Prädikat in einer Kongruenzbeziehung (im Präteritum). In unseren Daten gibt es aber Belege (wie (52) und (53)) für eine Verletzung der Kongruenz. Im Singular betrifft das alle drei Genera, im Plural wird das maskulin-personale mit dem nichtpersonal-belebten bzw. nichtpersonal-unbelebten Genus verwechselt:

- (52) tam tylko **część zostało** [została] (KL_GDA_GP_PL)

,es ist dort nur ein Teil geblieben‘

- (53) że jak z majątku **dziewczyny** musiały iść na rok do innego majątku / i **uczyli się** [uczyły się] od podstaw (KL_GDA_GP_PL)

,dass wenn die Mädchen aus einem Gut auf ein anderes Gut für ein Jahr gehen mussten / und dort von der Pike auf gelernt haben‘

Die Frequenz eines AA-Typs war in der Etappe der Annotation nicht relevant. Gekennzeichnet wurden auch seltene oder sogar einmalige sprachliche Erscheinungen, wie zum Beispiel die Verdoppelung der beweglichen Verbendung im Polnischen (Beispiel (54)). Diese Änderung könnte auch als ein Versprecher eingeordnet werden. Da sie aber deutlich von der Baseline abweicht, wurde sie annotiert:

- (54) to _ zawsze w środę i w niedzielę to po obiedzie **myśmy** w środę **przyszliemy** [przyszliśmy oder myśmy przyszli] ze szkoły obiad (KL_GDA_GP_PL)

,immer mittwochs und sonntags sind wir nach dem Mittagessen in der Schule am Mittwoch gekommen‘

Eine weitere Erscheinung ist die Spracherosion, die man auf der Basis des Abbaus von festen Wendungen erkennen kann. Im verbalen Bereich beruht die Abweichung auch auf der Verwendung eines untypischen Verbs im Rahmen eines Phraseologismus. In (55) und (56) verwendet der Sprecher eine Phrase *robić hobby* anstatt der typischen Kollokation *uprawiać hobby*, wie auch *pokazuje na* anstatt *wskazuje na*. Diese Innovationen sind auf kein Muster in der deutschen Sprache zurückzuführen:

- (55) sobota **mogę hobby robić** [uprawiać] nie no i to tak jest tego nie (SO_LAM_GD_PL)
 ‚am Samstag kann ich meinem Hobby nachgehen nicht und so ist es nicht‘
- (56) to nie nie **na** nic dobrego nie **pokazuje** [nie wskazuje] (SO_LAM_GD_PL)
 ‚das zeigt nichts Gutes an‘

Darüber hinaus können Abweichungen zum Teil als Beweis eines unvollständigen Spracherwerbs eingestuft werden. Abweichungen, die auf kein Muster in der Modellsprache zurückgeführt werden können, treten beispielsweise dann auf, wenn in der Modellsprache eine grammatische Kategorie fehlt, die in der Replikasprache vorhanden ist. In den polnischen und tschechischen Beispielen betrifft das die Abweichungen im Aspekt, der im Deutschen fehlt. In einem Interview erzählte eine Sprecherin über eine Sitte in ihrer Familie, dass die Enkelinnen unter Aufsicht der Großmutter regelmäßig auf Deutsch gelesen haben. Im zitierten Beispiel (57) verwendete die Sprecherin anstatt der imperfektiven Verbform *zaczynałyśmy*, das auf eine zyklische Tätigkeit verweist, die perfektive *zaczęłyśmy*, die eine einmalige Handlung bezeichnet:

- (57) pół godziny najpierw odpoczynek nie ? babcia się trochę położyła / a później przyszła / i już siedziałyśmy / i **zaczęłyśmy** [zaczynałyśmy] czytać (KL_GDA_GP_PL)
 ‚erstmal eine halbe Stunde Erholung nicht ? Oma hat sich kurz hingelegt / und später ist sie gekommen / und wir saßen schon / und wir begannen zu lesen‘

In dem analysierten Korpus treten jedoch keineswegs nur mehr oder weniger prototypische Fälle auf, sondern auch sprachliche Erscheinungsformen, die nicht eindeutig einer Kategorie zugeordnet werden können und Randphänomene etwa zwischen den Kategorien PAT und AA darstellen.

Grenzfälle dieser Art sind beispielsweise die von der deutschen Standardsprache abweichenden Verwendungen des Verbs *mögen* im Konjunktiv II. Bekanntlich wird der Konjunktiv Präteritum dieses Modalverbs, d. h. *möchte*, der nicht als Konjunktiv sondern als Indikativ empfunden wird, im Standarddeutschen voluntativ zum Ausdruck eines Wunsches in der Gegenwart gebraucht (vgl. Helbig & Buscha 1996: 132). In dem gesammelten Korpus kommen jedoch auch Äußerungen (Beispiele (58)–(61)) vor, in denen *möchte* konditional verwendet wird:

- (58) und wir sind ja auch etliche Male noch / hatten Urlaub / und so sind wir auch ganz Deutschland rumgereist / also ich **mechte** [würde] mich da nicht verkaufen lassen (SL_OPO_GP_DE)
- (59) ich **möchte mich** [würde mir] wünschen eine deutsche Beerdigung (SV_NAK_GP_DE)
- (60) dann wenn die # wenn [das] ganze Polen die Regierung **möchte i * # anders mit den Leuten sich umgehen** [anders mit den Leuten umgehen würde] / da fahren sie auch nicht raus (BQR_RAC_GP_DE)

- (61) # **wenn er möchte leben / dann möchte er** einen Kontakt mit meinem Vater # später schon machen ne # [wenn er leben würde, dann würde er später Kontakt mit meinem Vater aufnehmen] (LU_KLO_GP_DE)

In der einschlägigen Literatur werden diese sprachlichen Erscheinungen unterschiedlich gedeutet: Chmiel (1988: 123–124) stuft sie als grammatische Interferenz ein, Pelka (2006: 156–157) betrachtet sie als sprachkontaktbedingte Funktionsänderung von Lexemen mit syntaktischer Funktion. Obwohl Jorroch (2015: 172) zunächst nach einer sprachhistorischen Erklärung dafür sucht, zieht sie letztendlich den Schluss, es handle sich dabei um einen slavischen Einfluss: „Der Verwendungsbereich des potenziellen Konjunktivs im Polnischen hat ebenfalls zur Veränderung im Gebrauch von anderen sprachlichen Mitteln der Konjunktivform im Deutschen“ beigetragen. Zuweilen werden allerdings auch in den einschlägigen Wörterbüchern andere Bedeutungsvarianten von *möchte* angegeben, beispielsweise heißt es im Wörterbuch der deutschen Gegenwartssprache: *mögen* „drückt aus, daß der Sprecher den Inhalt des Inf. als Möglichkeit, Vermutung hinstellt“ und zur Veranschaulichung wird u. a. folgendes Beispiel angeführt: „spränge ich ihr nach, so möchte ich wohl beide noch einholen Le Fort Papst 283“ („mögen“ in: DWDS). Angesichts dieser unterschiedlichen Erklärungswege werden diese Konstruktionen nicht vorab als PAT, sondern als AA klassifiziert, unter dem Vorbehalt, dass sie einer tieferen Analyse bedürfen.

3.4.2 Nominalphrase

Im Polnischen und im Deutschen besteht eine große Gruppe von unbelebten Substantiven, bei denen das grammatische Genus mit dem natürlichen Geschlecht nicht im Einklang steht. Dabei stimmen sie in beiden Sprachen nicht immer überein. Auf dieser Basis kann es zu verschiedenen Abweichungen im Bereich der Genera kommen. Wenn die Informantinnen und Informanten den Substantiven ein falsches Genus unabhängig von der Kontaktsprache zugewiesen haben, wurde das als AA markiert. Beispielsweise hat die Sprecherin in (62) dem Nomen *Zug* das neutrale Genus gegeben, obwohl sein Äquivalent im Polnischen *pociąg* dasselbe maskuline Genus hat wie in der deutschen Baseline. Darüber hinaus wurden auch die ungewöhnlichen Genuszuweisungen als AA annotiert, die mit einem regionalen Gebrauch motiviert werden konnten. In (63) hat die Sprecherin *den Radio* gesagt. In Süddeutschland, Österreich und der Schweiz ist das Maskulinum korrekt, wenn mit dieser Wortform ein Rundfunkgerät gemeint ist („Radio“ in: DWDS). In der polnischen Baseline hat das Äquivalent für *Radio* – *radio* das neutrale Genus, was der deutschen Baseline gleich ist:

- (62) und **dieses** _ [**dieser**] **Zug** immer fuhr von Rastenburch nach Angerburch
(UM_JAN_GP_DE)
- (63) ich werde ausmachen **den** [das] **Radio** (UM_JAN_GP_DE)

Einige Abweichungen können unterschiedlich interpretiert werden und ihre Interpretation wird mit unterschiedlichen Sprachkontaktphänomenen begründet. Im Polnischen

der bilingualen Sprecherinnen und Sprecher kommt es zur Verwechslung des (in (64) deklinierten) maskulin-personalen Pronomens *oni* mit dem nichtmaskulin-personalen Pronomen *one*. Anstelle der nichtmaskulin-personalen Akkusativform *je* steht die maskulin-personale Form *ich*. Dies kann entweder als Resultat der Vereinfachung des Paradigmas gesehen werden oder dialektal begründet sein (Joachimiak 1999: 28) und auf einen Spracherwerb durch den Kontakt mit Sprecherinnen und Sprechern aus den ehemaligen polnischen nördlichen Ostgebieten hindeuten, die nach 1945 in die westlichen Territorien gekommen sind. In (64) kann diese Erscheinung als Resultat der Vereinfachung des Paradigmas verstanden werden. Das bestätigen die Verbformen. Zunächst verwendet die Sprecherin die mit dem Subjekt *dziewczyny* kongruente nichtmaskuline Form *musiały*. Im nächsten Satz vergisst sie das Genus desselben Subjekts *dziewczyny* und benutzt die maskuline Form *ucz yli*, was die Kongruenz verletzt. Beim Übergang zu den maskulinen Formen ist sie konsequent und benutzt weiter das maskuline Personalpronomen *ich*, das sich auf das Subjekt *dziewczyny* bezieht:

- (64) że jak z majątku **dziewczyny musiały** iść na rok do innego majątku / i **ucz yli** [ucz yły] się od podstaw / to znaczy **ich** [je] jak ścielić łóżko (KL_GDA_GP_PL)
 ‚dass wenn die Mädchen aus einem Gut auf ein anderes Gut für ein Jahr gehen mussten / und dort von der Pike auf gelernt haben / das heißt sie wurden unterrichtet wie man das Bett macht‘

Ebenfalls als AA werden Fälle annotiert, in denen im gesprochenen Polnisch Demonstrativpronomina vor Orts- oder Länderbezeichnungen stehen (Beispiele (65) und (66)). Weder sind im Polnischen solche Formulierungen usuell, noch ist dabei auf einen Einfluss des Deutschen zu schließen, wo bei solchen Konstruktionen der Nullartikel üblich ist:

- (65) bo ona się też urodziła tu jeszcze w **tym Opolu** w Zakrzowie (SL_OPO_GP_PL)
 ‚weil sie auch noch hier geboren wurde in Opole in Zakrzów‘
 (66) ale w **tym Dobrzyniu** ta moja ciotka pisała (BQ_RAC_GP_PL)
 ‚aber in Dobrzeń schrieb meine Tante‘

Auch im Tschechischen treten Abweichungen im Bereich der Deklination von Pronomina auf. Zum Beispiel verwendet der Sprecher in (67) die Pluralform des Demonstrativpronomens *ty* ‚die.F.PL.ACC‘ anstatt der Singularform *tu* ‚die.F.SG.ACC‘. Folglich kongruiert die gewählte Form nicht mit dem Substantiv *práci* ‚Arbeit.F.SG.ACC‘.

- (67) no jo _ tak s* dostal tam **ty** [tu] práci (KL_HEI_SP_CZ)
 ‚na ja _ so hat er dort die Arbeit bekommen‘

Darüber hinaus kommt es auch zur Bildung von normabweichenden Formen bei der Deklination der Nomina. Dabei ist jeder Beleg getrennt zu analysieren und schwer zu gruppieren. In (68) hat die Sprecherin die Endung *-s* zu dem Nomen *Pfarrer* hinzugefügt, das mit *-er* im Auslaut standardmäßig keine Endung im Plural bekommt (Helbig & Buscha 2005: 216). Aufschlussreich könnte hierbei eine Analyse am gesamten Korpus

in Hinsicht darauf werden, ob derartige Beispiele bei den Sprecherinnen und Sprechern vereinzelt vorkommen oder gar eine Tendenz zur deutlichen Pluralmarkierung aufzeigen, wie sie etwa für oberschlesische Sprachinselvarietäten des Deutschen verzeichnet wurde (vgl. Księżyk 2008: 208–209):

- (68) sonst sind nicht zu viel **Pfarrers** [Pfarrer] so dies (UM_JAN_GP_DE)

Das Beispiel (69) stellt einen Grenzfall zwischen AA und PAT dar. Die Sprecherin gibt in Kilometern bekannt, wie weit sich hinter Olsztyn die Ortschaft Szalstry befindet. In der Aussage der Sprecherin ist die normabweichende Form *kilometry* auffällig. Vergleichen wir das (auf die in beiden Sprachen auftretenden Kasus beschränkte) polnische und das deutsche Deklinationsmuster im Plural:

- N. *kilometry* Kilometer
 G. *kilometrów* Kilometer
 D. *kilometrom* Kilometer
 A. *kilometry* Kilometer

Das besprochene Substantiv sollte im Polnischen im Genitiv *kilometrów* gebraucht werden, im Deutschen sollte es im Nominativ stehen, wobei im Deutschen alle Formen konvergent sind. Es könnte hier entweder zu der Übertragung der Konvergenz (PAT) kommen oder zur Vereinfachung des Paradigmas (AA) oder aber zu einem standardabweichenden Gebrauch des Substantivs im Nominativ bzw. im Akkusativ (AA):

- (69) Szalstry dwadzieścia osiem *kilometry* [kilometrów] za Olsztynem (LA_MIK_GP_PL)
 ‚Szalstry ist achtundzwanzig Kilometer hinter Olsztyn‘

Andere Abweichungen stellen auch die folgenden sprachlichen Phänomene dar, bei denen ebenfalls keine direkte Nachbildung sprachlicher Muster aus der Kontaktsprache vorliegt – normgerecht müsste in (70) und (71) der Nullartikel gebraucht werden, die Abweichungen erscheinen somit als Konsequenz einer Hyperkorrektheit:

- (70) no und ich wu * habe mein Ziel im Leben gehabt / wollte **die Lehrerin** [wollte Lehrerin] werden (BQ_RAC_GP_DE)
 (71) Romane interessieren mich nicht / aber wenn das was **mit der Geschichte** [mit Geschichte] ist / bringt das nach Hause ja (RN_BRO_GP_DE)

Da eines unserer Annotationsprinzipien besagte, dass nur die vorhandenen Erscheinungen und nicht die fehlenden markiert werden dürfen (siehe Abschnitt 3.1), wurden in den Aussagen auch die Nominalphrasen markiert, anstelle derer sich eine Präpositionalphrase befinden sollte, wie in (72):

- (72) wir haben [am] **Anfang** haben wir was war da # acht # Milchkühe und und # dann die Kälbe (SO_LAM_GD_DE)

3.4.3 Adjektivphrase

Abweichungen im adjektivischen Bereich zeigen auf Unsicherheit der Sprechenden bezüglich Genera. Im Polnischen kommt es zur Inkongruenz zwischen den Wortarten, die in Hinblick auf das Genus veränderbar sind. Beispielsweise verwendet die Sprecherin KL_GDA_GP in (73) die standardgemäße nichtmaskulin-personale Verbalendung im Plural *chodziły* und in (74) die standardgerechte feminine Verbendung im Singular *byłam*. Trotz richtiger Verbalendungen verwendet sie aber inkongruente Adjektivendungen: in (73) nämlich eine maskulin-persönliche Endung im Plural *ubrani* und in (74) eine maskuline Endung im Singular *wykszałcony*.

(73) to **chodziły** bardzo skromnie **ubrani** [ubrane] (KL_GDA_GP_PL)
 ‚sie waren sehr bescheiden gekleidet‘

(74) a ja to wszystko to ja **byłam** super **wykszałcony** [wykształcona] (KL_GDA_GP_PL)
 ‚und ich das alles ich war super ausgebildet‘

Adjektive sind in beiden Sprachen deklinierbar und ihre Deklination ist variabel. Die große Vielfalt der Flexionsendungen ist ein Grund für die Anwendung des Ökonomieprinzips und Vereinfachung des Paradigmas sowohl im Deutschen als auch im Polnischen. Zum Beispiel benutzt der Sprecher in der deutschen Sprache in (75) wie auch in (76) im Plural und im Singular dieselbe Endung *-e*, was von der Baseline abweicht. In (77) gleicht er die Endung des Adjektivs dem Indefinitpronomen *alles* an, sodass die beiden Endungen (des Adjektivs und des Pronomens) konvergent sind:

(75) ja ja **die ganze[n]** **zwei** Jahre (SO_LAM_GD_DE)

(76) ja aber das ist kein **kein # richtige[s]** **Deutsch** (SO_LAM_GD_DE)

(77) und gorole das war **alles anderes** [andere] als Oberschlesien also aus Katowice Gleiwitz (SO_LAM_GD_DE)

Im Polnischen benutzt die Sprecherin in (78) anstatt der Adjektivform *lewo* die Form des Adverbs (w lewo/na lewo), dessen Endung *-o* konvergent mit der Endung des Substantivs *oko* ist. Es liegt hier ein ähnlicher Konvergenzprozess wie in der deutschen Sprache in (77) vor. In (79) ist die Pluralendung des Adjektivs *mazurskie* mit der Singularform des Substantivs inkongruent:

(78) no i komputerowo się okazuje że to **lewo** [lewe] oko (IL_PIL_II_GP_PL)
 ‚und es stellt sich per Computer heraus, dass es das linke Auge ist‘

(79) tak tak i bo tu był **język mazurskie** [mazurski] (LA_MIK_GP_PL)
 ‚ja ja und weil es hier eine masurische Sprache gab‘

Im Tschechischen kommt es in der Adjektivphrase u. a. zu anderen Abweichungen im Bereich der Wortbildung. In (80) verwendet der Sprecher das Adjektiv *azylovským* ‚Asyl-‘ mit einer erweiterten Suffixvariante *-ovský* (vgl. Karlík, Nekula & Rusínová 1995: 166–167), anstatt das Suffix *-ový* zu nutzen. Das Suffix *-ovský* wird in der Heimatland-Baseline jedoch vor allem bei Propria verwendet (Štícha et al. 2013: 204–205).

- (80) já jsem dokonce i s tím _ **azylovským** [azylovým] pasem jsem letěl do Maďarska (KC_ISM_MI_CZ)
 ‚ich bin sogar mit dem _ Asylpass nach Ungarn geflogen‘

Ein ähnlicher Fall ist die folgende Bildung *válkovej kámoš* ‚Kriegskamerad‘. Der Sprecher leitet von dem Substantiv *válka* ‚Krieg‘ ein in der Baseline nicht existierendes Adjektiv ab, das dem ersten Glied des deutschen Kompositums entspricht. Die in der Heimatland-Baseline zulässigen Varianten wären etwa *válečný kamarád* oder *kamarád z války* ‚ein Kamerad aus dem Krieg‘. Das von dem Respondenten eingesetzte Suffix *-ový* (bzw. seine umgangssprachliche Variante *-ovej*) ist in seiner Funktion komplementär zu dem in der Baseline-Variante vorgeschlagenen *-ný* (Štícha et al. 2013: 201). Wie jedoch Štícha et al. (ebd.) anmerken, wird *-ový* zurzeit häufiger als *-ný* bei Neologismen verwendet, was die Wahl dieses Suffixes in (81) teilweise erklären kann. Einen weiteren Faktor könnte die Morphonologie darstellen. In dem Adjektiv *válkovej* kommt es im Gegensatz zu *válečný* zu keiner Alternation *k–č*, was das Derivationsverfahren für den Sprecher weniger komplex macht.

- (81) jak mi to (řikal) tak řikal _ no jo _ **válkovej kámoš** [kamarád z války] tak řikal (KL_HEI_SP_CZ)
 ‚wie es mir (sagte) so sagte er _ na ja _ Kriegskamerad so sagte er‘

3.4.4 Präpositionalphrase

Auch in der Präpositionalphrase kommt es zur Bildung verschiedener Typen von Abweichungen, u. a. zur Verwendung einer von der Norm abweichenden Präposition, die nicht auf ein Muster in der Modellsprache zurückzuführen ist. In (82) benutzt die Sprecherin die Präposition *bei* anstatt *zu*, in (83) *an* anstelle von *in*:

- (82) aber wie ich [zu] **bei meine[r] Schwester gefahren bin** (UM_JAN_GP_DE)

- (83) und im Frühjahr nachdem haben wir ausgesammelt [in] **an der Mitte** (UM_JAN_GP_DE)

Diese Beobachtung bestätigt sich auch in den tschechischsprachigen Daten wie in (84), in denen zum Beispiel die Präposition *v* ‚in‘ mit Lokativ im Zusammenhang mit der Verwendung von *počítač* ‚Computer‘ verwendet wird, bei denen in der Heimatland-Baseline *na* ‚auf‘ mit dem gleichen Kasus vorkommt.

- (84) nebo teď taky # v [na] počítači že má* že # že si píšeme (CI_BUB_SP_CZ)
 ‚oder jetzt auch # im Computer dass wir h* dass wir uns schreiben‘

Es werden auch Präpositionalphrasen gebildet, die in keiner von beiden Sprachen vorkommen. In (85) bildet die Sprecherin eine aus einer Präposition *w* und einem Pronomen *kiedyś* gebaute Präpositionalphrase *w kiedyś* anstelle des alleinstehenden Pronomens *kiedyś*. Im Deutschen würde man diese Phrase mit dem Adverb *damals* übersetzen. Es besteht also kein Muster, worauf dies zurückzuführen wäre:

- (85) złotnik a **w kiedyś** [kiedyś] to nie że tylko jeden to pięć lat się uczyło po maturze (KL_GDA_GP_PL)
 ‚Goldschmied und damals hat man nicht nur ein Jahr sondern fünf Jahre lang nach dem Abitur gelernt‘

Zu einer ungewöhnlichen Abweichung kommt es in einer Aussage des Sprechers SO_LAM. In (86) benutzt er die Präpositionalphrase *na urzędnika* anstelle des Dativobjekts *urzędnikowi* in Verbindung mit dem Verb *zrobić*. Die Präposition *na* wurde wahrscheinlich von dem Idiom *nie poradzić nic na coś* genommen, was die Negation und das Pronomen *nic* andeuten:

- (86) no **na na urzędnika** [urzędnikowi] nie zrobisz nic (SO_LAM_GD_PL)
 ‚einem Beamten kannst du nichts anhaben‘

Es kann auch innerhalb einer PP zum abweichenden Kasusgebrauch kommen (Beispiele (87) und (88)). Hierbei können Fälle, in denen Akkusativformen anstelle von Dativformen (und umgekehrt) verwendet werden, auch auf eine Dialektkontinuität hindeuten (siehe Jorroch, Księżyk & Zielińska in diesem Band):

- (87) mit dem _ mit der Mutter haben wir Deutsch gesprochen bloß **mit die** [den] **Kindern** (UM_JAN_GP_DE)
 (88) dann fuhr ich **nach sechsendvierzig Jahre**[n] ihr [sie] besuchen (UM_JAN_GP_DE)

Der abwechselnde Gebrauch von Dativ und Akkusativ bei der Positions- und Richtungsangabe kann ebenfalls dialektal bedingt sein. An Stellen, an denen Dativ zu erwarten wäre, wie in (89), wird Akkusativ verwendet. An Stellen, an denen Akkusativformen zu erwarten wären, wie in (90), verwenden die Sprecherinnen und Sprecher Dativformen (siehe Jorroch, Księżyk & Zielińska in diesem Band):

- (89) aber sie **war in die** [der] **Kirche** wie war gewesen Konfirmation (UM_JAN_GP_DE)
 (90) die **kamen auf** [das] **dem Feld** (UM_JAN_GP_DE)

3.4.5 Adverbialphrase

In der „Generation Polen“ kommen häufig dialektal motivierte adverbiale Abweichungen vor. So wird das Adverb *dann* durch die Konjunktion *denn* wie in (91) ersetzt (vgl. Jorroch, Księżyk & Zielińska in diesem Band):

- (91) **denn** [dann] fahren wir mit dem Auto Anfang Deutsch dann immer Polni *
(UM_JAN_GD_DE)

Es kann auch zu anderen, nicht dialektal begründeten Normabweichungen kommen, wie zum Beispiel zur Anwendung des Adverbs in der Grundstufe *bardzo* anstatt des Adverbs *dużo* wie in (92) oder zur Verwendung des Interrogativums *wo* anstatt *woher* wie in (93):

- (92) **bardzo** [dużo] **wcześniej** tak jeszcze w oś * w dziewięćdziesiątym szóstym roku tak
(NA_HAM_GD_PL)

,viel früher so noch im ach * im Jahr sechsundneunzig ja‘

- (93) **von wo** [woher] die da kamen so (UM_JAN_GP_DE)

In den tschechischsprachigen Daten findet sich zum Beispiel in (94) die Verwendung von *nikde* ‚nirgendwo‘ statt *nikdy* ‚nie‘, die auf eine Unsicherheit im Paradigma der Negationswörter hindeutet.

- (94) a teďka jsem řekla potom sem _ tady _ nic _ tady _ sem _ už vůbec _ **nikde** _ [nikdy]
nepřijdeš (TU_HAG_SP_CZ)

,und jetzt sagte ich denn hierher _ hier _ nichts _ hier _ hierher _ gar nicht mehr _ nirgendwo _ wirst du kommen‘

3.4.6 Satzebene

Als AA wurden auf der Satzebene Fälle eingeschätzt, in denen der Satz zwar grammatisch korrekt ist, seine Entschlüsselung jedoch ohne einen breiteren Kontext und die Miteinbeziehung des Weltwissens durch die Rezipientin oder den Rezipienten unmöglich wäre. In (95) will die Sprecherin sagen, dass sie sich beim Lesen auf Deutsch konzentrieren und viel nachdenken muss. Stattdessen sagt sie *z głową muszę mieć*. Es kann hier eine Verwechslung zweier Idiome vorliegen: *robić coś z głową* und *mieć coś z głową*. Das erste Idiom mit dem Verb *robić/machen* bedeutet *mitdenken*. Das zweite Idiom mit dem Verb *mieć/haben* bedeutet hingegen *nicht recht bei Verstand sein*:

- (95) z * z głową muszę mieć [muszę się zastanowić] (IL_PIL_II_GP_PL)

,ich muss es mit dem Kopf haben [ich muss es mir überlegen]‘

In (96) betont die Sprecherin, dass die Brille, die sie gerade trägt, eine Fernbrille ist und ihr das Lesen dennoch nicht erleichtert. Sie will sagen, dass ihre Brille die Buchstaben nicht vergrößert, auf Polnisch – *przybliżać, powiększać*. Stattdessen sagt sie, dass die Brille nicht heranzieht, auf Polnisch – *przyciągać*:

- (96) teraz na ten to je * są okulary na dal / **nie przyciąga mi** [nie powiększają]
(IL_PIL_II_GP_PL)

,jetzt dafür das sin * das ist eine Weitsichtbrille / sie zieht es mir nicht‘

Die Label AA S wurden auch bei Normabweichungen bei Konjunktionen verwendet, zum Beispiel zur Kennzeichnung des Interrogativadverbs *wann* anstelle der Konjunktion *wenn*, wie in (97):

- (97) wann [wenn] war er hier # / der der kl * der jüngste Sohn / der war ein halbes Jahr
(FL_BON_GP_DE)

3.5 Codeswitching

Unter Codeswitching wird grundsätzlich das Wechseln von einer Sprache bzw. Varietät in eine andere innerhalb einer Äußerung verstanden. Was genau Codeswitching umfasst und wie es von anderen Sprachkontaktphänomenen abgegrenzt werden kann, ist umstritten (vgl. etwa Gardner-Chloros 2009: 10ff.). In unserem Projekt war in dieser Hinsicht vor allem die Differenzierung zwischen Codeswitching und Replikation des sprachlichen Materials (Abschnitt 3.3) wichtig, auf die weiter unten eingegangen wird.

Anhand der Prozesse, die für das Wechseln der Sprache bezeichnend sind, unterscheidet Muysken (2000: 3) zwischen dem Einfügen lexikalischen Materials in eine anderssprachige Äußerung (*insertion*) und dem Wechsel zwischen den Äußerungen bzw. Klausen (*alternation*). Sowohl *insertional* (in (98) in der Nominal- und (99) in der Präpositionalphrase) als auch *alternational* (in (100) auf Satzebene) *code-switching* kommen in unseren Daten vor.

- (98) dann kam ich mit sechs wie jeder andere auch in die Schule
damals hieß es **základní _ škola _ devítiletá základní devítiletá # letá škola**
(KC_ISM_MI_DE)
,damals hieß es **Grundschule _ neunjährige Grundschule neunjährige Schule**‘
- (99) ah wo ? ich konnte ja nicht **po polsku** [Polnisch] von wo ? (KQ_GOG_GP_DE)
- (100) a to jsem neměla hodně kontaktu k tomu proč **wieso warum _ weiß ich nicht**
(TU_HAG_SP_CZ)
,und da hatte ich nicht viel Kontakt dazu warum **wieso warum _ weiß ich nicht**‘

Beim Einfügen einzelner Wörter in die Matrixsprache handelt es sich unter anderem um Unika, die beispielsweise Institutionen oder bürokratische Prozedere bezeichnen, die mit dem Kontext der anderen Sprache verbunden sind. Wie Matras (2009: 107) erklärt, assoziiert eine solche Bezeichnung das ursprüngliche Setting, was die Authentizität der Beschreibung erhöht. Von solchen Beispielen wurden im Projekt diejenigen Fälle unterschieden, in denen auf Konzepte referiert wird, die in der Matrixsprache nicht vorhanden sind (vgl. Myers-Scotton 1997: 16), wie in (101). Unter Matrixsprache wird die Sprache verstanden, in die der Wechsel eingebettet wird. In (101) erwähnt die Sprecherin den Laden *Tuzex*, in dem zur Zeit des Sozialismus in der Tschechoslowakei exklusive Westware verkauft wurde. Diese mit der Kultur der eingebetteten Sprache zusammenhängenden Begriffe wurden in unserem Korpus nicht annotiert.

(101) aber das hat man dann nur in **Tuzex** gekriegt (EQ_HER_MI_DE)

Ähnlich wurden auch Orts- oder andere Eigennamen (Beispiele (102) und (103)) aus der Kontaktsprache nicht als CS eingestuft, auch wenn dafür in der Basissprache durchaus Entsprechungen vorliegen.

(102) na ja da muss man nach **Warszawa** [Warschau] fahren nicht (BQ_RAC_GP_DE)

(103) die Schule also in Cosel das war **Technikum Żegluga** [Schiffahrtstechnikum]
(NT_LUB_GD_DE)

Das Beispiel (98) oben zeigt das sog. funktionale Codeswitching (Riehl 2014: 25). Dieses wird aus diskursstrategischen Gründen eingesetzt (vgl. Auer 1995: 120) und kann – wie in (98) – von dem Sprecher explizit eingeleitet werden (*damals hieß es*). Das Vorkommen des konversationell bedingten Codeswitchings in unseren Daten hängt mitunter damit zusammen, dass die Interviews, die die Grundlage des Korpus bilden, mit bilingualen Interviewerinnen und Interviewern geführt wurden (vgl. Jorroch & Prawdzcic und Bučková & Nekula in diesem Band).

Das nicht-funktionale Codeswitching ist wiederum psycholinguistisch motiviert und wird von den Sprecherinnen und Sprechern nicht beabsichtigt (Riehl 2014: 29). Es wird von sog. *trigger words* ausgelöst (Clyne 1967: 84, 2003: 162ff.). Dies können beispielsweise in beiden Sprachen identisch oder ähnlich klingende Wörter sein, wie *perfekt* im Deutschen und *perfektni* im Tschechischen (siehe (104)).

(104) ty n* ty nejsou tak per* **perfekt** [perfektni] ne ? _ hmm (LS_ZEI_SP_CZ)
,die sind auch nicht so per* **perfekt** ne ? _ hmm‘

Das durch ein *trigger word* eingeleitete Codeswitching kann sich über mehrere aufeinanderfolgende Klausen erstrecken (für die Bestimmung von Klausengrenzen siehe Bučková & Prawdzcic in diesem Band). In solchen Fällen wurde CS S in allen solchen Klausen annotiert.

Die Abgrenzung von Codeswitching und Transfer (MAT) gehört zu den häufig diskutierten Themen auf dem Feld der Sprachkontaktforschung, wovon auch die Einführung unterschiedlicher Begriffe wie Ad-hoc-Entlehnung (Riehl 2014: 22) oder *nonce-borrowing* (Poplack, Sankoff & Miller 1988) zeugt. Als Kriterien der Unterscheidung nennt Matras (2009: 106) die höhere Frequenz von Entlehnungen im Kontrast zum spontanen Codeswitching und den Grad der morphologischen Integration von MAT in der Replikasprache. Letzteres war, wie oben beschrieben, für die Annotation von MAT im LangGener-Projekt ausschlaggebend.

Grundsätzlich kann die Beziehung zwischen Codeswitching und Entlehnung als diachron angesehen werden (Backus 2020: 192), das heißt Lexeme, die in der Sprache der Bilingualen aus einer Modell- in eine Replikasprache anfangs spontan übernommen werden, können mit der Zeit in der Replikasprache als Lehnwörter etabliert werden. Aufgrund des Forschungsdesigns des LangGener-Projektes war jedoch ein Verfolgen einer diachronen Perspektive, die eine klare Abgrenzung von MAT und CS ermöglichen könnte, nicht gegeben. Stattdessen orientierten wir uns an einer Reihe von Indizien im Sprachmaterial, die im Folgenden vorgestellt werden.

Bei Äußerungen, in denen der Wechsel in eine andere Sprache etwa durch metasprachliche Kommentare oder Häsitiation begleitet wurde, wurde CS annotiert. Es handelt sich um sog. *flagged code-switching* das unter anderem von Wortfindungsproblemen zeugen kann (Bolonyai 2009: 260). In (105) wechselt die Sprecherin beispielsweise ins Deutsche, worauf ein Häsitationsgeräusch und die an die Interviewerin gerichtete Frage, wie *Teppich* auf Tschechisch heißt, folgen. Falls der Wechsel auf diese oder eine ähnliche Weise hingegen nicht reflektiert wird, wurde MAT annotiert (vgl. (38) und (39)).

- (105) voni dělají **Teppich # jak se to jmenuje Teppich na česky ? #** (LS_ZEI_SP_CZ)
 ‚sie machen **Teppich # wie sagt man das Teppich auf Tschechisch ? #**‘

Ähnlich signalisiert das Zögern in der folgenden polnischsprachigen Äußerung (106) ein mit Wortfindungsproblemen verbundenes Codeswitching:

- (106) tam było wszystko w socjalizmie nie było jakiegoś # **Wettbewerb** jak się nazywa ?
 rywalizacja ? (XL_MUN_GD_PL)
 ‚Da gab es im Sozialismus alles, es gab keinen #Wettbewerb wie heißt das ? Rivalität ?‘

Wenn der Teil der Äußerung, in dem die Sprache gewechselt wurde, in beiden Sprachen nacheinander genannt wurde, wurde CS annotiert (vgl. (107)).

- (107) oder # so Plisseerock so **\$sterilinová\$ sukně** # das war absoluter Hit (EQ_HER_MI_DE)
 ‚oder # so Plisseerock so **Plisseerock** # das war absoluter Hit‘

Sowohl CS als auch MAT konnten ebenso beim Wechsel in eine Sprache annotiert werden, die nicht zu den im Projekt untersuchten Sprachen gehört. Beispielsweise wechselt in (108) die Sprecherin vom Deutschen ins Englische.

- (108) dann war ich _ dann bin ich zu den Amerikanern i* in diese # **personal personal office**
 (EQ_HER_MI_DE)

3.6 Wortstellung

Mit dem Marker Wortstellung (*WO/word order*) werden sämtliche Abweichungen von den Wortstellungsregularitäten der jeweiligen Sprache annotiert, unabhängig davon, ob sie aus Sprachkontakt resultieren oder anders bedingt sind. Die Entscheidung für eine gesonderte Kategorisierung der Wortstellungsabweichungen statt ihrer Zuordnung zu PAT oder AA erschien aus mehrfacher Hinsicht sinnvoll. Da sich das Deutsche und die polnische bzw. tschechische Sprache topologisch voneinander unterscheiden, auch im Hinblick auf die sog. Festigkeit der Wortstellung, die als entscheidend für die Nachbildung der Stellungsregularitäten gilt, erlaubt die gesonderte Klassifizierung der Wortfolgeabweichungen diesen Einflussfaktor am Korpus spezieller zu untersuchen und zudem den Anteil sprachkontaktgestützter im Verhältnis zu anderen Wortstellungsabweichungen zu ermitteln (vgl. Hinc 2010: 144). Was die Festigkeit der Wortstellung anbelangt, so ist die Wortfolge im Deutschen bekanntlich viel strenger geregelt als in der polnischen und tschechischen

Sprache (vgl. Engel et al. 2018: 391; Štícha 2003: 122). Die Tatsache, dass das Polnische über weit mehr Stellungsmöglichkeiten als das Deutsche verfügt, wird als der wichtigste Unterschied zwischen den beiden Sprachen betrachtet. Genau genommen geht es dabei darum, dass das Deutsche viel mehr grammatisch determinierte Abfolgeregeln als das Polnische und Tschechische aufweist, während die polnische und tschechische Sprache stärker kommunikativ-pragmatischen Prinzipien folgen (vgl. Gaca 2001: 114). Dies hängt sprachtypologisch damit zusammen, dass das Flexionssystem der Substantive im Polnischen und Tschechischen weit mehr erhalten ist als im Deutschen, weshalb die syntaktische Funktion nicht durch die Wortfolge, sondern durch die morphologische Markierung eindeutig wird (vgl. Engel et al. 2018: 392; Engel 2001: 64; Zeman 1992).

Besonders viele Unterschiede innerhalb der polnischen und deutschen Wortstellungsregeln ergeben sich bei der Stellung des Prädikats. Im Deutschen weist das Prädikat satztypabhängig eine fixe Position auf, das Polnische und das Tschechische sind hierbei durch eine größere Variabilität gekennzeichnet. Charakteristisch für das Deutsche ist das Prinzip der Satzklammer, einer zweigliedrigen Verbalstruktur, wohingegen das Prädikat im Polnischen und Tschechischen nicht klammerbildend ist: Im Polnischen und Tschechischen folgen die verbalen Elemente nacheinander, im Allgemeinen weisen sie also Kontaktstellung auf (vgl. Engel et al. 2018: 395, 398; Beneš 1968; Štícha 2003; Zeman 1992).

Im Folgenden wird auf die Wortstellungsabweichungen im Deutschen und im Polnischen eingegangen. Die Abweichungen in der tschechischen Sprache werden in Bučková & Nekula in diesem Band analysiert.

3.6.1 Wortstellungsabweichungen im Deutschen

In den durchgeführten sprachbiographischen Interviews betreffen die Abweichungen von der deutschen Normalstellung sehr oft die Abfolge im Verbalkomplex. Häufig beruhen diese Stellungsabweichungen darauf, dass bei Aussagesätzen nicht die Zweitstellung, sondern Erststellung des finiten Verbs (Beispiele (109)–(111)) eingesetzt wird:

(109) **hat er** [er hat] keine Probleme gehabt (BT_KLO_GP_DE)

(110) ja das waren also die ersten zwei drei Tage und dann hieß es ja ab **musst du** [du musst] nach Friedland (NT_LUB_GD_DE)

(111) nein nein **war ich** in Kindergarten nicht [ich war nicht im Kindergarten] (BQ_RAC_GP_DE)

Derartige Umstellungen des finiten Verbs sind im Deutschen entweder nicht möglich oder sie führen zur Transformation eines Aussagesatzes in einen Fragesatz (vgl. Engel et al. 2018: 402).

Ebenso treten häufig Abweichungen innerhalb des Verbalkomplexes in zusammengesetzten Sätzen auf. Während in der deutschen Standardsprache Nebensätze mit Ausnahme der uneingeleiteten Nebensätze bzw. der *weil*-Nebensätze im gesprochenen Deutsch prinzipiell die Verbendstellung aufweisen (vgl. Pittner & Berman 2004: 82), unterscheiden sich Aussagesätze und Nebensätze im Polnischen in der Grundfolge nicht voneinander

(vgl. Morciniec 2014: 70). Beispiele für Abweichungen von der Verbendstellung in deutschen Nebensätzen veranschaulichen Belege (112)–(114):

- (112) also ich denke schon **dass die meisten sind** in Deutschland [dass die meisten in Deutschland sind] (NT_LUB_GD_DE)
- (113) und **da ich bin** kein guter Verkäufer [da ich kein guter Verkäufer bin] (XL_MUN_GD_DE)
- (114) manchmal die Kinder wussten nicht / **wie sollen sie** sagen nicht [manchmal wussten die Kinder nicht, wie sie es sagen sollen nicht] (BQ_RAC_GP_DE)

Bei Fällen dieser Art, die im Projekt mit WO annotiert wurden, liegt ihre Deutung als Musterreplikationen (PAT) nahe. Speziell im schlesischen Teilkorpus ist allerdings zu bedenken, dass die Verbzweitstellung in Nebensätzen auch den schlesischen Dialekten des Deutschen eigen war, wie bereits Hanke (1913: 61f.) feststellt: „Im Schlesischen hat der Nebensatz auch gar nicht selten, man könnte beinahe sagen überwiegend, Hauptsatzstellung“.

Andere Abweichungen von der Wortfolge im Deutschen, etwa was die Stellung der adverbialen oder adjektivischen Phrase angeht, treten in dem Korpus seltener auf.

3.6.2 Wortstellungsabweichungen im Polnischen

Im Hinblick darauf, dass das Polnische über mehr Stellungsmöglichkeiten verfügt als das Deutsche, treten Wortstellungsabweichungen in den polnischen Aufnahmen weit seltener auf. Diese betreffen ebenfalls unterschiedliche Phrasentypen:

– die Verbalphrase (Beispiele (115) und (116)):

- (115) i ja myślę_jak to po śląsku **nazywa się**? [jak to się nazywa po śląsku] (XL_MUN_GD_PL)
,und ich überlege, wie es auf Schlesisch heißt‘
- (116) no gdyż bardzo dobrze **usytuowani byli** [gdyż byli bardzo dobrze usytuowani] nie w Łodzi (LU_KLO_GP_PL)
,na weil sie sehr betucht waren nicht in Łódź‘

– die Adverbialphrase (Beispiel (117)):

- (117) tam tyż się to spaliło **wtedy** [wtedy spaliło] (LU_KLO_GP_PL)
,dort ist es damals abgebrannt‘

– die Adjektivphrase (Beispiel (118)):

- (118) tym pożar był **słynny który** [ten słynny pożar był, który] (LU_KLO_GP_PL)
,es gab diesen berühmten Brand, der‘

– die Nominalphrase (Beispiele (119) und (120)):

(119) a z tyłu jest potem **Mickiewicza** ulica [ulica Mickiewicza] (BQ_RAC_GP_PL)
,und hinten ist dann die Mickiewicz-Straße‘

(120) i jechaliśmy do Noroka do do **mojej mamy** rodziców [do rodziców mojej mamy]
(BQ_RAC_GP_PL)
,wir sind nach Norok gefahren zu zu den Eltern meiner Mutter‘

Mehrfach lassen sich dabei ebenfalls potenzielle Nachbildungen der entsprechenden Stellungsregularitäten im Deutschen aufzeigen, etwa die Verbendstellung in Nebensätzen (vgl. (115) und (116)) oder die Voranstellung von Genitivattributen (vgl. (119) und (120)). Die eingangs erwähnte unterschiedliche Festigkeit der Wortstellung erweist sich somit nicht als ausschließlicher Einflussfaktor der Wortstellungsabweichungen, vielmehr dürfte hier auch die Korrelation mit soziolinguistischen Faktoren, wie etwa Bildungsgrad, von Belang sein.

3.7 Selbstkorrektur

Als Selbstkorrektur (*SC/self-correction*) werden Ausdrücke oder Formulierungen verstanden, die die Sprecherinnen und Sprecher selbst als falsch oder unpassend werteten und sie deswegen durch einen Ausdruck ersetzen. Die von ihnen vorgenommene Korrektur soll „beim Hörer mental an die Stelle des Bezugsdrucks treten“ (vgl. Fiehler 2011: 104) und damit auch eine eventuelle Fremdkorrektur verhindern. Die Selbstkorrekturen sind nicht nur eine Manifestation der Normorientierungen des Sprechenden (vgl. Schneider et al. 2018: 17) wie auch „das Ergebnis von seiner bewussten Sprachreflexion“ und dem expliziten Grammatikwissen (vgl. Kocianová 2005: 78, 97), sondern sie können auch seine sprachliche Unsicherheit offenlegen.

Als Indikatoren für Selbstkorrekturen werden Abbrüche, Dehnungen, Pausen, gefüllte Pausen (#) sowie Ausdrücke wie *nein*, *oder*, *also*, *beziehungsweise* im Deutschen (vgl. Schneider et al. 2018: 217) und *no*, *jakby*, *właśnie* im Polnischen (vgl. Majewska-Tworek 2014: 291, 306, 313, 357) angeführt. Die aufgezählten Erscheinungen kommen in den durchgeführten Interviews öfters vor. In unserem Projekt wurden aber nur die Selbstkorrekturen annotiert, deren Fehlen erkennbar zu einer Musterreplikation (PAT) oder zu einer anderen Abweichung (AA) führt. Wie bei den anderen Kategorien wurden die betreffenden Konstruktionen sowohl mit dem Label SC versehen als auch einem entsprechenden Phrasentyp zugeordnet. Die Beispiele dafür folgen in der Übersicht weiter unten.

3.7.1 Verbalphrase

Die Selbstkorrekturen im verbalen Bereich sind vielfältig und in den deutschsprachigen Interviewteilen stark präsent. Unter anderem betreffen sie den Gebrauch des Modalverbs *können* vs. *kennen*. In der „Generation Polen“ wurde die Entrundung des Stammvokals *ö* nicht eindeutig als eine Musterreplikation aus dem Polnischen (*znac język*) interpretiert,

sondern als ein dialektales Merkmal gesehen (vgl. Jorroch, Księżyk & Zielińska in diesem Band). In der „Generation Deutschland“ ist die Anlehnung an die polnische Kollokation in einigen Beispielen unumstritten, da von den Sprecherinnen und Sprechern diese finiten Verbformen korrigiert wurden, in denen der Lautwechsel (ö:e) gar nicht vorkommt (Beispiele (121) und (122)):

(121) und _ und dann habe ich also si * sicherlich **ich kannte ich konnte Deutsch** (NT_LUB_GD_DE)

(122) _ oder gleichzeitig # gleichwertig Deutsch und Wasserpolnisch ne # _ und ich bin aber mittlerweile auch froh **dass ich Polnisch kenne ne kann** (XL_MUN_GD_DE)

Die Verben der Möglichkeit involvieren auch in den tschechischen Daten Selbstkorrekturen. Dabei geht es um die Polysemie des deutschen *können*, die im Tschechischen durch zwei Verben ausgedrückt wird: *moci* ‚die Möglichkeit haben, etwas zu tun‘ und *umět* ‚die Fähigkeit haben, etwas zu beherrschen‘ (siehe auch Bučková & Nekula in diesem Band). Der Sprecher in (123) ist sich dieses Unterschieds bewusst, wodurch es zu einer Reparatur in seiner Äußerung kommt:

(123) my jsme **mohli mluvit uměli mluvit** (II_ROD_SP_CZ)
,wir konnten sprechen konnten sprechen‘

Auch die Wahl des richtigen Hilfsverbs im Deutschen bereitet in der „Generation Deutschland“ einige Probleme, die den Interviewten bewusst sind (Beispiele (124) und (125)):

(124) und so **ist es # so hat das angefangen** (UC_LUC_GD_DE)

(125) so **ich habe** neunzehnhundertneunzig **bin ich** aus Polen **weggefahren** (UC_LUC_GD_DE)

Sprachliche Unsicherheit rufen auch die Partizip-II-Formen hervor, die entweder abweichend gebildet (126) oder übermäßig verwendet (127) wurden, was Korrekturen nach sich zieht:

(126) sie haben **gebitten ge* gefragt** wer möchte (UC_LUC_GD_DE)

(127) wenn du wenn du **gebeten willst beten willst** (BK_AUG_GD_PL)

Im Polnischen treten die selbstkorrigierten Abweichungen vor allem bei der Konjugation auf, wie in (128), wo die Sprecherin die Flexionsendung der IV. anstelle der VI. Konjugationsklasse lt. Tokarski (Nagórko 2012: 223) bei der Konjugation des Verbs *podszkolić* verwendet:

(128) i mnie troszkę **podszkoluje podszkoli** mi ten # ten mój polski (NA_HAM_GD_PL)
,und [er] bildet mich ein bisschen in dem # in meinem Polnischen fort‘

3.7.2 Nominalphrase

Im nominalen Bereich handeln die Sprecherinnen und Sprecher u. a. bei der Genuszuweisung an Substantive unsicher. In (129) weist der Sprecher im Deutschen dem femininen Substantiv *Spur* das maskuline Genus zu. Die Zuweisung des maskulinen Genus ist auf das Polnische zurückzuführen, in dem das polnische wörtliche Äquivalent von *Spur*, nämlich *ślad*‘ eben maskulin ist:

(129) dass sie **den ein ein eine Spur** noch machen (BT_KLO_GP_DE)

Auch beim Numerus begegnet man Selbstkorrekturen. Die Sprecherin in (130) will zunächst das Substantiv *noviny* ‚Zeitung‘ entsprechend dem Numerus des deutschen Äquivalents im Singular verwenden, was aus dem Adjektiv *českou* ‚tschechisch.F.SG.ACC‘ ersichtlich wird. Dann bricht sie ihre Äußerung ab und zögert, was sich an dem Häsitationsgeräusch zeigt. Schließlich ändert sie die Form zu *české* ‚tschechisch.F.PL.ACC‘, was im Einklang mit der Baseline steht.

(130) a _ přečtu **českou nov*** # **české noviny** (CI_BUB_SP_CZ)
 ‚und _ ich lese tschechische Zei* # tschechische Zeitungen‘

Manchmal mischen die Interviewten phonetisch ähnliche Nomen, die etwas anderes bedeuten – wie in (131) den *Ausflug* mit dem *Abflug*. Es kommt auch zu Neubildungen. In (132) bildet der Sprecher ein neues Substantiv *Zuziedler* mithilfe eines ungewöhnlichen Präfixes, anstatt das gewöhnliche Nomen *Aussiedler* zu verwenden:

(131) wo die meisten dann mal **den _Ausflug oder den Abflug** gemacht haben ja (HM_STE_GD_DE)

(132) die die welche sind dann # **die Zuz* Zuziedler also die die Aussiedler** (SO_LAM_GD_DE)

3.7.3 Präpositionalphrase und Satzebene

Die Selbstkorrekturen im präpositionalen Bereich betreffen u. a. den semantischen Transfer. Im Beispiel (133) ist die Sprecherin unsicher, welche Präposition angemessen ist. Die deutsche Präposition *von* lässt sich kontextabhängig mit verschiedenen Präpositionen wie *o*, *przez*, *z*, *od* auf Polnisch wiedergeben. Die Sprecherin verwendet zunächst die von der Baseline abweichende Präposition *z* und korrigiert sie dann auf *od*:

(133) ponieważ był inwalidą # **z od urodzenia** (IT_ELB_GP_PL)
 ‚denn er war Invalide # von Geburt an‘

In (134) wählt die Sprecherin zuerst die Präposition *na* ‚auf‘, die im gegebenen Kontext (*kostel byl na cestě* ‚die Kirche war auf dem Weg‘) andeuten würde, dass die Kirche auf einem Weg oder auf einer Straße gebaut wurde. Dann korrigiert sie sich und kodiert durch die Präposition *po* ‚entlang, unterwegs‘ die Bedeutung, dass sie die Kirche unterwegs gesehen hat.

- (134) a my jsme šli do př* _ nějakýho kostela co byl právě **na _ na cestě _ po cestě**
(CI_BUB_SP_CZ)

,und wir sind in _ eine Kirche gegangen die gerade auf _ auf dem Weg _ auf dem Weg war‘

3.7.4 Adjektivphrase und Adverbialphrase

Die Selbstkorrekturen im Bereich des Adjektivs sind entweder eine Folge der falschen Genuszuweisung in (135) und (136) oder einer richtigen Genuszuweisung, jedoch falscher Flexionsendung (137). In (135) kann sich die Sprecherin nicht entscheiden, ob das polnische maskuline Substantiv *procent* neutral, feminin oder doch maskulin ist. Je nach der Entscheidung dekliniert sie das Indefinitpronomen *jakiś* nach dem neutralen (*jakieś*), femininen (*jakąś*) und schließlich maskulinen (*jakiś*) Genus:

- (135) że będą ściągać # **jakieś jakąś jakiś procent** (IT_ELB_GP_PL)

,dass sie # irgendeines, irgendeine, irgendein Prozent einziehen werden‘

In (136) weist die Sprecherin dem neutralen Substantiv *Blut* zunächst das feminine Genus zu, was auf das polnische Äquivalent *krw* (feminin) zurückzuführen ist. Infolgedessen dekliniert sie auch das Adjektiv *rein* zunächst nach dem femininen Genus und nach der Selbstkorrektur nach dem neutralen Genus:

- (136) also dies das ist ja **eine polni poln * polnische Blut bis ja polnisches Blut reines rein * polnisches Blut** bis Jahrhunderte zurück (NA_HAM_GD_DE)

In (137) verwendet die Sprecherin den von der Baseline im Polnischen abweichenden Kasus, nämlich den Nominativ (*angielski*) anstelle von Genitiv (*angielskiego*), was auf das Valenzmuster im Deutschen zurückzuführen ist. Nach der Aussage kommt das Pausenzeichen, danach wird die Selbstkorrektur vorgenommen:

- (137) i **angielski** uczyła się też prywatnie _ **angielskiego** (CH_BER_GD_PL)

,und Englisch hat sie privat gelernt _ Englisch‘

Am seltensten kommen die Selbstkorrekturen im adverbialen Bereich vor. In (138) verwendet die Sprecherin zunächst ein Adverb *często* ‚oft‘ im Kontext des Sprechens mit dem sog. fremden Akzent. Danach korrigiert sie sich und sagt nach der Hässitation *długo* ‚lang‘, was im Einklang mit dem Verb *mieć* ‚haben‘ ist:

- (138) **bardzo często _ długo** miałam akcent _ ten # niemiecki (KL_GDA_GP_PL)

,sehr oft _ lange hatte ich Akzent _ den # deutschen‘

4 Fazit und Ausblick

Die im vorliegenden Kapitel vorgestellte sprachstrukturelle Annotationsweise soll eine Möglichkeit aufzeigen, wie mit gesprochenen Daten bilingualer Sprecherinnen und Sprecher verfahren werden kann. Dabei sollte jenseits einer klassischen Typologie, welche auf einer Dreiteilung in die phonetisch-phonologische, lexikalisch-semantische und grammatische Ebene basiert und so bereits bei Weinreich (1977) zu finden ist, ein Ansatz aufgezeigt werden, der zwischen materiellem Transfer und der Replikation von Mustern aus der einen Sprache in die andere unterscheidet. Neben diesen beiden Kategorien war es jedoch nötig, weitere Kategorien, wie AA, CS, WO oder SC einzuführen, um bei der Annotation möglichst keine morphosyntaktischen Abweichungen von der Baseline auszulassen, auch wenn die dadurch erfassten Phänomene über die im LangGener-Projekt gesetzten Untersuchungsziele hinausgehen. Dies war dem Zweck geschuldet, dass das im Rahmen des Projekts erstellte Korpus zur deutsch-polnischen bzw. deutsch-tschechischen Zweisprachigkeit nach Projektabschluss auch anderen Forscherinnen und Forschern mit womöglich anderen Forschungsinteressen zugänglich gemacht werden soll. So gesehen stellt beispielsweise die sehr umfangreiche Kategorie der anderen Abweichungen (AA), die bei der Typologisierung zusätzlich aufgestellt wurde und Phänomene versammelt, in denen weder eine MAT-, noch eine PAT-Replikation vorlag, ein breites Feld für weitere Untersuchungen dar, die ebenso wertvolle Informationen über die Abweichungen in der Sprache Bilingualer liefern könnten. Interessante Untersuchungsmöglichkeiten eröffnet auch die Kategorie der Wortstellung, dank der die annotierten Abweichungen ebenfalls etwa im Zusammenhang mit der Art der Zweisprachigkeit, der Sprachkontaktinduziertheit oder im Rahmen ausgewählter Stellungsregularitäten analysiert werden können.

Durch die entstandene Typologie mussten in Bezug auf das Korpus pragmatisch motivierte und zu einem gewissen Grad arbiträre, allerdings möglichst einheitliche Festlegungen bezüglich der Einordnung einiger Beispiele in diese Typologie getroffen werden. Es galt also, einen Grad zwischen einer genauen Annotation der Sprachkontakthänomene und einer Vereinheitlichung, die quantitativen Analysen dient, zu finden. Neben den prototypischen Fällen gab es daher Entscheidungen, die anhand einer Reihe zusätzlicher Kriterien getroffen wurden. So musste beispielsweise festgelegt werden, wie mit dem Verlust von Kategorien umgegangen wird. Somit zeigt dieses Kapitel Grenzen der Typologie und weitere Felder, die theoretischer Überlegungen bedürfen, auf. Ferner dürften die hier zahlreich angeführten Belege veranschaulicht haben, dass anders als etwa bei einer morphologischen Annotation, die hier beschriebene sprachformbezogene Annotation nicht automatisch umsetzbar ist, sondern manuell – verbunden mit einer genauen linguistischen Prüfung – durchgeführt werden muss.

Das vorliegende Kapitel ist mit einem der Hauptziele des LangGener-Projekts verbunden – der Entwicklung einer Typologie morphosyntaktischer Musterreplikationen. Ein weiterer, sich davon ableitender Projektzweck war die Untersuchung der Nachbildungen morphosyntaktischer Muster in Abhängigkeit von den interviewten Generationen. Diesen Zweck verfolgen insbesondere zwei Dissertationen, die parallel zu diesem Band entstehen: Aneta Bučková untersucht in ihrer Arbeit syntaktische Musterentlehnungen im Bereich der Argumentstruktur. Carolin Centner erforscht hingegen lexikalische Replikationen auf Verbalphrasenebene und ordnet die Replikationen in eine Klassifikation ein, die sowohl qualitativ beschrieben als auch quantitativ ausgewertet wird. Erste Ergebnisse der

PAT-Untersuchung sind bereits in diesem Band enthalten: Bučková & Nekula beschreiben nicht nur das deutsch-tschechische Teilprojekt, sondern geben auch einen Überblick über die sprachkontaktbedingte Attrition des Tschechischen.

Eine Besonderheit des Projekts und auch des daraus entstehenden Korpus ist die Verbindung der soziolinguistischen mit der sprachformbezogenen Annotation. Dadurch soll ein wertvoller Beitrag zur interdisziplinären Betrachtungsweise von Sprachkontakt geleistet werden. Die interdisziplinäre Reichweite beschränkt sich dabei nicht auf die Kontakt- und Soziolinguistik; auch Wissenschaftlerinnen und Wissenschaftler aus anderen linguistischen Disziplinen sowie aus den Geistes- und Sozialwissenschaften sollen von dem Korpus profitieren.

Literatur

- Archiwum Państwowe w Katowicach: <https://www.katowice.ap.gov.pl/art,440,raciborz-poczatek-nowego-swiata-1945-1946>, letzter Zugriff am 11.06.2021.
- Auer, Peter (1995): The pragmatics of code-switching. A sequential approach. In: Lesley Milroy, Pieter Muysken (Hgg.), *One Speaker, two Languages. Cross-Disciplinary Perspectives on Code-Switching*. Cambridge: Cambridge University Press, 115–135.
- Bachmann, Armin (2013): Das gesprochene Prager Deutsch in seiner letzten Phase. In: Anja Ference, Libuše Spáčilová (Hgg.), *Deutsch als Sprache der (Geistes) Wissenschaften. Linguistik*. Brunn: Tribun EU, 11–19.
- Backus, Ad (2020): Code-switching as a reflection of contact-induced change. In: Anthony P. Grant, Ad Backus (Hgg.), *The Oxford Handbook of Language Contact*. Oxford: Oxford University Press, 192–214.
- Balhar, Jan (1995): Jak mluví dnes vídeňští Češi. *Naše řeč* 78.4, 189–196.
- Bartnicka, Barbara/Hansen, Björn/Klemm, Wojtek/Lehmann, Volker/Satkiewicz, Halina (2004): *Grammatik des Polnischen*. München: Otto Sagner.
- Basler, Helena et al. (Hg.) (2006): *Die Wiener Tschechen 1945-2005. Zur Geschichte einer Volksgruppe*. Wien: Tschechisches kulturhistorisches Institut.
- Beneš, Eduard (1968): Die funktionale Satzperspektive im Deutschen im Vergleich mit dem Tschechischen. In: Bohuslav Havránek (Hg.), *Deutsch-tschechische Beziehungen im Bereich der Sprache und Kultur*. Berlin: Akad.-Verlag, 57–69.
- Berger, Tilman (2008): *Studien zur historischen Grammatik des Tschechischen*. München: Lincom.
- Berger, Tilman (2014): The convergence between Czech and German between the years 900 and 1500. In: Juliane Besters-Dilger, Cynthia Dermarkar, Stefan Pfänder, Achim Rabus (Hgg.), *Family Effects in Language Contact. Modeling Congruence as a Factor in Contact Induced Change*. Berlin, Boston: De Gruyter, 184–198.
- Betsch, Michael (2000): *Diskontinuität und Tradition im System der tschechischen Anredepronomen (1700-1850)*. Bern: Peter Lang.
- Blahak, Boris (2015): *Franz Kafkas Literatursprache. Deutsch im Kontext des Prager Multilingualismus*. Köln, Weimar, Wien: Böhlau.
- Blankenhorn, Renate (2003): *Pragmatische Spezifika der Kommunikation von Russlanddeutschen in Sibirien. Entlehnung von Diskursmarkern und Modifikatoren sowie Code-switching*. Frankfurt/Main: Peter Lang.
- Błaszczyc, Izabela (2018): *Wenn man die polnische Sprache erbt...Beschreibung und empirische Analyse zum Irrealis-Marker by im Polnischen und seiner Verwendung bei Herkunftssprechern mit Deutsch als dominanter Sprache*. Regensburg: Dissertation an der Universität Regensburg. Online verfügbar unter: <https://epub.uni-regensburg.de/37566/>, letzter Zugriff am 11.06.2021.

- Bolonyai, Agnes (2009): Code-switching, imperfect acquisition, and attrition. In: Almeida Jacqueline Toribio, Barbara E. Bullock (Hgg.), *Cambridge Handbook of Linguistic Code-switching*. Cambridge: Cambridge University Press, 253–269.
- Brehmer, Bernard/Czachór, Agnieszka (2012): The formation and distribution of the analytic future tense in Polish-German bilinguals. In: Kurt Braunmüller, Christoph Gabriel (Hgg.), *Multilingual Individuals and Multilingual Societies*. Amsterdam, Philadelphia: John Benjamins, 297–314.
- Brehmer, Bernard/Rothweiler, Monika (2012): The acquisition of gender agreement marking in Polish. In: Kurt Braunmüller, Christoph Gabriel (Hgg.), *Multilingual Individuals and Multilingual Societies*. Amsterdam, Philadelphia: John Benjamins, 81–100.
- Bučková, Aneta (2016): Der Idiolekt von Lenka Reinerová. *brücken – Germanistisches Jahrbuch Tschechien – Slowakei* 24.1-2, 121–162.
- Chmiel, Peter (1988): Zum Problem der sprachlichen Integration deutscher Aussiedler aus Oberschlesien. *Oberschlesisches Jahrbuch* 4, 117–128.
- Clyne, Michael G. (1967): *Transference and Triggering*. The Hague: Nijhoff.
- Clyne, Michael G. (2003): *Dynamics of Language Contact. English and Immigrant Languages. Cambridge Approaches to Language Contact*. Cambridge: Cambridge University Press.
- Digitales Wörterbuch der deutschen Sprache (DWDS): <https://www.dwds.de/wb/wdg/mögen>, letzter Zugriff am 17.09.2020.
- Digitales Wörterbuch der deutschen Sprache (DWDS): <https://www.dwds.de/wb/Radio>, letzter Zugriff am 11.06.2021.
- Engel, Ulrich (1994): *Syntax der deutschen Gegenwartssprache*. Berlin: Erich Schmidt.
- Engel, Ulrich (1996): *Deutsche Grammatik*. Heidelberg: Groos.
- Engel, Ulrich (2001): Zur Beschreibung der Wortstellung auf Grund der Abhängigkeitsstruktur. Ein Vorschlag für sprachvergleichende Untersuchungen. In: Peter Bassola, Ulrich Engel, Alicja Gaca, Marc van de Velde (Hgg.), *Wortstellung im Sprachvergleich (deutsch-niederländisch-polnisch-ungarisch)*. Tübingen: Julius Groos Verlag, 63–89.
- Engel, Ulrich/Rytel-Kuc, Danuta/Cirko, Lesław/Dębski, Antoni (1999): *Deutsch-polnische kontrastive Grammatik*. Band 1 & 2. Heidelberg: Julius Groos Verlag.
- Engel, Ulrich/Cirko, Lesław/Jurasz, Alina/Rytel-Schwarz, Danuta (2018): *Deutsch-polnische kontrastive Grammatik*. Bd. 2. Hildesheim: Georg Olms Verlag.
- Feldmeier, Beata (2014): *Anrede im Sprachkontakt. Reflexion und Gebrauch von Anredestrategien durch tschechische Migranten im deutschsprachigen Umfeld*. München, Berlin, Washington: Verlag Otto Sagner.
- Fiehler, Reinhard (2011): Mündliche Verständigung und gesprochene Sprache. In: Sandro M. Moraldo (Hg.), *Deutsch aktuell 2. Einführung in die Tendenzen der deutschen Gegenwartssprache*. Rom: Carocci, 83–107.
- Fischer, Gero (1993): The language of the Czech minority in Vienna. In: Eva Eckert (Hg.), *Varieties of Czech. Studies in Czech Sociolinguistics*. Amsterdam: Rodopi, 254–262.
- Gaca, Alicja (2001): Thema und Satzanfang. Die pronominalen Kasusergänzungen im Deutschen und im Polnischen. In: Peter Bassola, Ulrich Engel, Alicja Gaca, Marc van de Velde (Hgg.), *Wortstellung im Sprachvergleich (deutsch-niederländisch-polnisch-ungarisch)*. Tübingen: Julius Groos Verlag, 91–119.
- Gardner-Chloros, Penelope (2009): *Code-Switching*. Cambridge: Cambridge University Press.
- Grosjean, François (2006): The bilingual's language modes. In: Janet L. Nicol (Hg.), *One Mind, two Languages. Bilingual Language Processing. Explaining Linguistics*. Bd. 2. Malden, Mass.: Blackwell, 1–22.
- Hanke, Lothar (1913): *Die Wortstellung im Schlesischen*. In: Wort und Brauch 11. Breslau: M. & H. Marcus.
- Helbig, Gerhard/Buscha, Joachim (1996): *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. 17. Aufl. Leipzig, Berlin u. a.: Langenscheidt.

- Helbig, Gerhard/Buscha, Joachim (2005): *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. 5. neubearb. Aufl. Berlin, München u. a.: Langenscheidt.
- Hinc, Jolanta (2010): Englisch als Interferenzquelle bei der Aneignung der Wortstellung des Deutschen. *Glottodidactica* 36, 143–54.
- Hlavac, Jim (2006): Bilingual discourse markers: Evidence from Croatian–English code-switching. *Journal of Pragmatics* 38.11, 1870–1900.
- Joachimiak, Joanna (1999): Regionalne cechy północnokresowe we współczesnej polszczyźnie mieszkańców Białorusi. *Zeszyty Naukowe Wyższej Szkoły Pedagogicznej w Bydgoszczy. Studia Filologiczne. Filologia Rosyjska* 45 (19), 21–36.
- Jorroch, Anna (2015): *Die deutsche Sprache der dreisprachigen Altgläubigen in Masuren*. Warszawa: Instytut Germanistyki Uniwersytetu Warszawskiego.
- Karlík, Petr/Nekula, Marek/Rusínová, Zdena (Hgg.) (1995): *Průruční mluvnice češtiny*. Praha: Nakladatelství Lidové noviny.
- Karszniewicz-Mazur, Alicja (1988): *Zapożyczenia leksykalne ze źródła niemieckiego we współczesnej polszczyźnie*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Kocianová, Maria (2005): *Metasprachliche Fähigkeiten zweisprachiger Kinder: Zum Zusammenhang von sprachlicher und metasprachlicher Leistungsfähigkeit und die damit einhergehenden Implikationen für eine adäquate Förderung der Russisch-Deutsch sprechenden Kinder im Grundschulalter*. Inaugural-Dissertation zur Erlangung des Doktorgrades der Philosophie an der Ludwig-Maximilians-Universität München.
- Księżyk, Felicja (2008): *Die deutsche Sprachinsel Kostenthal. Geschichte und Gegenwart*. Berlin: trafo Verlag.
- Majewska-Tworek, Anna (2014): *Niepłynność wypowiedzi w oficjalnej odmianie polszczyzny. Propozycja typologii*. Wrocław: Quaestio.
- Matras, Yaron (1998): Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36.2, 281–332.
- Matras, Yaron (2009): *Language Contact*. Cambridge: Cambridge University Press.
- Matras, Yaron/Sakel, Jeanette (2007): Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31.4, 829–865.
- Michalewska, Maria T. (1991): *Polszczyzna osób bilingwalnych w Zagłębiu Ruhry w sytuacji oficjalnej*. Kraków: Uniwersytet Jagielloński.
- Morcinek, Norbert (2014): *Gramatyka kontrastywna: Wprowadzenie do niemiecko-polskiej gramatyki kontrastywnej*. Wrocław: Wydawnictwo Wyższej Szkoły Filologicznej.
- Muysken, Pieter (2000): *Bilingual Speech. A Typology of Code-Mixing*. Cambridge, New York: Cambridge University Press.
- Myers-Scotton, Carol (1997): *Duelling Languages. Grammatical Structure in Code-Switching*. Oxford: Clarendon Press.
- Nagórko, Alicja (2012): *Podręczna gramatyka języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Nekula, Marek (1996a): *System der Partikeln im Deutschen und Tschechischen unter besonderer Berücksichtigung der Abtönungspartikeln*. Tübingen: Niemeyer.
- Nekula, Marek (1996b): Textuelle Funktionen des deutschen Artikels und des tschechischen Aspekts. In: Wolfgang Gladrow, Sonja Heyl (Hgg.), *Slawische und deutsche Sprachwelt*. Frankfurt/Main: Peter Lang, 161–172.
- Nekula, Marek (2003): *Franz Kafkas Sprachen: „...in einem Stockwerk des innern babylonischen Turmes...“*. Tübingen: Niemeyer.
- Nekula, Marek (2021): Varieties of German in the Czech Republic. In: *Wiener Slavistisches Jahrbuch* N.F. 9, 69–103.
- Nekula, Marek/Rychnovská, Lucie (2013): Jazyková biografie Bedřicha Smetany a jazyk jeho korespondence se zřetelem ke gramatickým jevům. In: Zdeňka Hladká, Robert Adam, Dana

- Hlaváčková (Hgg.), *Soukromá korespondence jako lingvistický pramen*. Brunn: Masarykova univerzita, 99–136.
- Nekula, Marek/Rychnovská, Lucie (2016): Jazyk Smetanových dopisů v dobovém kontextu / The Language of Smetana's Letters and Diaries in the Context of His Time. / Die Sprache von Smetanas Briefen im Kontext seiner Zeit. In: Olga Mojžíšová, Milan Pospíšil (Hgg.), *Bedřich Smetana, Korespondence I (1840–1862). Kritická edice*. Praha: Národní muzeum – Koniasch Latin Press, 28–73, 123–181, 235–291, 429–444.
- Nekula, Marek/Šichová, Kateřina/Valdrová, Jana (Hgg.) (2013): *Bilingualer Sprachvergleich und Typologie: Deutsch – Tschechisch*. Tübingen: Stauffenburg.
- Nekula, Marek/Večerka, Radoslav (2017): Bohemismus. In: Petr Karlík, Marek Nekula, Jana Pleskalová (Hgg.), *CzechEncy – Nový encyklopedický slovník češtiny*. Online verfügbar unter: <https://www.czechency.org/slovník/BOHEMISMUS>, letzter Zugriff am 17.02.2021.
- Newerkla, Stefan Michael (2013): *Sprachkontakte Deutsch – Tschechisch – Slowakisch. Wörterbuch der deutschen Lehnwörter im Tschechischen und Slowakischen: historische Entwicklung, Beleglage, bisherige und neue Deutungen*. 2. Aufl. Frankfurt/Main: Peter Lang.
- Newerkla, Stefan Michael (2017): Čeština v Rakousku. In: Petr Karlík, Marek Nekula, Jana Pleskalová (Hgg.), *CzechEncy – Nový encyklopedický slovník češtiny*. Online verfügbar unter: https://www.czechency.org/slovník/ČEŠTINA_V_RAKOUSKU, letzter Zugriff am 17.02.2021.
- Nowowiejski, Bogusław (1996): *Zapożyczenia leksykalne z języka niemieckiego w polszczyźnie XIX wieku*. Białystok: Dział Wydawnictw Filii UW.
- Pelka, Daniela (2006): *Der deutsch-polnische Sprachkontakt in Oberschlesien am Beispiel der Gegend von Oberglogau*. Berlin: trafo Verlag.
- Pittner, Karin/Berman, Judith (2004): *Deutsche Syntax: Ein Arbeitsbuch*. Tübingen: Narr.
- Poplack, Shana/Sankoff, David/Miller, Christopher (1988): The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26.1, 47–104.
- Půda, Aleš (2010): *Zur Theorie der Lehnprägung im deutsch-tschechischen Sprachkontakt. Eine historisch-vergleichende Untersuchung im innerslavischen und europäischen Kontext*, Frankfurt am Main: Peter Lang.
- Riehl, Claudia M. (2014): *Sprachkontaktforschung. Eine Einführung*. 3., überarbeitete Aufl. Tübingen: Narr.
- Sakel, Jeanette (2008): Types of loan. Matter and pattern. In: Yaron Matras, Jeanette Sakel (Hgg.), *Grammatical Borrowing in Cross-Linguistic Perspective*. Berlin, New York: Mouton de Gruyter, 15–31.
- Skála, Emil (1966): Das Prager Deutsch. In Tradition und Ursprünglichkeit. In: Werner Kohlschmidt, Herman Meyer (Hgg.), *Akten des III. Internationalen Germanistenkongresses 1965 in Amsterdam*. Bern, München: Francke Verlag, 156–158.
- Schneider, Jan Georg/Butterworth, Judith/Hahn, Nadine (2018): *Gesprochener Standard in syntaktischer Perspektive. Theoretische Grundlagen – Empirie – didaktische Konsequenzen*. Tübingen: Stauffenburg Verlag.
- Synowicz, Helena (1992): *Sprawność językowa uczniów w śląskim środowisku gwarowym. Problemy, badania, konsekwencje dydaktyczne*. Katowice: Uniwersytet Śląski.
- Štícha, František (2003): *Česko-německá srovnávací gramatika*. 1. Aufl. Praha: Argo.
- Štícha, František/Vondráček, Miroslav/Kolářová, Ivana/Hoffmannová, Jana/Bílková, Jana/Svobodová, Ivana (2013): *Akademická gramatika spisovné češtiny*. Praha: Academia.
- Trost, Pavel (1962): Das späte Prager Deutsch. *Germanistica Pragensia* 2, 31–39.
- Trost, Pavel (1964): Franz Kafka und das Prager Deutsch. *Germanistica Pragensia* 3, 29–37.
- Trost, Klaus/Jelínek, Milan (Hgg.) (2000): *Deutsch-tschechische Sprachbeziehungen. Germanismen, Personennamen, Ortsnamen*. Regensburg: Roderer.
- Weinreich, Uriel (1977): *Sprachen in Kontakt. Ergebnisse und Probleme der Zweisprachigkeitsforschung*. München: Beck.

- Weiss, Daniel (1982): Deutsch-polnische Lehnbeziehungen im Bereich der Passivbildung, in: Eberhard Reißner (Hgg.), *Literatur- und Sprachentwicklung in Osteuropa im 20. Jahrhundert*. Berlin: Berlin Verlag, 197–218.
- Weiss, Daniel (1987): Polsko-niemieckie paralele w zakresie czasowników modalnych (na tle innych języków zachodniosłowiańskich). In: Gerd Hentschel (Hgg.), *Sprach- und Kulturkontakte im Polnischen*. München: Sagner, 131–156.
- Wójtowiczowa, Janina (1995): *Wybór tekstów polonijnych z Zagłębia Ruhry*. Warszawa: Energeia.
- Zeman, Dalibor (2011): Tschechische Einflüsse im Wienerischen auf phonologischer, phraseologischer sowie lexikalischer Ebene. Einige Bemerkungen zu den österreichisch-tschechischen Sprachbeziehungen. *brücken – Germanistisches Jahrbuch Tschechien – Slowakei* 19.1-2, 43–91.
- Zeman, Jaromír (1992): Wortstellungsschemata im Deutschen und im Tschechischen, Versuch einer vergleichenden Darstellung. *DAAD Dokumentation und Materialien* 25, Bonn, 17–34.

Praktische Lösungen

Praktische Durchführung der Feldforschung im deutsch-polnischen Team

Anna Jorroch – Irena Prawdzic

Abstract:

The practical side of the fieldwork of the German-Polish team

This chapter reports on the fieldwork carried out in the German-Polish project. It recapitulates the research questions of the overall project and describes the practical side of the fieldwork in Poland and Germany. It deals with the way of getting in contact with the bilingual informants belonging to two different generations. Further, the authors give an outline of how to conduct an interview in a maximally natural setting and discuss unforeseen problems the researchers had to tackle during the fieldwork. The final section contains a list of conversational topics used in the interviews with the bilingual speakers. This chapter can be seen as a hands-on report and might be helpful for future research projects on bilingualism which strive for environments generating natural data.

1 Das Forschungsvorhaben

Laut der Arbeitspläne sollten wir 30 Personen in Polen und 30 Personen in Deutschland interviewen. In Polen haben wir Untersuchungen in Schlesien, Pommern, Masuren und in dem Teil der heutigen Woiwodschaft Großpolen durchgeführt, der in den Jahren 1922–1938 zur Provinz Grenzmark Posen-Westpreußen und in den Jahren 1938–1945 zur Provinz Pommern¹ gehörte. Alle Forschungsregionen wurden 1945 an Polen angeschlossen und liegen im historischen deutsch-polnischen Grenzgebiet. Die Kontaktpersonen sind demzufolge unter natürlichen Umständen der Zweisprachigkeit aufgewachsen. Jede Region, in der wir geforscht haben, unterscheidet sich von der anderen wesentlich. Die in den genannten Gebieten durchgeführten Interviews erlaubten uns, nicht nur die Sprachkontaktphänomene im Sprachgebrauch der Bilingualen, sondern auch Dialektmerkmale des Polnischen in Schlesien und des Deutschen (siehe Jorroch, Księżyk & Zielińska in diesem Band) zu dokumentieren. In Deutschland haben wir die Gespräche in Bayern, Nordrhein-Westfalen, Berlin und in Hamburg aufgenommen. Wir haben insgesamt mit 124 anstatt mit 60 Personen gesprochen, weil sich zum einen mehr Kontaktpersonen gemeldet haben, als wir angenommen hatten. Zum anderen haben wir die Suche nach

¹ <https://ome-lexikon.uni-oldenburg.de/regionen/grenzmark-posed-westpreussen>, letzter Zugriff am 05.03.2021.

Respondentinnen und Respondenten immer fortgesetzt, da wir nicht voraussehen konnten, ob die Gespräche aufgrund der von uns festgelegten Profilkriterien ins Korpus einbezogen werden können. Ins Korpus haben wir insgesamt ca. 78 Gesprächsstunden mit 58 Respondentinnen und Respondenten aufgenommen. Die Liste der Kontaktpersonen mit ihrer regionalen Zugehörigkeit ist im Anhang 4 dieses Bandes angegeben.

2 Die Vorbereitung auf die Feldforschung: Wege der Kontaktaufnahme

Die vorgenommenen Analysen zu Veränderungen im Sprachgebrauch der Bilingualen basieren auf aufgenommenen Gesprächen. Der erste Schritt, den wir tun mussten, um die Projektziele zu realisieren, war dementsprechend die Kontaktaufnahme zu den Sprecherinnen und Sprechern. Die Feldforschung in Polen wurde für das erste Projektjahr 2018 und jene in Deutschland für das darauffolgende Jahr geplant.

Das Interviewer-Team zählte insgesamt 10 Personen, von denen zwei nicht Projektmitarbeiter, sondern geschulte Hilfskräfte waren. Einige von uns hatten ihre Privatkontakte unter Bekannten und Familie oder dank früherer wissenschaftlicher Tätigkeit für die Akquise von Interviewpartnerinnen und -partnern genutzt. Diese Ressourcen ermöglichten uns die Gespräche in Westpommern dank Prof. Dr. Anna Zielińska, in Masuren dank Dr. Anna Joroch und in Schlesien dank Dr. habil. Felicja Księżyk. Hauptsächlich haben wir aber die zweisprachigen Personen über verschiedene Institutionen gesucht, darunter über:

- den Verband der Deutschen Sozial-Kulturellen Gesellschaften in Polen (Związek Niemieckich Stowarzyszeń Społeczno-Kulturalnych w Polsce),
- die Sozial-Kulturelle Gesellschaft der Deutschen Minderheit Ortsgruppe Stargard (Towarzystwo Społeczno-Kulturalne Mniejszości Niemieckiej Koło Stargard),
- die Sozial-Kulturelle Gesellschaft der Deutschen Minderheit Ortsgruppe Arnswalde (Towarzystwo Społeczno-Kulturalne Mniejszości Niemieckiej Koło w Choszczynie),
- die Sozial-Kulturelle Gesellschaft der Deutschen Minderheit "Pomerania" in Köslin (Towarzystwo Społeczno-Kulturalne Mniejszości Niemieckiej „Pomerania“ w Koszalinie),
- die Deutsche Sozial-Kulturelle Gesellschaft in Schneidemühl (Niemieckie Towarzystwo Społeczno-Kulturalne w Pile),
- die Deutsche Sozial-Kulturelle Gesellschaft Ortsgruppe Flatow (Niemieckie Towarzystwo Społeczno-Kulturalne w Pile – Koło Terenowe w Złotowie),
- den Sozial-Kulturellen Bund der Deutschen Minderheit (Stowarzyszenie Społeczno-Kulturalne Mniejszości Niemieckiej w Świdwinie),
- den Verband der Deutschen Gesellschaften in Ermland und Masuren (Związek Stowarzyszeń Niemieckich Warmii i Mazur),
- die Danziger Deutsche Minderheit (Gdańska Mniejszość Niemiecka),
- die Gesellschaft der Deutschen Minderheit Stadt und Kreis Elbing (Stowarzyszenie Mniejszości Niemieckiej Miasta i Powiatu Elbląg),
- die Gesellschaft der Deutschen Minderheit "Masuren" in Lyck (Stowarzyszenie Mniejszości Niemieckiej Mazury w Elku),

- die Sozial-Kulturelle Gesellschaft der Deutschen Minderheit in Stettin (Towarzystwo Społeczno-Kulturalne Mniejszości Niemieckiej w Szczecinie),
- den Deutsch Sozial-Kulturellen Verein in Lötzen (Niemieckie Stowarzyszenie Kulturalno-Socjalne w Giżycku),
- das Pfarramt der Evangelisch-Augsburgischen Kirche zur Heiligen Dreifaltigkeit in Stettin (Parafia Ewangelicko-Augsburska Świętej Trójcy w Szczecinie),
- das Haus der Deutsch-Polnischen Zusammenarbeit (Dom Współpracy Polsko-Niemieckiej).

An die meisten dieser Institutionen haben wir uns mit der Bitte gewandt, offizielle Briefe in beiden Sprachen unter ihren Mitgliedern zu verbreiten, in denen wir über das Projekt informiert und um Kontaktaufnahme mit uns gebeten haben. An die anderen Interviewpartnerinnen und -partner haben wir uns persönlich gewandt. Dr. habil. Felicja Księżyk und Dr. Irena Prawdzic haben auch am VI. Kulturfestival der Deutschen Minderheit in Polen teilgenommen, das im September 2018 in Wrocław stattgefunden hat. Wir haben es geschafft, unter den Besuchern einige Kontaktpersonen für die Forschung kennenzulernen und ihr Interesse zu wecken.

In Westpommern hat uns unerwartet eine Deutschlehrerin in Choszczno (ehemals Arnswalde), Frau Teresa Wiśniewska, über eine uns bekannte Englischlehrerin zu einem Fest der deutschen Pommern in Dobrzany (Jacobshagen) eingeladen. Infolge dieser Einladung knüpften wir wertvolle Kontakte mit den Sprecherinnen und Sprechern aus Stargard (Stargard) und Świnoujście (Swinemünde). Auf diesem Fest wunderten wir uns über die starke Präsenz der Gäste aus Piła (Schneidemühl), wo wir später zweimal hingefahren sind. Bei dem zweiten Aufenthalt in Piła haben wir Kontakt zum Vorsitzenden der deutschen Minderheit in Złotów (Flatow) aufgenommen und dort auch sehr interessante Interviews durchgeführt.

Dank der im „Wochenblatt“, einer Wochenzeitung der deutschen Minderheit in Opateln, veröffentlichten Anzeige haben uns mehrere Menschen aus ganz Polen und sogar aus Deutschland auf eigene Faust kontaktiert. Wir haben alle Adressen gesammelt und mehrmals mit den Freiwilligen telefoniert. Als eine gute Informationsquelle mit Zugang zu den Informantinnen und Informanten erwiesen sich auch die Museen in Stargard (Muzeum Archeologiczno-Historyczne w Stargardzie) und in Złotów (Muzeum Ziemi Złotowskiej w Złotowie), bei denen wir uns vor Ort erkundigt haben.

Versuche der Kontaktaufnahme verliefen vielgeisig und passten sich den lokalen Umständen wie auch der gesuchten Kontaktgruppe flexibel an. Während jedes Interviews oder Gesprächs mit einer Kontaktperson fragten wir nach anderen bilingualen Sprechern vor Ort oder in der Familie bzw. im Bekanntenkreis, um mit ihnen sprechen zu können. Eine Sprecherin oder ein Sprecher konnte uns einer anderen Person weiterempfehlen, wenn wir das erste Interview angemessen durchgeführt hatten. Diese sogenannte Schneeball-Methode (vgl. Babbie 2009: 213) hat sich als bewährte Methode bei Feldforschungen erwiesen, sodass sie auch im LangGener-Projekt erfolgreich angewandt wurde.

Die Interviews mit der „Generation Deutschland“ (vgl. Centner, Hansen & Nekula in diesem Band) wurden von Dr. habil. Felicja Księżyk ausnahmsweise in der Weihnachtszeit auch in Schlesien durchgeführt, als die Repräsentanten der „Generation Deutschland“ ihre Familien in der Heimat besuchten. Die Kontaktaufnahme in Bayern und in Nordrhein-Westfalen stützte sich auf die guten privaten Beziehungen zweier

Forscherinnen, Dr. Anna Jorroch und Carolin Centner. Die Informantinnen und Informanten in Hamburg gewannen wir im Gegensatz dazu über Prof. Dr. Patrick Zeller von der Universität Hamburg. Durch ihn kamen wir mit Aleksandra Jeszke-Zillmer in Kontakt, der stellvertretenden Vorsitzenden der Deutsch-Polnischen Gesellschaft Hamburg, die uns eine Reihe von Informantinnen und Informanten vermitteln konnte.

In Berlin sind wir direkt zu zwei Familienangehörigen unserer Respondentinnen und Respondenten aus Polen gefahren. Jedoch stellte sich die Schneeball-Methode (und andere Suchmethoden) in der Hauptstadt als nutzlos heraus, weil die dortige polnischsprachige Gemeinde nahe der Grenze zu Polen aus sog. Pendelmigranten bestand, was den projektinternen Suchkriterien nicht entsprach.

Bevor wir zur Feldforschung fuhren, stellten wir eine Liste der Kontaktpersonen zusammen, um später niemanden zu vergessen. Während der Forschung selbst haben wir die Liste modifiziert und um neue Informationen ergänzt. Das Interesse für unser wissenschaftliches Vorhaben (vor allem in Schlesien) übertraf sogar unsere Erwartungen und Möglichkeiten, sodass wir leider wegen der großen Entfernungen sowie der finanziellen und zeitlichen Begrenzung der Dienstreisen nicht alle Personen besuchen konnten.

3 Die Durchführung der Feldforschung

Die Forschung wurde innerhalb von zwei Gruppen bilingualer Sprecherinnen und Sprecher durchgeführt, die bereits charakterisiert wurden. Wir haben versucht, für das Interview Personen zu gewinnen, die den Profilbestimmungen am nächsten entsprechen (siehe Centner, Hansen & Nekula in diesem Band).

Zu den Interviews wurden wir in die Wohnungen, Häuser und Hausgärten unserer Informantinnen und Informanten eingeladen. Wir hatten einen geringeren Einfluss darauf, wer noch am Gespräch teilnehmen würde und wurden manchmal von einer größeren Gesellschaft oder einem unerwarteten Besuch überrascht. Einerseits hat es uns geholfen, eine weitgehende Natürlichkeit des Gesprächs zu erlangen oder ungezwungene Sprachsituationen zu beobachten, andererseits konnten wir manchmal kein für das Korpus akustisch brauchbares Material sammeln. So ist es beispielsweise in Kętrzyn (Rastenburg) passiert, als Prof. Dr. Björn Hansen und Dr. Anna Jorroch ein Interview im Büro der Gesellschaft der Deutschen Minderheit durchführen wollten und insgesamt 5 Kontaktpersonen zusammengekommen sind, sodass kein gesteuertes Gespräch zu einem bestimmten Thema möglich war. Darüber hinaus wollten sich alle zugleich an der Unterhaltung beteiligen, sodass die Aufnahmen für eine spätere Auswertung ungeeignet waren. Dieses aus unserer Sicht zuerst misslungene Treffen war allerdings ein Sprungbrett zu einer weiteren Kontaktaufnahme mit einer der bilingualen Beteiligten.

Wir sind auch Personen begegnet, die den Profilbestimmungen nicht entsprechen, was wir vor dem Gespräch nicht wussten. Auch wenn das der Fall war und ein Interview nicht ins Korpus aufgenommen werden konnte, ist für uns jede Geschichte von großem Wert, weil sie wertvolle historische, soziolinguistische, dialektale, soziologische und linguistische Angaben beinhaltet und eine breite Perspektive schuf. Wir haben also versucht, mit möglichst vielen bilingualen Sprecherinnen und Sprechern ein Interview zu führen. Insgesamt haben wir 124 Personen interviewt, von denen 47 zur „Generation Deutschland“, der jüngeren Generation in Deutschland und 77 zur „Generation Polen“, den

älteren Sprecherinnen und Sprechern in Polen, gehören. Von dem gesammelten Material haben wir Gespräche mit 58 Bilingualen ausgesucht und 78 Stunden der Aufnahmen in beiden Sprachen in das Korpus aufgenommen.

3.1 Die Herangehensweise bei den Interviews

Am Anfang des Interviews haben wir uns vorgestellt und die Forschungsziele erklärt. Wir waren verpflichtet, die Sprecherin bzw. den Sprecher um Einwilligung zu bitten, das Interview für die Forschung verwenden zu dürfen. Dabei haben wir eine von einem deutschen und einem polnischen Juristen in beiden Sprachen erstellte Einwilligungserklärung (beide Dokumente befinden sich im Anhang dieses Bandes) zur Erhebung und Verarbeitung personenbezogener Daten für Forschungszwecke gemäß Art. 13 I a) der europäischen Datenschutz-Grundverordnung (DS-GVO) verwendet. Die Interviewten erklärten sich damit einverstanden, dass die Verarbeitung von personenbezogenen Daten auf Grundlage der europäischen DS-GVO und ihrer nationalen Ausgestaltungen erfolgt. Die Rechtsgrundlage für die Verarbeitung ist ihre Einwilligung gemäß Art. 6 I a) und Art. 9 II a) DS-GVO. Die Interviewten haben mit ihrer Unterschrift bestätigt, dass sie damit einverstanden sind, dass die Daten nach der Aufnahme alsbald pseudonymisiert werden. Das bedeutet, dass aus den Daten die direkten Bezüge auf ihre Person entfernt werden, so dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr der betroffenen Person zugeordnet werden können (Art. 4 Nr. 5 DS-GVO). Wir haben die Kontaktpersonen auch darüber informiert, dass die Unterschriften auf beiden (dem polnischen und dem deutschen) Formularen gesetzlich verlangt werden und die Fragen zu den Erklärungen beantwortet.

Bei den Interviews haben wir uns der bewährten Methode *der teilnehmenden Beobachtung* bedient, die u. a. von Riehl (2014: 48f.) verwendet wurde: Wir haben versucht, einen Kontakt so aufzubauen, dass die Sprecherin oder der Sprecher sich nicht als ein Untersuchungsobjekt fühlte. Die Kontaktperson sollte uns nicht als einen Fremdkörper empfinden und sich in unserer Gegenwart natürlich verhalten können. Diese Einstellung erlaubte den Sprecherinnen und Sprechern mehr von sich selbst zu erzählen (siehe Kaufmann 2010: 28f.). Es ist zwar zeitaufwändig, liefert aber die besten natürlichen Ergebnisse einer Feldforschung.

Nach jeder Dienstreise haben wir einen Brief mit Danksagung vom Institut der Slavistik der Polnischen Akademie der Wissenschaften auf Polnisch und einen Brief auf Deutsch von der Universität Regensburg an jede einzelne Person verschickt. Dies hat uns auch bei der Rückkehr weitere Kontaktaufnahmen erleichtert. Bei einem Besuch bei den Respondentinnen und Respondenten haben wir meistens einige Kleinigkeiten, zum Beispiel Ansichtskarten aus Regensburg, Süßigkeiten oder, wie in Hamburg, Amazon-Gutscheine, mitgebracht.

Bei dem Interview baten wir die Sprecherin oder den Sprecher, sich vorzustellen. Sie oder er sollten uns den Vor- und Nachnamen, das Geburtsdatum und den Geburtsort angeben. Wir haben uns bei dem Gespräch auf den zuvor vorbereiteten Fragebogen gestützt (Anhang 1 dieses Bandes). Während des Gesprächs haben wir darauf geachtet, welche Frage gestellt werden sollte, ohne dass der rote Faden verloren geht. Das Gespräch sollte natürlich verlaufen. Die Informationen für die Sprachbiographien wurden uns allmählich

im Laufe des Interviews gegeben. Wir haben die Kontaktpersonen nicht abgefragt. Falls wir etwas nicht verstanden haben, haben wir um weitere Erklärungen gebeten (vgl. Kaufmann 2010: 72–78). Wir haben emotional belastende Themen (u. a. Kriegserlebnisse, Nachkriegszeitverfolgungen) möglichst vermieden, es sei denn, die Sprecherin oder der Sprecher hat von selbst angefangen, darüber zu erzählen.

Während des Gesprächs wurde das Diktiergerät herausgeholt und auf den Tisch hingelegt. Die Respondentinnen und Respondenten waren sich dessen bewusst, dass das Interview aufgenommen wurde. Dem Gerät haben wir nicht zu viel Aufmerksamkeit geschenkt. Um eine noch höhere Tonqualität zu sichern, hat das Forschungsteam Prof. Dr. Björn Hansen, Carolin Centner und Dr. Irena Prawdzcic bei den Aufnahmen in Hamburg und in Bayern ein Aufnahmegerät mit getrenntem Mikrofon verwendet, das am Kopf der Kontaktperson befestigt wurde. Diese Vorgehensweise war bei den Vertretern der „Generation Deutschland“ möglich, da diese bereits viel Erfahrung mit der Technik gemacht haben (siehe Bučková & Centner in diesem Band).

Wir haben uns vorgenommen, mit jeder Sprecherin oder mit jedem Sprecher ein Interview auf Polnisch und auf Deutsch jeweils 60 Minuten lang zu führen. Mit einigen Sprecherinnen und Sprechern war das nicht möglich und es hat viel kürzer gedauert, als wir geplant hatten. Bei den anderen hatten wir Probleme, die Sprache des Interviews zu wechseln, sodass wir mehr Zeit gebraucht haben. Wir haben unterschiedliche Methoden beim Sprachenwechsel verwendet. Manche Interviewerinnen und Interviewer aus dem LangGener-Team haben versucht, die Kontaktpersonen mit Fragen auf Deutsch bzw. Polnisch zum Sprechen in der einen oder anderen Sprache zu bringen. Die Sprecherinnen und Sprecher haben sich meistens schrittweise auf eine natürliche Art und Weise unserer Sprache angepasst. Wenn diese Methode sich als erfolglos erwies, haben wir einfach die Kontaktperson gebeten, die Sprache zu wechseln. Eine weitere Option war es, dass sich das Interviewer-Team geteilt hat; zum Beispiel hat Prof. Dr. Björn Hansen in Hamburg mit einigen Befragten zunächst das Interview auf Deutsch durchgeführt und im Anschluss kam Dr. Irena Prawdzcic hinzu, um das Interview auf Polnisch weiterzuführen. Wir haben immer versucht, uns frei zu unterhalten, das Interview gleichzeitig aber auch zu steuern. Allerdings haben wir uns bemüht, dass keine künstlichen Situationen entstehen und die Kontaktpersonen nicht die Lust am Gespräch verlieren.

Wir haben die Verwendung von schwierigem Fachwortschatz reduziert und über möglichst natürliche Alltagsthemen gesprochen. Wir haben auch Interesse am Gespräch gezeigt und versucht ein gemeinsames Thema zu finden, auch wenn es nicht im Fokus der Forschung lag. Fotos sind ein guter Anlass zu interessanten Gesprächen. Manchmal ist es uns gelungen, die Kontaktperson dazu zu ermuntern, uns alte Familienfotos, Diplome oder andere Dokumente zu zeigen. Es hat uns ermöglicht, sehr gute, informative Redepassagen aufzunehmen. Zusätzliche Fragen haben die Kontaktaufnahme sowie die Annäherung erleichtert und es ermöglicht, neue Gesprächsbereiche zu finden. Während der Feldforschung haben wir auch allein die Gegend besucht: nahe gelegene Kirchen und Kirchenorganisationen, Freizeitgemeinden, Bibliotheken, den Friedhof, den Markt usw. Dies ergänzt das Bild über die Umgebung, in der die Kontaktpersonen leben und liefert uns möglicherweise wichtige Informationen, die wir meistens mit den Sprecherinnen und Sprechern angesprochen haben.

3.2 Die Teams der Feldforschung

Bei der Feldforschung haben wir in deutsch-polnischen Paaren gearbeitet. Da die nach 1945 an Polen angeschlossenen Gebiete weitläufig sind und die polnischsprachigen Gruppen in ganz Deutschland verstreut sind, haben wir jedem Forschungspaar einzelne Regionen zugeteilt. Im Folgenden werden unsere Erfahrungen beschrieben.

3.2.1 Juli 2018 – Forschungsteam Anna Jorroch & Peter Jorroch – Woiwodschaft Ermland-Masuren

Dr. Anna Jorroch und ihr Ehemann Peter haben die Aufnahmen mit insgesamt 3 Personen in Elk (Lyck) und Giżycko (Lötzen) durchgeführt. Der Kontakt zu den Sprecherinnen und Sprechern wurde über die bereits angesprochene Anzeige im Wochenblatt und über die Verwandten der Mitarbeiterinnen und Mitarbeiter am Sitz der Gesellschaft der Deutschen Minderheit „Masuren“ in Lyck aufgenommen. Sie wurden sehr herzlich und offen empfangen.

3.2.2 Juli/August 2018 – Forschungsteam Carolin Centner & Irena Prawdzcic – Woiwodschaft Westpommern

Im Projekt begann die Feldforschung mit einer Dienstreise im Juli 2018 nach Stettin (parallel zu der Dienstreise nach Elk vom Team Anna und Peter Jorroch), in der insgesamt 11 Personen interviewt wurden. In der Hauptstadt der Woiwodschaft Westpommern haben wir nur ein Interview im Büro der Sozial-Kulturellen Gesellschaft der Deutschen Minderheit in Stettin durchgeführt, denn nur eine Person hat uns ihr Vertrauen geschenkt und gestattete das Treffen. Auch während der Teilnahme am bereits erwähnten Fest der deutschen Pommern in Dobrzany (Jacobshagen) lehnten mehrere Personen den Vorschlag, mit uns zu sprechen, wegen der polnischen Einsprachigkeit ab, die sie deklariert haben. Da wir erklärt haben, dass wir nach zweisprachigen Personen suchen, konnte dies auch der wahre Grund der Ablehnung sein. Die Einstellung der Feiernden uns gegenüber hat sich aber sichtbar von einer lockeren, ungezwungenen zu einer misstrauischen geändert, als sie erfuhren, dass wir Wissenschaftlerinnen sind. Trotzdem gelang es uns, mehrere Kontakte zu knüpfen und Gesprächstermine zu vereinbaren. Wir haben sogar ein Interview direkt während des Festes aufgenommen, das aber akustisch nicht von großem Wert ist, da das Ereignis im Freien mit Musik stattfand. Wir hatten auch die Gelegenheit, die starken Beziehungen zwischen Verbliebenen und Vertriebenen zu beobachten und mit den letzteren zu sprechen. Sie waren uns gegenüber im Gegensatz zu den Verbliebenen sehr aufgeschlossen. Als wir das Vertrauen der Respondentinnen und Respondenten gewonnen hatten, wurden wir überall sehr offen empfangen und sogar mit verschiedenen Unterlagen und Zeitschriften beschenkt, sodass zum Beispiel die gebürtige Stettinerin Dr. Irena Prawdzcic später noch zweimal nach Swinemünde gefahren ist, um sich noch einmal mit den dort kennengelernten Frauen zu treffen.

3.2.3 August 2018 – Forschungsteam Anna Jorroch & Björn Hansen – Woiwodschaft Ermland-Masuren

Insgesamt hat das Team 10 Personen in der Gegend von Kętrzyn (Rastenburg), Giżycko (Lötzen) und Ruciane-Nida (Rudczanny-Nieden) interviewt. Die Kontakte wurden über eine Anzeige im Wochenblatt, über früher gemachte Bekanntschaften und über den privaten Bekannten- und Verwandtenkreis aufgenommen. Alle Termine wurden von Dr. Anna Jorroch vereinbart, und nur einmal ist sie anfangs misstrauisch behandelt worden. Die Gespräche wurden bei den Kontaktpersonen zu Hause, in den Örtlichkeiten der Deutsch-Polnischen Vereine und auch unerwartet auf dem Stocherkahn eines Bilingualen aufgenommen.

3.2.4 August/September/Dezember 2018 – Forschungsteam Felicja Księżyk & Izabela Błaszczyk, Felicja Księżyk & Anna Księżyk – Woiwodschaft Oppeln

In Schlesien haben sich bei uns die meisten Freiwilligen gemeldet. Insgesamt wurden 26 Personen von der „Generation Polen“ und 13 Personen von der „Generation Deutschland“ interviewt. Ausnahmsweise hat die Last der Feldforschung in der Region hauptsächlich Dr. habil. Felicja Księżyk getragen, die mehrere Gespräche in beiden Sprachen allein aufgenommen hat. Teilweise wurde sie bei der Arbeit von ihrer Schwägerin und von einer Kollegin unterstützt. Die Interviews wurden in privaten Wohnungen und Häusern durchgeführt. Als zu Weihnachten die jüngeren Menschen ihre Eltern besucht haben, wurde diese Gelegenheit für weitere Aufnahmen genutzt.

3.2.5 September 2018 – Forschungsteam Carolin Centner & Barbara Alicja Jańczak – Woiwodschaft Westpommern, Pommern und Ermland-Masuren

Die nächste Dienstreise, während der insgesamt 15 Personen interviewt wurden, umfasste ein ausgedehntes Gebiet (von Koszalin/Köslin über Świdwin/Schivelbein, Gdańsk/Danzig bis Elbląg/Elbing). Alle Treffen wurden noch vor der Reise vereinbart und die Routen, Unterkünfte sowie Verkehrsmittel vorher ausgewählt und gebucht. Einige Personen, die die Forscherinnen während dieser Feldforschung besuchten, wurden auch während der Dienstreise in der Stettiner Gegend als gesprächsbereite Gesprächspartnerinnen und -partner von den Sprecherinnen und Sprechern aufgezeigt. Die Interviewerinnen wurden überall erwartet und offen in den privaten Wohnungen und Häusern empfangen. Trotz der guten Vorbereitung wurden die Forscherinnen nicht immer vor unerwarteten Situationen verschont. In Danzig zum Beispiel fand das Treffen im Büro der Danziger Deutschen Minderheit statt, zu dem um dieselbe Uhrzeit mehrere Frauen gekommen waren. Das Team hat die Ankommenden gebeten, an dem Interview der Reihe nach teilzunehmen, musste aber für jedes Gespräch einen sehr kurzen Zeitrahmen bestimmen, damit die Kontaktpersonen nicht zu lange warten mussten.

3.2.6 Oktober 2018 und Oktober 2019 – Forschungsteam Anna Zielińska & Irena Prawdzcic – Woiwodschaft Großpolen und Woiwodschaft Westpommern

Insgesamt wurden 11 Personen während zweier Dienstreisen in die Region interviewt. Das heutige Großpolen war bei der Antragstellung nicht im Forschungsprojekt vorgesehen und die Recherchen erfolgten aufgrund früher geführter Feldforschungen. Es hatte sich herausgestellt, dass eine relativ zahlreiche Gruppe von Bilingualen in der Region lebt, die in der Zeit von 1918 bis 1939 ein deutsch-polnisches Grenzgebiet war. Aus diesem Grund konnten wir auch diesmal nicht alle Gespräche ins Korpus einbeziehen, da die Geburtsorte der Respondentinnen und Respondenten, die sich ein paar Kilometer nach Westen bzw. nach Osten befinden, sich dies- oder jenseits der damaligen Grenze wiederfanden; denn die Biographien der in Polen (also nach 1945) geborenen Personen passten nicht in das von uns skizzierte Profil. Trotz dieses formalen Kriteriums haben wir mehrere wertvolle Interviews durchgeführt, die in den privaten Wohnungen und Häusern oder im Büro der Deutschen Sozial-Kulturellen Gesellschaft in Schneidemühl stattfanden.

3.2.7 Dezember 2018/April 2019 – Forschungsteam Carolin Centner & Irena Prawdzcic und Björn Hansen & Carolin Centner – Bundesland Bayern

Insgesamt hat das Team mit 12 Informantinnen und Informanten gesprochen. Der Zugang zu den Kontaktpersonen in Deutschland stützte sich hauptsächlich auf private Beziehungen. Die Immigrantinnen und Immigranten aus Polen bilden zwar eine zahlreiche Migrantengruppe in Deutschland, wurden aber rechtlich nie als Minderheit anerkannt. Es gibt also keine vom Staat unterstützte Institution, an die wir uns hätten wenden können. Darüber hinaus haben wir vermutet, dass nicht alle Repräsentantinnen und Repräsentanten der Zielgruppe sich um die polnischen Organisationen konzentrieren würden (zum Beispiel um die römisch-katholischen Kirchen mit den polnischsprachigen Gottesdiensten), weil sie sich auf Grund ihrer Familiengeschichte nicht unbedingt mit der polnischen Kultur identifizieren.

Bei der Organisation der Dienstreise in Bayern setzte sich stark die Forscherin und Regensburgerin Carolin Centner ein, deren Eltern teilweise einen polnischsprachigen Bekanntenkreis haben. Dank der Unterstützung durch eine befreundete Familie in Augsburg wurden zwei Personen an einem Tag in die Wohnung von dieser Familie eingeladen, mit denen (und mit dem Gastgeber) die Forscherinnen der Reihe nach sprechen konnten. An diesem Tag wurde auch ein zusätzliches Interview mit einem jungen Ehepaar aus der Nachbarschaft durchgeführt. Die Augsburger Gruppe widersprach den ursprünglichen Vermutungen, da sie sowohl mit der Kirche mit den polnischsprachigen Gottesdiensten als auch mit der polnischen Kultur stark verbunden ist.

Die restlichen Aufnahmen wurden in Regensburg und in kleineren Ortschaften in der Gegend ausschließlich in privaten Häusern und Wohnungen durchgeführt.

3.2.8 Februar 2019 – Forschungsteam Irena Prawdzcic & Björn Hansen – Bundesland Hamburg

In Hamburg wurden insgesamt 9 Personen interviewt, die sich untereinander sehr gut kannten. Der Kontakt wurde, wie bereits erwähnt, über die Universität Hamburg hergestellt. Die vorher vereinbarten Gespräche haben wir sowohl in den privaten Wohnungen bzw. Häusern der Respondentinnen und Respondenten geführt als auch in Büros, in einer Schule und in einem Hotel. Während der Interviews haben wir ein Headset-Mikrofon genutzt, das die Informantinnen und Informanten gar nicht störte, da manche von ihnen auch beruflich ähnliche Geräte benutzen und daran gewöhnt waren. Die Schneeball-Methode hat uns vor Ort zu zwei neuen Personen gebracht, die bei der ersten Anfrage an einer Teilnahme am Projekt nicht interessiert waren.

3.2.9 Februar 2019 – Forschungsteam Iga Kościółek & Barbara Alicja Jańczak – Bundesland Berlin

Die Dienstreise nach Berlin resultierte in 3 Interviews. Die Suche nach weiteren Kontaktpersonen misslang, weil alle Bekannten der Informantinnen und Informanten ständig abwechselnd sowohl in der deutsch- als auch in der polnischsprachigen Umgebung in beiden Ländern leben.

3.2.10 Februar 2019 – Forschungsteam Anna Jorroch & Peter Jorroch – Bundesland Nordrhein-Westfalen

Insgesamt wurden 11 Informantinnen und Informanten interviewt. Die Kontaktpersonen wurden im privaten Bekanntenkreis von Peter Jorroch gesucht, die in Wuppertal wohnhaft sind. Sie haben uns zu Hause empfangen und sich gern mit uns unterhalten. Die freundschaftlichen Beziehungen haben die Möglichkeit eines natürlichen Gesprächs geschaffen, bei dem auf andere Themen als mit den anderen, bisher fremden Kontaktpersonen eingegangen wurde.

4 Realisierung der Forschungspläne und die Ergebnisse

Die Feldforschungspläne haben wir in der Praxis an manchen Stellen modifiziert. Wir wollten in gemischten Forschungspaaren arbeiten, in denen eine Person das deutsche und die andere das polnische Team repräsentiert. Diese Konstellation sollte den Gesprächspartnerinnen und -partnern möglichst natürliche und sprachlich günstigste Bedingungen bieten. Letztendlich haben wir einige Male diese Vorgabe nicht befolgt. In und um Piła/Schneidemühl haben nur die Forscherinnen aus Polen gearbeitet. In Schlesien hat überwiegend Dr. habil. Felicja Księżyk allein oder mit ihrer Schwägerin, Anna Księżyk, recherchiert. In Bayern haben die deutschen Forscherinnen und Forscher zwei Gespräche aufgenommen. In Wuppertal und in Masuren war Dr. Anna Jorroch tätig, die von ihrem Mann, Peter Jorroch, unterstützt wurde.

Auch die Dauer der Interviews ließ sich nur schwer kontrollieren. Gemäß unserem Vorhaben sollten wir uns eine Stunde auf Deutsch unterhalten und eine Stunde auf Polnisch. Darüber hinaus sollten wir uns mit jeder Informantin und jedem Informanten zweimal getrennt und unter vier Augen treffen. Das war in der Realität nicht umsetzbar.

Im Rahmen der Feldforschung wurden die Aufnahmen gesammelt, die nicht nur den Ansprüchen für das Korpus genügen, sondern auch zu einem wichtigen Objekt weiterer soziolinguistischer Forschung werden dürften.

Literatur

- Babbie, Earl (2009): *Podstawy badań terenowych*. Warszawa: Wydawnictwo Naukowe PWN.
- Kaufmann, Jean-Claude (2010): *Wywiad rozumiejący*. Warszawa: Oficyna Naukowa.
- Online-Lexikon zur Kultur und Geschichte der Deutschen im östlichen Europa: <https://ome-lexikon.uni-oldenburg.de/regionen/grenzmark-poses-westpreussen>, letzter Zugriff am 05.03.2021.
- Riehl, Claudia Maria (2014): *Sprachkontaktforschung. Eine Einführung*. 3., überarbeitete Aufl. Tübingen: Narr Verlag.

Transkriptionskonventionen

Aneta Bučková – Irena Prawdzcic

Abstract:

Transcription conventions

The chapter introduces the transcription rules used in the LangGener project. It briefly summarizes the project goals, which are essential for the described principles. Although the rules are in principle based on the orthography of the transcribed languages, phenomena typical of spoken language are taken into account as well. These include disfluencies, such as rendered topic, insertion, anacoluthon, false starts etc.

Das Korpus der bilingualen Sprache besteht aus drei unterschiedlichen Subkorpora:

- 1) dem deutsch-polnischen Korpus der „Generation Polen“,
- 2) dem deutsch-polnischen Korpus der „Generation Deutschland“,
- 3) dem deutsch-tschechischen Korpus der Spätaussiedlerinnen und Spätaussiedler und der Migrantinnen und Migranten.

Alle drei Subkorpora umfassen verschiedene Sprachenpaare (Deutsch-Polnisch, Deutsch-Tschechisch) und dienen teilweise unterschiedlichen Forschungszielen. Diese sind im Einzelnen: 1. Dokumentation der aussterbenden deutschen Dialekte in Polen, 2. Untersuchung der Typen von morphosyntaktischen Replikationen im Rahmen jeder einzelnen Generation und Feststellung, ob die bei der „Generation Polen“ beobachteten Replikationstypen auch von der „Generation Deutschland“ genutzt werden, sowie die Analyse und der Abgleich dieser Replikationen bei den Gruppen der Spätaussiedlerinnen und Spätaussiedler und der Migrantinnen und Migranten aus der Tschechoslowakei, 3. Untersuchung der Sprachbiographien anhand der soziolinguistischen Annotation (siehe Jańczak, Nekula & Zielińska in diesem Band), 4. Untersuchung der Beziehung zwischen Sprachbiographie und morphosyntaktischen Replikationen, 5. Korrelation der Typen von Replikationen in den Sprachpaaren Deutsch-Polnisch und Deutsch-Tschechisch, um eine translinguale Typologie der Replikationen zu entwickeln (in diesem Sinne dient das deutsch-tschechische Subkorpus als ein Kontrollkorpus für das deutsch-polnische Subkorpus und umgekehrt). Dessen ungeachtet sind die Subkorpora gleichzeitig einer größeren Einheit untergeordnet und realisieren somit auch gemeinsame Ziele (Untersuchung der Replikationen, siehe Bučková, Centner, Księżyk & Prawdzcic in diesem Band). Deswegen mussten wir einen Kompromiss zwischen den einzelnen Besonderheiten und der aus der Nutzerperspektive nötigen Einheitlichkeit schließen und gemeinsame Transkriptionskonventionen erarbeiten.

Im Wesentlichen wurden die Regeln anhand der Korpora für gesprochene Sprache des Tschechischen Nationalkorpus erstellt (Komrsková 2020; Poukarová 2020). Es mag verwunderlich sein, warum nicht etwa auf das in der deutschsprachigen Korpuslinguistik und Gesprächsanalyse etablierte Transkriptionssystem GAT bzw. GAT2 zurückgegriffen wurde (vgl. Hirschmann 2019: 76). Ein Grund, weshalb wir uns gegen diese Konventionen entschieden haben, war ihre Komplexität, die zwar für die Gesprächsanalyse geeignet, aber für die Forschungsfragen in unserem Projekt nicht notwendig war (siehe Hansen & Nekula sowie Bučková & Nekula in diesem Band). Außerdem galt es, nicht nur die Besonderheiten des gesprochenen Deutsch, sondern auch des Tschechischen und des Polnischen zu berücksichtigen. Darüber hinaus wäre etwa eine phonetische Transkription aufgrund der unterschiedlichen Qualität einzelner Aufnahmen zum Teil nur bedingt möglich gewesen (zu den technischen Aspekten siehe Bučková & Centner in diesem Band).

Eines unserer Hauptforschungsziele war, wie oben erwähnt, die Dokumentation der aussterbenden deutschen Dialekte in Polen. Deswegen wurde die halbphonetische Transkription bei den deutschsprachigen Interviewteilen innerhalb der „Generation Polen“ (gebildet von Personen geb. in den 1930er und 1940er Jahren) eingeführt. Weiter wurde die Anwendung der halbphonetischen Transkription innerhalb der „Generation Polen“ bei den polnischen Interviewteilen festgelegt, jedoch nur bei den Aufnahmen mit den schlesischen Sprecherinnen und Sprechern, weil die dialektalen Merkmale nur in ihrer Sprache häufig aufgetreten sind. Die Sprecherinnen und Sprecher außerhalb Schlesiens bedienen sich der polnischen Umgangssprache. Obwohl die Standardisierung und Kodifizierung des Schlesischen (Wyderka 2018) weit fortgeschritten ist und der schlesische Dialekt selbst seit 2007 sogar mit dem Kode **szl** im Rahmen der Norm ISO639-3 versehen ist, richteten wir uns nach den projektinternen Transkriptionsregeln, weil wir die Gespräche mit den Interviewten nicht auf Schlesisch, sondern auf Standardpolnisch durchgeführt haben. Deswegen wäre die Anwendung der schlesischen Orthographie im Gegensatz zum Vermerk der dialektalen Merkmale unangemessen gewesen. Die genaue Auflistung der vermerkten schlesischen Merkmale im Polnischen und der dialektalen Merkmale im Deutschen wird in von Jorroch, Księżyk & Zielińska (in diesem Band) veranschaulicht und detailliert besprochen (vgl. auch Księżyk & Zielińska 2020: 8f.). Da die Abfrage der annotierten sprachlichen Erscheinungen nur anhand der orthographischen Transkription abrufbar wird, wurden die Transkripte normalisiert und an die standardsprachlichen orthographischen Konventionen in der jeweiligen Sprache angepasst. Das bedeutet, dass das Korpus nur über die normalisierte Orthographie durchsuchbar ist. Es besteht aber die Möglichkeit, von der orthographischen Aufzeichnung auf die halbphonetische Aufzeichnung umzuschalten und umgekehrt.

Die orthographische Transkription wurde in den deutschen und polnischen Interviewteilen der „Generation Deutschland“, in den polnischen Interviewteilen der „Generation Polen“ sowie in den deutschen und tschechischen Interviewteilen der Migrantinnen und Migranten und der Spätaussiedlerinnen und Spätaussiedler angewandt. Sie hält sich grundsätzlich an die aktuellen Orthographieregeln der jeweils transkribierten Sprache. Das bedeutet, dass die Sprachdaten bis zu einem gewissen Grad normalisiert wurden. So wurden beispielsweise Reduktionen in der Aussprache nicht berücksichtigt, wie in *weil ich in Tschechisch einen Vierer gehabt habe* (TU_HAG_SP_DE). Bei dieser Klausel geht aus der Aufnahme hervor, dass die Sprecherin den ə-Laut am Ende des Wortes *habe* nicht ausspricht. Phonetisch wäre an dieser Stelle daher [hap] zu transkribieren, was der Wort-

form *hab* entspräche. Eine solche Reduktion ist aber weder für die soziolinguistischen noch für die morphosyntaktisch orientierten Forschungsfragen des LangGener-Projekts relevant (vgl. Dresing & Pehl 2015: 26). Die Normalisierung macht in diesem sowie in analogen Fällen das Korpus besser durchsuchbar, denn die Ergebnisse einer Abfrage schließen somit etwa mehrere phonetische Varianten einer Wortform ein.

Zur Transkriptionskonvention gehört auch die Grenzsetzung für Segmente. Wir segmentierten die Aufnahme grundsätzlich in Einheiten, die ein konjugiertes Verb beinhalten, einschließlich Interjektionen und Häsitationen (siehe unten), zum Beispiel *wir hatten vier Semester # _Matfyz gehabt* (KG_MUN_MI_DE) (Sonderzeichen werden unten erläutert; eine tabellarische Übersicht befindet sich am Ende des Kapitels). Daraus ergeben sich morphosyntaktische Einheiten, die sich für die Identifikation der fokussierten Sprachkontakterscheinungen auf der Ebene der syntaktischen Phrasen gut eignen (siehe siehe Bučková, Centner, Książek & Prawdziej in diesem Band). Darin besteht eine Ähnlichkeit mit Brazils *A Grammar of Speech* (1995). Dieser Autor (ebd.: 42ff.) schlägt als eine minimale syntaktische Anforderung für eine Einheit der gesprochenen Sprache ein nominales und ein verbales Element vor. Im LangGener-Projekt wurde auf das nominale Element verzichtet, weil die untersuchten Sprachen Tschechisch und Polnisch im Unterschied zu dem von Brasil (1995) behandelten Englisch nicht immer eine Nominalphrase in der Subjektfunktion verlangen. Eine syntaktisch bedingte Segmentierung der gesprochenen Sprache anhand von finiten Verbformen schlägt für das Tschechische auch Müllerová (1994: 42ff.) vor. Andere Segmentierungskriterien finden sich etwa im Tschechischen Nationalkorpus, in dem sich die Segmentgrenzen grundsätzlich an der Wortanzahl orientieren (Kopřivová et al. 2017). Weitere Segmentierungsweisen gehen zum Beispiel nach semantischen Kriterien (Passonneau & Litman 1997) oder nach der Prosodie (Edlund et al. 2005) vor. Zusätzlich zu der syntaktischen Anforderung an den Segmentinhalt gab es ein weiteres Kriterium. Ein Segment durfte nicht mit einer Pause beginnen, aber falls möglich sollten seine Grenzen eben am Ende einer Pause gesetzt werden.

In der gesprochenen Sprache kommen aber verschiedene Erscheinungen vor, die die Grenzsetzung komplexer machen. Diese wurden bei den Entscheidungen über Segmentgrenzen stets berücksichtigt. Darunter unterscheiden wir u. a.: ein wiedergegebenes Thema, retrospektive elementare Diskurseinheiten, diskursstrukturierende Einheiten, Parenthese, Substitution, Fehlstart, Ellipse sowie Wiederholungen und partielle Wiederholungen, die im Folgenden jeweils kurz vorgestellt werden.

Unter einem wiedergegebenen Thema (*rendered topic*) verstehen Kibrik & Podlesskaja (2003: 7) eine Nominalphrase, die syntaktisch nicht in der Klausel integriert ist. Wie aus (1) ersichtlich wird, referiert die Sprecherin auf die Nominalphrase *moje kamarádka* ‚meine Freundin‘ mit dem Personalpronomen *ona* ‚sie‘, welches an dieser Stelle als resumptives Pronomen fungiert (siehe Bučková & Nekula in diesem Band). Beide Nominalphrasen wurden als Teil desselben Segments interpretiert.

- (1) # **moje** _ **kamarádka** **ona** bydlela _ v Čechách
 1SG.POSS Freundin 3SG wohn.3SG.PST in Böhmen.LOC
meine Freundin sie hat in Böhmen gewohnt
 (TU_HAG_SP_CZ)

Als retrospektive elementare Diskurseinheiten (retrospective edus) werden kleinere Einheiten als Klausen bezeichnet, die nach der Phrase oder Klausen folgen, auf die sie sich beziehen (Kibrik & Podlesskaja 2003: 10). Im Unterschied zu *rendered topic* fehlt bei der retrospektiven elementaren Diskurseinheit das anaphorische Pronomen, das auf die Nominalphrase referiert. Stattdessen greift die Sprecherin in (2) die Phrase *u babiččiny sestry* ‚bei Omas Schwester‘ mit dem Nachtrag *u mojí pratety* ‚bei meiner Großtante‘ wieder auf. Auch hier wurden beide nacheinander folgenden Präpositionalphrasen in einem Segment belassen.

- (2) **u _ u babiččiny_ sestry** u mají **u mojí pratety _**
 bei Oma.ADJ Schwester.GEN bei – bei 1SG.POSS.GEN Großtante.GEN
 se ubytovala
 REFL unterkomm.3SG.PST
bei Omas Schwester bei meiner Großtante ist sie untergekommen
 (NT_GAU_MI_CZ)

Diskursstrukturierende Elemente (*discourse structuring elements*) sind Wörter oder Phrasen, die syntaktisch eliminierbar sind, das heißt sie enthalten keinen propositionalen Inhalt, können nicht erfragt oder negiert werden (Kolaković et al. 2022: 181; Birzer 2015: 85f.). Das Segment in (3) enthält zum Beispiel die Verbform *říkám* ‚ich sage‘, die nicht ein Teil des propositionalen Inhalts der Klausen *takže to trvalo těch pět let* ‚also es hat diese fünf Jahre gedauert‘ ist. Die Aussage des Sprechers wurde nicht etwa in *takže to trvalo / říkám / těch pět let* geteilt (die Schrägstriche signalisieren potenzielle Segmentgrenzen), sondern das diskursstrukturierende Element wurde als ein Bestandteil miteinbezogen.

- (3) **takže to trvalo říkám těch pět let**
 also DEM dauer.3SG.PST sag.1SG DEM.GEN.PL fünf Jahr.GEN.PL
also hat es wie gesagt diese fünf Jahre gedauert
 (LQ_UNT_MI_CZ)

Bei Anakoluth handelt es sich um einen Bruch in der syntaktischen Konstruktion, der zu einer syntaktisch oder semantisch fehlerhaften Äußerung führt (Karlík 2017; Kolaković et al. 2022: 182). In (4) fängt die Sprecherin den Satz mit *die waren* an, dann bricht sie ihn aber ab und fährt mit *hatten die Schwierigkeiten* fort. Auch hier wurde auf eine Aufteilung nach der ersten finiten Verbform *waren* verzichtet, sondern die abgebrochene und neu angefangene Aussage wird als eine Einheit betrachtet.

- (4) **die waren politisch** also a* **politisch hatten** die Schwierigkeiten
 (MQ_UNT_MI_DE)

Konstruktionen mit abhängigen Infinitivphrasen werden zwar in der Literatur als satzwertig bezeichnet (vgl. Wöllstein 2016: 859), aber in unserem Korpus wurden sie in einem Segment zusammen mit dem regierenden Satz transkribiert (vgl. (5)). Diese Entscheidung ist aufgrund der Analyse einzelner Segmente gefallen, die ohne eine Infinitivphrase wenig Platz für eine Interpretation bieten.

- (5) na ja und da da wuchs in mir die Sehnsucht **auch etwas mehr zu sehen zu wissen**
(KN_MUN_MI_DE)

Eine Parenthese (*parenthetical insertion*) bedeutet nach Crible (2018: 75) „propositional segments [...] located in a sequence of fluencemes to which it adds some background information without directly modifying the content of the utterance“. In (6) schob die Sprecherin zwei Fragen ein *no jak ?* ‚na wie ?‘ und *jak moge ?* ‚wie kann ich ?‘, weil ihr ein Wort fehlte. Als ihr das fehlende Wort *Gefangene* vorgesagt wurde, hat sie ihre Aussage fortgesetzt. Wenn die Parenthesen als selbstständige Einheiten betrachtet würden, erhielten wir zwei Segmente, die getrennt unverständlich wären, nämlich *Kriegs ** und *Gefangene aber aber a Franzose* und eine angeblich nicht komplette Klausel *er war eigentlich #*, die in der Tat mit der Nominativ-Ergänzung vervollständigt wurde:

- (6) er war eigentlich # **no jak ?** Kriegs * **jak moge ?** Gefangene aber aber a Franzose
Franzose
,er war eigentlich # **na wie ?** Kriegs * **wie kann ich ?** Gefangene aber aber und
Franzose Franzose‘
(IL_PIL_II_GP_DE)

Peter Öhl (2009: 400) versteht unter Substitution „die Ersetzung von Elementen in ihren funktionalen Paradigmen“. Im Beispiel (7) korrigiert sich die Sprecherin selbst und substituiert das konjugierte Verb *sind* mit *gingen* und später mit der Singularform *ist sägen gegangen*. Dabei ändert sie das Subjekt von *sie* auf das zusammengesetzte *mein Vater und ich*. Die Grenzen innerhalb dieses Segments könnten auch anders gesetzt werden, zum Beispiel *na und dann sind sie nach Feierabend/gingen für _/aha nach Feierabend ist mein Vater und ich noch immer Holz sägen gegangen für ein Stück trockenes Brot und saure Milch* oder *na und dann sind sie/nach Feierabend gingen für _/aha nach Feierabend ist mein Vater und ich noch immer Holz sägen gegangen für ein Stück trockenes Brot und saure Milch*. Bei derartiger dreifacher Gliederung müsste hier eine zusätzliche Replikation (das Fehlen des Subjektes in der Klausel *nach Feierabend gingen für _*) annotiert werden. Diese Annotation ist überflüssig, wenn die Substitution berücksichtigt wird:

- (7) na und dann **sind sie** nach Feierabend **gingen** für _ aha nach Feierabend **ist mein Vater und ich** noch immer Holz **sägen gegangen** für ein Stück trockenes Brot und saure Milch
(HD_SWI_GP_DE)

Unter einem Fehlstart versteht Crible (2018: 73) „interruptions that leave a segment syntactically and/or semantically incomplete and where no elements from the previous, abandoned context are taken up in what follows“. Im Segment (8) beginnt die Sprecherin ihre Aussage mit dem Modalverb *müssen*, bricht sie aber ab und drückt den Gedanken mit einer anderen Formulierung aus. Die Bildung eines getrennten Segments aus dem Fehlstart *und musste* würde zur Annotation einer Replikation führen (das Fehlen des Subjektes), was in diesem Fall falsch wäre, weil es zur Umformulierung gekommen ist:

- (8) **und musste** die erste und zweite Klasse habe ich zusammen gemacht
(IL_PIL_II_GP_DE)

Im Allgemeinen versteht man unter Ellipse Auslassungen von Satzgliedern. Wie aus (9) ersichtlich wird, fehlen in dem Segment zwei Verben *ging an* und *kamen*. Man könnte diese Aussage anders gliedern und in drei Segmente teilen: *da kamen schon die Russen/ und dann der Kriech/und die Polen*. Jedoch wären die beiden kurzen Segmente ohne die erste Klausel unverständlich.

- (9) da kamen schon die Russen und dann der Kriech und die Polen
,da kamen schon die Russen und dann [**ging**] der Kriech [**an**] und die Polen [**kamen**]*
(PQ_GDY_II_GP_DE)

In einem Segment werden auch Wiederholungen (d. h. Verdoppelung/Verdreifachung etc. eines Wortes, zum Beispiel *musiał musiał* ‚musste musste‘) und partielle Wiederholungen (d. h. abgebrochene und reparierte Worte, zum Beispiel *wm * wmurować* ‚einm* einmauern‘, *le* lese*) erfasst. Wie aus (10) ersichtlich ist, würde das Setzen der Grenze nach dem wiederholten Wort *musiał* das zusammengesetzte Prädikat *musiał wmurować/musste einmauern* trennen, wodurch die Segmente unverständlich wären: *to takie pręty musiał/musiał wm * wmurować tam wszystko*.

- (10) *to takie pręty musiał musiał wm * wmurować tam wszystko* nie ?
,solche Montageeisen **musste musste er einm * einmauern** dort alles nicht ?
(IL_PIL_II_GP_DE)

Eine andere Gliederung würde in (11) zu einer überflüssigen Annotation führen (das Fehlen des Subjektes): *und mit wenn ich le */Deutsch lese*:

- (11) und mit wenn ich **le *** Deutsch **lese**
(IL_PIL_II_GP_DE)

In die Transkripte haben wir zur Kennzeichnung relevanter Phänomene Sonderzeichen eingeführt. Diese werden im Folgenden jeweils kurz beschrieben. Bei Wörtern, die in dem Transkriptionsprozess den Transkribentinnen und Transkribenten unbekannt waren, oder speziellen Ausdrücken wurde das Wort zwischen \$-Symbole geschrieben, zum Beispiel *sedmdesát čtyři _ <\$>seme<\$> byli s št* moje p* čtyři nejlepší kámoši jsme byli vod _prvního až do c* desátýho května*, vierundsiebzig <\$>seme<\$> waren wir v* meine F* vier beste Freunde waren wir vom _ersten bis z* zehnten Mai* (KL_HEI_SP_CZ). Die spitzen Klammern wurden bei jeglichen Symbolen (siehe unten) verwendet, damit diese Symbole bei der Arbeit mit den weiteren Tools (siehe Bučková & Centner in diesem Band) erhalten bleiben und nicht anders dekodiert werden. Am Ende des Transkriptionsprozesses wurden die spitzen Klammern automatisch gelöscht.

Es wurde alles transkribiert, d. h. auch Versprecher, Wiederholungen, Selbstkorrekturen etc. Ein abgebrochenes Wort wurde durch * gekennzeichnet und nicht ergänzt, zum Beispiel in *Tick* Ticket; aber wir konnt* ich # ich wusste es nicht* (EQ_HER_MI_DE). Die Transkribentinnen und Transkribenten verzichteten auf Verbesserungen oder Korrek-

turen, um den transkribierten Text möglichst realitätsgetreu wiederzugeben. So wurden auch ungewöhnliche oder aus der Sicht der Grammatik von der Baseline abweichende Konstruktionen unverändert transkribiert, zum Beispiel *und ich wollte mit die reden* (CH_BER_GD_DE). Diese Vorgehensweise entspricht auch anderen variationslinguistisch orientierten Korpora, wie es beispielsweise Zimmer et al. (2020: 221) in ihrem Vorgehen bei der Erstellung des DNam-Korpus für das Deutsche in Namibia beschreiben. Darüber hinaus wurde die Wortstellung nicht geändert.

Das Inventar an Interpunktion ist reduziert worden. Es wurden keine . , ; ! benutzt. Für Ergänzungsfragen und Entscheidungsfragen wurden Fragezeichen ? nach einem Leerzeichen verwendet, zum Beispiel # *_jak se jmenuje Škule na čes*_česky ?*, # *_wie heißt Schule auf Tsch* Tschechisch ?* (LS_ZEI_SP_CZ). Der Unterstrich ₁ kennzeichnet eine Pause innerhalb eines Segments. Es wurde nicht zwischen kurzen und langen Pausen unterschieden und es wurde keine minimale Pausenlänge festgelegt. Bei Teilwörtern in deutscher Sprache wurde der Bindestrich - benutzt (zum Beispiel *EDV-Programme*).

Schwer verständliche Passagen wurden in Klammern transkribiert, zum Beispiel *aber #so Aufsatz schreiben konnte (er) in Tschechisch nicht* (KI_HEI_SP_DE). Bei völlig unverständlichen Passagen wurde in Klammern eine Schätzung der Anzahl der unverständlichen Wörter aufgeschrieben, zum Beispiel *manchmal ist so dass ich mit die Deutsche im Haus (2) Polnisch mit die Leute die was da wohnen sprechen* (XL_PIL_GP_DE). Wenn die Anzahl nicht zu erkennen war, zählte eine unverständliche Passage als ein Wort.

Großbuchstaben wurden nur bei Eigennamen geschrieben, nicht am Anfang einer Replik. Im Deutschen wurden im Einklang mit der Norm die Substantive großgeschrieben. Mit der Orthographie der Modellsprache wurden auch Wörter notiert, die nicht adaptiert (d. h. zum Beispiel unflektiert) in einer anderssprachigen Aussage benutzt wurden (zum Beispiel *to się za *to było vorbei* - LC_BUR_GD_PL). Wenn die sprachlichen Übernahmen aus der Modellsprache adaptiert (d. h. zum Beispiel flektiert) wurden, transkribierten wir sie nach den orthographischen Regeln der Replikasprache (zum Beispiel *färbt die Derme* (SO_NIE_GP_DE); *musiałbym zdawać pryfungi* - LC_BUR_GD_PL). Diese sowie andere Sprachkontakterscheinungen wurden später auch als solche annotiert (siehe Bučková, Centner, Księżyk & Prawdzc in diesem Band).

Paraverbale Einheiten wurden einheitlich wiedergegeben. Im deutsch-tschechischen Subkorpus wurde *hmm* bei Geräuschen als Reaktion auf den Gesprächspartner bei Zustimmung und *emm* bei Nichtübereinstimmung, Ablehnung geschrieben. Dies steht im Einklang mit der Transkriptionskonvention der gesprochensprachlichen Korpora des Tschechischen Nationalkorpus, in dem die Transkripte des deutsch-tschechischen Subkorpus veröffentlicht werden. Im deutsch-polnischen Subkorpus wurde auf die Transkription dieser Geräusche hingegen konsequent verzichtet. Häsitationsgeräusche wurden in allen Subkorpora transkribiert, und zwar als #.²

¹ Ursprünglich wurden für die Kennzeichnung einer Pause zwei Punkte <.> festgelegt. Dies stellte sich für die spätere Korpusuche als ungünstig heraus, denn Punkte haben in der Korpusabfragesprache CQL bereits eine andere Funktion.

² Das anfangs vorgesehene Zeichen für Häsitationsgeräusche war das At-Zeichen <@>. Ähnlich wie bei der Pause (siehe oben) wurde dies in <#> umgewandelt, weil @ in CQL bereits anderweitig belegt ist.

Zusammenfassend lässt sich sagen, dass die erstellten Transkriptionskonventionen sowohl die unterschiedlichen im Projekt enthaltenen Forschungsbereiche berücksichtigen als auch eine benutzerfreundliche Einheitlichkeit gewährleisten. Ausgehend von den Regeln des Tschechischen Nationalkorpus eignen sie sich für die orthographische bzw. halb-phonetische Transkription gesprochener Sprache in allen untersuchten Varietäten. Die gewählte Segmentierung erlaubt eine klausenbezogene Annotation von Sprachkontakterscheinungen, deren Typologisierung eines der Forschungsziele des LangGener-Projekts darstellt.

Sonderzeichen	Funktion
\$	unbekannte Wörter/spezielle Ausdrücke
*	abgebrochene Wörter
?	nach Ergänzungs- und Entscheidungsfragen
_	Pause innerhalb eines Segments
#	Häsitationsgeräusch
-	innerhalb deutscher Komposita
()	schwer verständliche Passagen

Tabelle 1: Sonderzeichen in den LangGener-Transkriptionskonventionen

Literatur

- Birzer, Sandra (2015): Generally speaking, connectivity and conversation management combined: the functions of Russian *voobšče govorja* and Polish *ogólnie mówiąc*. *Russian Linguistics* 39.1, 81–115.
- Brazil, David (1995): *A Grammar of Speech. Describing English language*. Oxford: Oxford University Press.
- Crible, Ludivine (2018): *Discourse Markers and (Dis)fluency. Forms and functions across languages and registers*. Amsterdam: John Benjamins.
- Dresing, Thorsten/Pehl, Thorsten/Schmieder, Christian (Hgg.) (2015). Manual (on) Transcription. Transcription Conventions, Software Guides and Practical Hints for Qualitative Researchers. 3. engl. Aufl. Marburg: dr. dresing et pehl GmbH. Online verfügbar unter: <http://www.audiotranskription.de/english/>, letzter Zugriff am 13.08.2020.
- Edlund, Jens/Heldner, Mattias/Gustafson, Joakim (2005): Utterance segmentation and turn-taking in spoken dialogue systems. In: Bernhard Fisseni, Hans-Christian Schmitz, Bernhard Schröder, Petra Wagner (Hgg.), *Computer Studies in Language and Speech. Beiträge zur GLDV-Tagung 2005 in Bonn*. Frankfurt/Main: Peter Lang, 576–587.
- Hirschmann, Hagen (2019): *Korpuslinguistik. Eine Einführung*. Stuttgart: J.B. Metzler.
- Karlík, Petr (2017): ANAKOLUT. In: Petr Karlík, Marek Nekula, Jana Pleskalová (Hgg.), *CzechEncy - Nový encyklopedický slovník češtiny*. Online verfügbar unter: <https://www.czechency.org/slovník/ANAKOLUT>, letzter Zugriff am 24.11.2020.
- Kibrik, Andrej/Podlesskaja, Vera (2003): K sozdaniju korpusov ustnoj ruskoj reči: principy transkribovanija. *Naučno-techničeskaja informacija* 2.10, 5–13. Online verfügbar unter: http://www.philol.msu.ru/~otipl/new/main/people/kibrik-aa/files/Corpora_speech_transcription@S&I_2003.pdf, letzter Zugriff am 20.11.2020.
- Kolaković, Zrinka/Jurkiewicz-Rohrbacher, Edyta/Hansen, Björn/Filipović-Đurđević, Dušica/Fritz, Nataša (2022): *Clitics in the wild: Empirical studies on microvariation of the pronominal, reflexive and verbal clitics in Bosnian, Croatian and Serbian*. Berlin: Language Science Press.

- Komrsková, Zuzana (2020): *Transkripce v korpusu ORTOFON. Příručka ČNK*. Online verfügbar unter: <https://wiki.korpus.cz/doku.php/cnk:ortofon:pravidla>, letzter Zugriff am 13.08.2020.
- Kopřivová, Marie/Benešová, Lucie/Škarpová, Marie/Komrsková, Zuzana/Lukeš, David/Poukarová, Petra (2017): Instrukce k přepisu ve stopě ort, 1–11. Online verfügbar unter: <https://trnka.ff.cuni.cz/mluvka2/wiki/doku.php?id=ort>, letzter Zugriff am 16.01.2022.
- Księżyk, Felicja/Zielińska, Anna (2020): Language shifts in the language biographies of immigrants from Upper Silesia residing in Germany. *Multilingua: Journal of Cross-Cultural and Interlanguage Communication*, Band 40, Berlin/Boston: Walter de Gruyter Mouton, 675–706. DOI: 10.1515/multi-2019-0127.
- Müllerová, Olga (1994): *Mluvený text a jeho syntaktická výstavba*. Prag: Academia.
- Öhl, Peter (2009): Sprachwandel und kognitive Ökonomie: Zur Grammatikalisierung und Substitution von Satzkonnectoren. *Linguistische Berichte* 220/2009, 393–438. Online verfügbar unter: http://oehl.gesuss-info.de/CV/Writings/09_KONN%26Oekonomie_LB.pdf, letzter Zugriff am 12.12.2020.
- Passonneau, Rebecca/Litman, Diane (1997): Discourse segmentation by human and automated means. *Computational Linguistics* 23.1, 103–139. Online verfügbar unter: <https://www.aclweb.org/anthology/J97-1005>, letzter Zugriff am 20.11.2020.
- Poukarová, Petra (2020). *Transkripce v korpusu ORAL. Příručka ČNK*. Online verfügbar unter: <https://wiki.korpus.cz/doku.php/cnk:oral:pravidla>, letzter Zugriff am 13.08.2020.
- Wöllstein, Angelika (Hg.) (2016): *Die Grammatik. Unentbehrlich für richtiges Deutsch*. 9., vollständig überarbeitete und aktualisierte Aufl. Berlin: Dudenverlag.
- Wyderka, Bogusław (2018). O standaryzacji języka śląskiego. *Kwartalnik Opolski* 2018,1, 3–18.
- Zimmer, Christian/Wiese, Heike/Simon, Horst/Zappen-Thomson, Marianne/Bracke, Yannic/Stuhl, Britta/Schmidt, Thomas (2020): Das Korpus Deutsch in Namibia (DNam): Eine Ressource für die Kontakt-, Variations- und Soziolinguistik. *Deutsche Sprache* 3, 210–232. DOI: 10.37307/j.1868-775X.2020.03.03, letzter Zugriff am 26.09.2020.

Vom Gespräch zu annotierten Daten

Aneta Bučková – Carolin Centner

Abstract:

Data processing: from conversation to annotated data

This chapter aims at presenting a technical, semi-automated pipeline processing audio-data to an annotated ELAN-file. Therefore, the recording technique, the cutting process as well as the tools used within the project (OCTRA, WebMAUS, EXMARaLDA and ELAN) are introduced and discussed with regard to their advantages and disadvantages in processing oral data. This chapter can be considered a description of best practices and is therefore intended to facilitate future attempts to create spoken language corpora.

1 Einleitung

Dieses Kapitel konzentriert sich auf die technische Verarbeitung der narrativen Interviews. Diese wurden während der Feldforschung (siehe Jorroch & Prawdzc in diesem Band) aufgenommen und das Ziel dieses Textes ist es, den Weg von der Aufnahme über die technische Bearbeitung und Transkription bis hin zu ihrer Überführung ins Korpus (siehe Woźniak in diesem Band) nachzuzeichnen. Die folgenden Abschnitte berichten über die beim Aufnehmen verwendete Technik, das Schneiden, die Transkription einschließlich der Tools OCTRA, WebMAUS, EXMARaLDA und ELAN und über die Aufbereitung der Metadaten zwecks Archivierung und Bereitstellung für Korpusnutzerinnen und -nutzer. Die weitere Verarbeitung der annotierten Datei bis hin zum Korpus-Interface wird von Woźniak (in diesem Band) besprochen.

Mit dieser Beschreibung und Diskussion des technischen Weges von einem Gespräch im Feld bis hin zu einer korpusreifen Datei möchten wir die empirischen und die technischen Aspekte der linguistischen Forschung ansprechen, die in der Literatur häufig nicht explizit gemacht werden. Sie stellen jedoch einen wesentlichen Teil der Arbeit mit natürlichen Sprachdaten und somit auch einen nicht zu vernachlässigenden Teil des LangGener-Projektes dar. Diese wollen wir im Sinne der guten wissenschaftlichen Praxis bewusst explizit machen. Darüber hinaus sind wir der Meinung, dass andere, auf gesprochene Sprache orientierte Projekte von unseren in diesem Text aufgeführten Erfahrungen profitieren können.

2 Aufnahmetechnik

Während die inhaltliche Seite der Interviews in dem Kapitel „Praktische Durchführung der Feldforschung im deutsch-polnischen Team“ (Jorroch & Prawdziej in diesem Band) beschrieben wird, konzentriert sich dieser Abschnitt auf die technische Umsetzung der Aufnahmen. Diese wurden mit unterschiedlichen Aufnahmegegeräten durchgeführt. In der ersten Phase der Feldforschung wurde der Audio-Recorder Olympus LS-P1 verwendet. Sein Vorteil ist, dass er verhältnismäßig klein und dadurch unauffällig ist (siehe Abbildung 1). Er kann während des Interviews beispielsweise auf den Tisch gelegt werden und wirkt nicht ablenkend. Gleichzeitig befindet sich das Mikrofon aber dadurch nicht in unmittelbarer Nähe der Sprecherin oder des Sprechers, was eine niedrigere Aufnahmequalität bedeuten kann.

Für die weiteren Phasen der Feldforschung wurde die technische Ausrüstung um weitere Aufnahmegegeräten und Mikrofone ergänzt. Es wurde mit dem Aufnahme-Recorder Zoom H-4PRO gearbeitet, der entweder mit einem Headset oder mit einem Lavalier-Mikrofon verkabelt wurde. Neben dem jeweiligen angeschlossenen Mikrofon, das die Sprecherin oder der Sprecher mittels eines Headsets am Kopf trug bzw. durch eine Klammer am Kragen befestigt hatte, wurde gleichzeitig auch mit dem Mikrofon des Aufnahmegegeräts aufgenommen, das auf dem Tisch zwischen der Sprecherin oder dem Sprecher und der Interviewerin oder dem Interviewer lag.¹ Auf diese Weise sind zwei Aufnahmen gleichzeitig entstanden, die vor dem Schneiden zu einer Tonspur heruntergemischt wurden (siehe Abschnitt 3).

Die Verwendung des Headsets und des Lavalier-Mikrofons bringt eine eindeutig bessere Qualität der Aufnahme mit sich. Eine gute Aufnahmequalität bedeutet eine erhebliche Erleichterung des Transkriptionsprozesses einschließlich seiner automatisierten Schritte (siehe unten), denn die Sprache ist dadurch besser verständlich und lässt sich einfacher bearbeiten. Dadurch ermöglicht sie auch eine genauere linguistische Analyse, beispielsweise aus Sicht der Phonetik. Der Nachteil dieser Mikrofone besteht im Falle unseres Forschungsprojektes darin, dass sie durch ihre Befestigung am Körper bzw. an der Kleidung der Interviewten die Aufnahmesituation unnatürlicher machen. Die Interviewten werden dadurch ungewollt stets daran erinnert, dass sie aufgenommen werden, was zu einer erhöhten Selbstkontrolle führen kann. Sie können sich nicht spontan im Raum bewegen, wenn sie beispielsweise Fotos holen, die mit der erzählten Lebensgeschichte zusammenhängen, oder einen Kaffee kochen möchten. Bei den Interviewten, die durch ihre Kriegserlebnisse traumatisiert sind und eine negative Erfahrung mit der Überwachung seitens eines totalitären Regimes gemacht haben, kann eine solche Aufnahmetechnik noch hemmender wirken. Aus diesem Grund wurde vor allem bei den älteren Interviewten der „Generation Polen“ (siehe Centner in diesem Band) auf das Headset bzw. Lavalier-Mikrofon verzichtet. Abbildung 1 zeigt eine Aufnahmesituation mit Prof. Dr. Björn Hansen und zwei Interviewten. Das Aufnahmegegerät ist auf dem Tisch platziert (linke untere Ecke im Bild) und zeigt zu den Interviewten.

¹ Im deutsch-tschechischen Teilprojekt wurde außerdem bei einigen Interviews auf die Aufnahmesoftware SpeechRecorder zurückgegriffen (Jänsch & Draxler 2020). Für die Details bezüglich dieses Vorgehens im Kontext der Sprachbiographieforschung siehe Bučková (in Vorbereitung).



Abbildung 1: Aufnahmesituation ohne Headset
(Interviewer B. Hansen in Gızıcko, August 2018)

Nicht alle während der Feldforschung entstandenen Aufnahmen sind in das Korpus eingeflossen. Die Auswahl bezog sich außer auf die Fragestellung und die daraus resultierende Abgrenzung der untersuchten Gruppen (siehe Bučková & Nekula und Centner in diesem Band) auch auf korpuslinguistische Kriterien. Ein Grund für den Ausschluss einer Aufnahme war auch mangelnde technische Qualität, die beispielsweise durch viele Störgeräusche verursacht wurde.

Für die Unterscheidung einzelner Subkorpora (Hansen & Nekula sowie Bučková & Nekula in diesem Band) war es außerdem wichtig, dass der tschechisch- bzw. polnischsprachige und der deutschsprachige Interviewteil voneinander abgegrenzt werden können. Eine hohe Frequenz der Sprachmischung bedeutet bei der gegebenen Fokussierung auf Musterentlehnungen sog. *data noise*, der eine quantitative Analyse der im deutsch- oder slavischsprachigen Subkorpus vertretenen Sprachvarietäten problematisch machen würde. Kurze Passagen, in denen die Sprache gewechselt wurde, stellten aber keinen Ausschlussgrund dar, solange dieser Sprachwechsel nicht ständig vorgekommen ist. Sie bilden die Realität mehrsprachiger natürlicher Sprachdaten ab, die durch die Narration elizitiert werden sollten.

Des Weiteren wurden diejenigen Aufnahmen nicht ins Korpus aufgenommen, in denen es mehrere Sprecherinnen und Sprecher gab, die sich oft ins Wort fielen, und zwar insbesondere dann, wenn nicht alle beteiligten Personen zum untersuchten Sample gehörten. Die Anwesenheit mehrerer Sprecherinnen und Sprecher bedeutete eine erhebliche

Erschwerung des Transkriptionsprozesses und der darauffolgenden sprachformbezogenen Analyse der Sprachproduktion einzelner Sprecherinnen und Sprecher.

3 Schneiden der Aufnahmen

Die Originalaufnahmen wurden unkomprimiert und in voller Länge auf einer externen Festplatte gespeichert. Aus den in Abschnitt 2 angeführten Gründen war eine Auswahl von zu transkribierenden Passagen und damit das Schneiden der Aufnahmen notwendig. Im Laufe des Projekts wurden für die einzelnen Schritte des Transkriptionsprozesses kürzere, ungefähr zehn Minuten lange Abschnitte als *best practice* etabliert, denn sie sind handlicher und erlauben eine effizientere Bearbeitung innerhalb des semiautomatischen Transkriptionsprozesses (siehe Abschnitt 4). Der erste Schritt bei der Bearbeitung der durchgeführten Interviews war daher das Schneiden.

Herausgeschnitten wurden alle Störungen, die während des Interviews aufgetreten sind und mit aufgenommen wurden. Diese sind beispielsweise das Handyklingeln und Gespräche mit Dritten über Telefon, lange Hustenanfälle, Zwischenrufe dritter Personen und die Reaktion auf diese. Des Weiteren wurden diejenigen Passagen herausgeschnitten, in denen der Übergang von einer Sprache in die andere von den Interviewerinnen und Interviewern eingeleitet wurde oder in denen die Kontaktdaten zu anderen Interviewten vermittelt wurden. Nicht zuletzt sind Interviewabschnitte weggefallen, in denen von den Interviewten persönliche Informationen der Interviewerinnen und Interviewer erfragt wurden, und die Reaktionen auf diese Fragen.

Auch der Teil des Interviews, in dem das Forschungsprojekt vorgestellt wurde, konnte herausgeschnitten werden. Das lag im Ermessen der Bearbeiterinnen. Der Grund für ein etwaiges Behalten im Interview war beispielsweise eine individuelle Beeinflussung der Interviewten durch diese einführenden Worte, die sich am Anfang der darauffolgenden Narration etwa an einer direkten Reaktion auf eine Formulierung feststellen lässt. Ähnlich vorgegangen wurde auch beim Anbieten von Kaffee und Kuchen. Dieser Moment ist grundsätzlich nicht für die autobiographische Erzählung relevant, es sei denn, er führt zum Erzählen einer persönlichen Geschichte oder zum Mitteilen von Rezepten, die unter anderem aus lexikalischer Sicht wertvoll sein können.

Im Gegensatz dazu wurden beim Schneiden diejenigen Teile der Interviews belassen, die Angaben wie Eigennamen, Ortsnamen, Geburtsdaten etc. enthalten und die Interviewten betreffen. Diese sollen aus Datenschutzgründen nicht veröffentlicht werden, sind jedoch für eine soziolinguistische Analyse im Rahmen des Projektes bedeutsam. Sie wurden in den weiteren Schritten des Bearbeitungs- und Transkriptionsprozesses anonymisiert (siehe Abschnitt 4) und in den Metadaten des Korpus verallgemeinert festgehalten (siehe Abschnitt 6).

In der ersten Phase wurden die Interviews in dem Programm Audacity manuell geschnitten (Audacity® 2020). Dieses Programm wurde ausgewählt, weil es kostenlos und benutzerfreundlich ist. Nach dem Importieren der Rohaufnahmen im WAV-Format wurden diese zu Monoaufnahmen heruntergemischt, im Falle von einer gleichzeitigen Aufnahme aus mehreren Mikrofonen (siehe Abschnitt 2) wurde aus allen Tonquellen eine Monospur erstellt. Das Format der Aufnahme wurde zudem auf 24-bit PCM reduziert, wodurch die weitere Verarbeitung in OCTRA ermöglicht wurde. Diese Schritte dienen der Reduzierung

der Dateigröße und sind für die semiautomatische Bearbeitung essenziell (siehe Abschnitt 4). Nach der Komprimierung wurde die Aufnahme in Audacity abgespielt und an geeigneten Stellen, das heißt in ungefähr zehnminütigen Abständen und in den Pausen, geschnitten. Die so entstandenen Abschnitte wurden wieder als WAV-Dateien exportiert.

Beim Exportieren wurde jeder Audiodatei ein Name zugewiesen. Dieser beginnt mit einem Kürzel, in dem zwei Buchstaben, der Wohnort der Interviewten und ihre Zugehörigkeit zu einer der untersuchten Generationen und Gruppen (siehe Centner sowie Bučková & Nekula in diesem Band) kodiert werden. Dann folgen ein Label für die Sprache, die in der Datei gesprochen wird (bzw. die überwiegt), und eine Nummer, die die Reihenfolge des Abschnitts im Interview festlegt (siehe Abbildung 2).

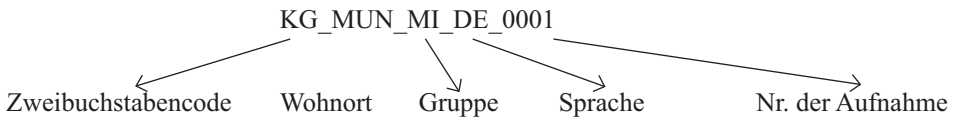


Abbildung 2: Audiodateiname

Der Prozess des Schneidens wurde in einer Tabelle, der sog. Masterliste, dokumentiert. In der Masterliste wurden alle Audiodateinamen zusammen mit den Zeitangaben (Anfang und Ende innerhalb des Interviews, Dauer), der Information über die Sprache und darüber, ob die Audiodatei transkribiert werden soll, festgehalten. Zusätzlich wurde der Inhalt jeder Audiodatei mittels festgelegter Labels zusammengefasst (beispielsweise die Gesprächsthemen Arbeit, Kindheit, Familie usw.). Bei Audiodateien, die nicht transkribiert werden sollten, informierten andere Labels (zum Beispiel Abwesenheit des Interviewten, Störgeräusche) über die Gründe, warum die Passage ausgeschlossen wurde. Abbildung 3 zeigt einen Ausschnitt aus der beschriebenen Masterliste.

Dateiname	Label 1	Label 2	Label 3	von	bis	Dauer	Sequenz #	Transkription	Sprache
NT_GAU_MI_DE_0001	Familie	Kindheit	totalitäres Regime	0:00:00.000	0:14:02.219	0:14:02.219	01	ja	de
NT_GAU_MI_DE_0002	Schule	totalitäres Regime	Familie	0:14:02.219	0:32:21.395	0:18:19.176	02	ja	de
NT_GAU_MI_DE_0003	Störgeräusche			0:32:21.395	0:32:45.848	0:00:24.453	03	nein	de
NT_GAU_MI_DE_0004	Sprache	Auswanderung	Schule	0:32:45.848	0:45:56.869	0:13:11.021	04	ja	de
NT_GAU_MI_DE_0005	Schule	Sprache	totalitäres Regime	0:45:56.869	0:51:33.355	0:05:36.487	05	ja	de
NT_GAU_MI_DE_0006	Störgeräusche			0:51:33.355	0:51:39.114	0:00:05.759	06	nein	de
NT_GAU_MI_DE_0007	Familie	Reisen/Urlaub	totalitäres Regime	0:51:39.114	1:06:12.010	0:14:32.896	07	ja	de
NT_GAU_MI_DE_0008	Familie	totalitäres Regime	Kindheit	1:06:12.010	1:21:39.498	0:15:27.487	08	ja	de
NT_GAU_MI_DE_0009	Kindheit	Reisen/Urlaub	totalitäres Regime	1:21:39.498	1:36:42.296	0:15:02.798	09	ja	de
NT_GAU_MI_DE_0010	Auswanderung	Arbeit	totalitäres Regime	1:36:42.296	1:43:57.808	0:07:15.512	10	ja	de
NT_GAU_MI_CZ_0011	Übergangsversuche ins CZ			1:43:57.808	1:44:04.644	0:00:06.836	11	nein	cz
NT_GAU_MI_CZ_0012	Auswanderung	Familie	totalitäres Regime	1:44:04.644	1:58:19.973	0:14:15.328	12	ja	cz
NT_GAU_MI_CZ_0013	andere Themen			1:58:19.973	2:01:00.769	0:02:40.796	13	nein	cz
NT_GAU_MI_CZ_0014	Auswanderung	totalitäres Regime	Bekannte	2:01:00.769	2:12:31.308	0:11:30.539	14	ja	cz
NT_GAU_MI_CZ_0015	Alltag	totalitäres Regime		2:12:31.308	2:23:29.249	0:10:57.941	15	ja	cz
NT_GAU_MI_CZ_0016	Schule	Sprache	totalitäres Regime	2:23:29.249	2:43:46.803	0:20:17.554	16	ja	cz
NT_GAU_MI_CZ_0017	andere Themen	Schule		2:43:46.803	3:01:06.499	0:17:19.696	17	ja	cz
NT_GAU_MI_CZ_0018	Studium	Kontakte zu Heimat	Bekannte	3:01:06.499	3:14:19.543	0:13:13.044	18	ja	cz
NT_GAU_MI_CZ_0019	Bekannte	Arbeit	Krankheiten	3:14:19.543	3:25:32.147	0:11:12.604	19	ja	cz
NT_GAU_MI_CZ_0020	Abwesenheit des Probanden			3:25:32.147	3:25:42.175	0:00:10.028	20	nein	cz
NT_GAU_MI_CZ_0021	Kontakte zu Heimat	Identität	Bekannte	3:25:42.175	3:40:16.866	0:14:34.691	21	ja	cz
NT_GAU_MI_CZ_0022	andere Themen			3:40:16.866	3:40:55.349	0:00:38.483	22	nein	cz
NT_GAU_MI_CZ_0023	Sprache			3:40:55.349	3:59:38.093	0:18:42.744	23	ja	cz
NT_GAU_MI_CZ_0024	Sprache			3:59:38.093	4:11:49.045	0:12:10.952	24	ja	cz

Abbildung 3: Masterliste (Schneiden in Audacity)

Im Laufe der Transkriptionsarbeiten wurde das Schneiden ebenfalls im Transkriptionstool OCTRA (OCTRA – modern web-application for the orthographic transcription 2020) ermöglicht. Da das Schneiden in OCTRA die Anzahl der manuell durchzuführenden Schritte im Vergleich zu Audacity reduziert, sind wir im deutsch-tschechischen

Teilprojekt teilweise zum Schneiden in OCTRA übergegangen. Die folgenden Abschnitte beziehen sich auf diese alternative Vorgehensweise beim Schneiden mithilfe von OCTRA, die im Laufe des Projektes in der Verarbeitung deutsch-tschechischer Aufnahmen das oben beschriebene manuelle Schneiden mit Audacity ersetzt.

Wie oben erwähnt, ist das Schneiden mit OCTRA im Vergleich zu dem ursprünglichen Vorgehen effizienter. Dennoch war es aber nicht möglich, auf Audacity (oder ein vergleichbares Programm) zugunsten von OCTRA völlig zu verzichten. Der Grund dafür ist, dass die Rohdaten, das heißt Interviews in voller Länge in einem Stereoformat oder aufgenommen durch mehrere Mikrofone gleichzeitig, nicht in OCTRA hochgeladen werden können. Deshalb mussten die ersten Schritte des Schneideprozesses trotzdem in Audacity durchgeführt werden. Die Originalaufnahme wurde komprimiert und in größere Abschnitte (etwa nach Sprache) geschnitten, sodass die entstandenen Dateien die Größe von 300 MB nicht überschritten.

Die in Audacity vorbereiteten Audiodateien wurden anschließend schrittweise in OCTRA hochgeladen. Jede Datei wurde in dem 2D-Editor abgehört und es wurden nach den oben beschriebenen Kriterien Segmentgrenzen² gesetzt (für den Umgang mit OCTRA siehe Abschnitt 4). In diesem Schritt wurde der Inhalt der Segmente nicht transkribiert. Statt einer Transkription wurde notiert, ob das Segment transkribiert werden soll (JA bzw. NEIN) und was es beinhaltet (dabei wurde auf die oben erwähnten Labels zurückgegriffen). Nach einer derartigen Segmentierung wurde die Audiodatei von OCTRA anhand der Segmentgrenzen automatisch geschnitten. Der Output sind nicht nur WAV-Dateien, sondern auch eine TXT-Datei mit Metainformationen (siehe Abbildung 4).

Die von OCTRA automatisch produzierte TXT-Datei hat die gleiche Funktion wie die oben beschriebene Masterliste. Sie enthält die Namen der einzelnen Audiodateien, den Namen der Inputdatei, die geschnitten wurde, den Anfang und die Dauer jeder Audiodatei (in Millisekunden und Sample) und die Transkription (sprich in diesem Kontext die Information darüber, ob die Datei transkribiert werden soll und was ihr Inhalt ist).

Name	File	SecondsStart	SecondsDuration	SampleStart	SampleDuration	SampleRate	Transcript
IL_REG_SP_DE_0001.wav	IL_REG_SP_DE.wav	0	4.410657596	0	194510		44100 NEIN, Vorstellung des Projektes
IL_REG_SP_DE_0002.wav	IL_REG_SP_DE.wav	4.410657596	479.0874376	194510	21127756		44100 JA, Sprache, Familie, Religion
IL_REG_SP_DE_0003.wav	IL_REG_SP_DE.wav	483.4980952	450.4943084	21322266	19866799		44100 JA, Schule, Familie, totalitäres Regime
IL_REG_SP_DE_0004.wav	IL_REG_SP_DE.wav	933.9924036	399.581746	41189065	17621555		44100 JA, totalitäres Regime, Familie
IL_REG_SP_DE_0005.wav	IL_REG_SP_DE.wav	1333.57415	9.476938776	58810620	417933		44100 NEIN, Übergangsversuche ins Tschechische
IL_REG_SP_CZ_0006.wav	IL_REG_SP_CZ.wav	0	4.8669161	0	214631		44100 NEIN, Übergangsversuche ins Tschechische
IL_REG_SP_CZ_0007.wav	IL_REG_SP_CZ.wav	4.8669161	562.4334694	214631	24803316		44100 JA, Schule, Sprache, Identität
IL_REG_SP_CZ_0008.wav	IL_REG_SP_CZ.wav	567.3003855	535.9315646	25017947	23634582		44100 JA, Arbeit, Sprache, Familie
IL_REG_SP_CZ_0009.wav	IL_REG_SP_CZ.wav	1103.23195	548.2129252	48652529	24176190		44100 JA, Kontakte zu Heimat, Familie, Sprache
IL_REG_SP_CZ_0010.wav	IL_REG_SP_CZ.wav	1651.444875	513.8022676	72828719	22658680		44100 JA, Familie, Sprache, Kontakte zur Heimat
IL_REG_SP_CZ_0011.wav	IL_REG_SP_CZ.wav	2165.247143	432.8897506	95487399	19090438		44100 JA, Kontakte zu Heimat, Familie, Sprache
IL_REG_SP_CZ_0012.wav	IL_REG_SP_CZ.wav	2598.136893	13.31741497	114577837	587298		44100 NEIN, andere Themen

Abbildung 4: Masterliste (Schneiden in Audacity und OCTRA)

Auf diese Weise wurde in OCTRA jeder in Audacity bearbeitete Audioabschnitt geschnitten. Das heißt, dass dadurch mehrere Masterlisten für ein Interview entstanden und dass die

² Im Zusammenhang mit dem Schneiden in OCTRA sind längere Abschnitte einer Audiodatei entstanden, die anschließend transkribiert werden sollten. In den Tools EXMARALDA und ELAN wurden innerhalb dieser Aufnahmeabschnitte nach syntaktischen Kriterien Segmente anderer Art erstellt. Diese werden von Bučková & Prawdziej in diesem Band vorgestellt.

Nummerierung nicht durchgehend war, sondern bei jedem Audioabschnitt wieder bei 0001 begonnen wurde. Die einzelnen Masterlisten mussten daher manuell zusammengeführt und die Nummerierung in der Tabelle sowie in den Dateinamen angepasst werden. Hinzu kam noch die Tatsache, dass die MS Office-Version des Betriebssystems Windows in der von OCTRA erstellten Tabelle die Zeitangaben in Sekunden nicht korrekt erkennt. Durch wenige einfache Schritte in MS Excel lässt sich dieses Problem jedoch schnell beheben.

OCTRA wurde nicht nur zum Schneiden, sondern auch im nächsten Schritt zur orthographischen Transkription verwendet. Die ungefähr zehn Minuten langen Dateien, die als Output bei dem eben beschriebenen Prozess entstanden sind, dienten als Input für diese Transkription, die im Abschnitt 4.1 beschrieben wird. Sie wurden nach dem Schneiden heruntergeladen und gespeichert und zum Zweck der Transkription später wieder einzeln in OCTRA hochgeladen.

4 Tools zur Transkription und Annotation

Bei der Wahl der technischen Verarbeitungsmöglichkeiten sollte auf die Anforderungen, die das spezifische Projekt mit sich bringt, eingegangen werden. Das Ziel des Projekts und damit auch Prämisse bei der Überlegung zur Korpusstruktur war, die erhobenen Daten auch für andere Wissenschaftlerinnen und Wissenschaftler zur Verfügung zu stellen. Daher muss bei der Erstellung einer Pipeline zur Verarbeitung der Daten darauf geachtet werden, dass diese möglichst auch dauerhaft einem breiteren wissenschaftlichen Publikum zugänglich sind.

Im vorliegenden Projekt wurden viele Daten gesprochener Sprache erhoben, die bei einer rein manuellen Transkription, beispielsweise in ELAN, einen großen Aufwand dargestellt hätten, sodass wir um eine möglichst automatisierte Lösung bemüht waren. Die automatisierte Transkription bringt jedoch nicht nur Vorteile mit sich, da dadurch auch Fehler entstehen. Daher war es wichtig, den Grad an Automatisierung zu finden, der die Arbeit mit den mündlichen Daten erleichtern würde, ohne zu viele Fehler zu produzieren, sodass die Korrektur nicht länger dauern würde als eine manuelle Transkription. Eine ähnliche Erfahrung spricht Schmidt (2018: 222) an, der für das FOLK Gesprächskorpus³ bei der Spracherkennungstechnologie auf eine Worterkennungsrate von unter 50% kommt. Er weist darauf hin, dass auch weitere automatisierte Verfahren wie beispielsweise *silence detection* oder *speaker diarization* so hohe Fehlerquoten aufweisen, dass der manuelle Transkriptionsaufwand als niedriger geschätzt wurde.

Bei den im LangGener-Projekt vorliegenden Daten konnte keine automatische Spracherkennung eingesetzt werden, die die Aufnahmen in Transkripte umwandeln würde, da die bilingualen Interviewten oft von der Standardnorm abwichen, die solchen Programmen zugrunde liegt, sodass die Software mit der automatisierten Transkription Probleme hatte. Dennoch wurden einige Programme eingesetzt, die die Arbeit mit den mündlichen Aufnahmen erleichtert haben und sich so für ähnliche Projekte mit einer analogen Zielsetzung empfehlen.

³ Das Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK) ist ein Korpus des IDS Mannheim.

Wie Kisler et al. (2017: 326f.) beobachten, macht es die fortschreitende Stabilität und Funktionalität des Internets möglich, auf Programme, die lokal auf Rechnern installiert werden, zu verzichten und stattdessen vermehrt auf webbasierte Tools zu setzen. Neben den technischen Vorteilen lässt sich so auch eine bessere Vernetzung herstellen, wenn nicht nur eine Person an der Umsetzung der technischen Pipeline beteiligt ist, sondern, wie im vorliegenden Fall, ein ganzes Team. Diese webbasierten Tools werden oft innerhalb einer Plattform zusammengefasst, so beispielsweise bei ‚Common Language Resources and Technology Infrastructure‘ (CLARIN), innerhalb derer wiederum das ‚Bavarian Archive for Speech Signals‘ (BAS) integriert ist, dessen Tools für das Projekt genutzt wurden.

Insgesamt wurden in dem Projekt fünf verschiedene Tools im Bearbeitungs-, Transkriptions- und Annotationsprozess verwendet. Die nachstehende Graphik zeigt, wie die Programme aufeinander aufbauen und welcher Output jeweils generiert wird:

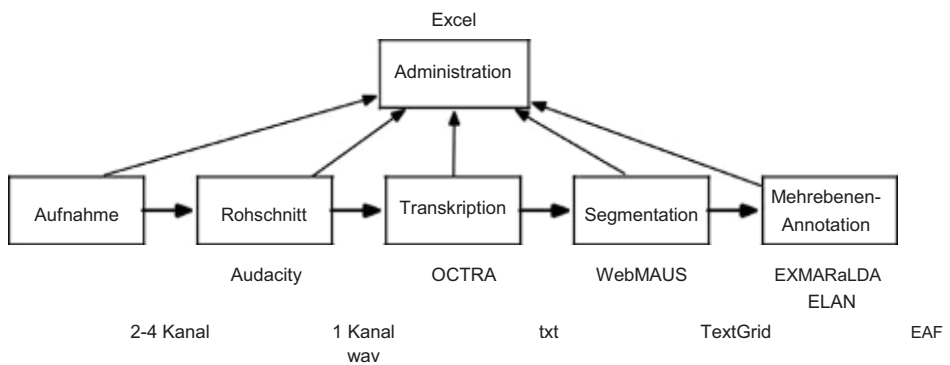


Abbildung 5: Tools im Transkriptions- und Annotationsprozess⁴

Im Folgenden soll der Transkriptions- und Annotationsteil der technischen Pipeline, wie sie im Projekt gehandhabt wurde, mit ihren Vor- und Nachteilen skizziert werden.

4.1 OCTRA

OCTRA (kurz für **O**rthographic **T**ranscription) ist eine Web-App, die vom BAS der LMU München zur Verfügung gestellt wird. Das LangGener-Team wurde dankenswerterweise während des ganzen Transkriptionsprozesses von Mitarbeiterinnen und Mitarbeitern der LMU betreut. Nachdem die Aufnahmen bereits in Audacity oder teilweise auch in OCTRA (siehe Abschnitt 2 zur Aufnahmetechnik) geschnitten wurden, entstand in OCTRA das manuell erstellte Rohtranskript. Dieses orientiert sich an den in den Transkriptionskonventionen (vgl. Bučková & Prawdziej in diesem Band) vorgestellten Transkriptionsregeln. In diesem Schritt wurde noch nicht auf die Unterteilung nach unterschiedlichen

⁴ Diese Abbildung wurde freundlicherweise von Dr. Christoph Draxler vom Institut für Phonetik und Sprachverarbeitung der LMU München für das Projekt erstellt und zu Publikationszwecken zur Verfügung gestellt.

Sprecherinnen und Sprechern geachtet, alles wurde in der gleichen Spur transkribiert. Dies ist dem Umstand geschuldet, dass OCTRA nur eine lineare Transkription zulässt; zugleich ist dies aber auch zeitsparend. Die Aufteilung der Transkription nach Sprecherinnen und Sprechern wurde anschließend in EXMARaLDA vorgenommen.

Um eine Transkription in OCTRA zu beginnen, wird zunächst eine Audiodatei, die 300 MB nicht überschreitet, im WAV-Format in OCTRA geladen. Daraufhin öffnet sich das Fenster, in dem transkribiert wird. Dabei kann man nach persönlicher Präferenz zwischen dem *Dictaphone Editor*, dem *Linearen Editor* und dem *2D-Editor* wählen. OCTRA ist mit verschiedenen Features ausgestattet, die die Transkription erleichtern. Dazu gehören Schaltflächen für *Abspielen*, *Stopp*, *Wiederholung*, *Zurück* sowie Regler für Geschwindigkeit und Lautstärke beim Abspielen der Aufnahme. Diese Features sind ebenfalls über Tastenkombinationen, die über eine Pop-Up-Liste einsehbar sind, wählbar. Auch die projektspezifischen Transkriptionsregeln können vom BAS-Team in das Tool eingebunden werden und sind für die Transkribentinnen und Transkribenten jederzeit aufrufbar. Eine Übersichtsfunktion zeigt die bereits transkribierten Segmente innerhalb der Datei.

Nach dem Hochladen der Audiodatei in die Web-App wird diese zunächst in kleinere Segmente unterteilt. Dabei orientiert man sich an der visuellen Darstellung der Tonspur (siehe Abbildung 6), um Pausen im Gespräch festzumachen. Der Vorteil der kleineren Segmente liegt darin, dass sich die Transkribentinnen und Transkribenten nicht zu lange Passagen merken müssen, ehe sie niedergeschrieben werden. Über Tastenkombinationen oder die Schaltflächen kann zwischen den Segmenten hin- und hergeschaltet werden. In der Übersicht erscheinen die bereits transkribierten Segmente grün, die noch nicht transkribierten Segmente wiederum rot. In Abbildung 6 ist die graphische Darstellung der Tonspur im 2D-Editor zu sehen sowie eine Unterteilung in Segmente anhand der Darstellung der Tonspur.

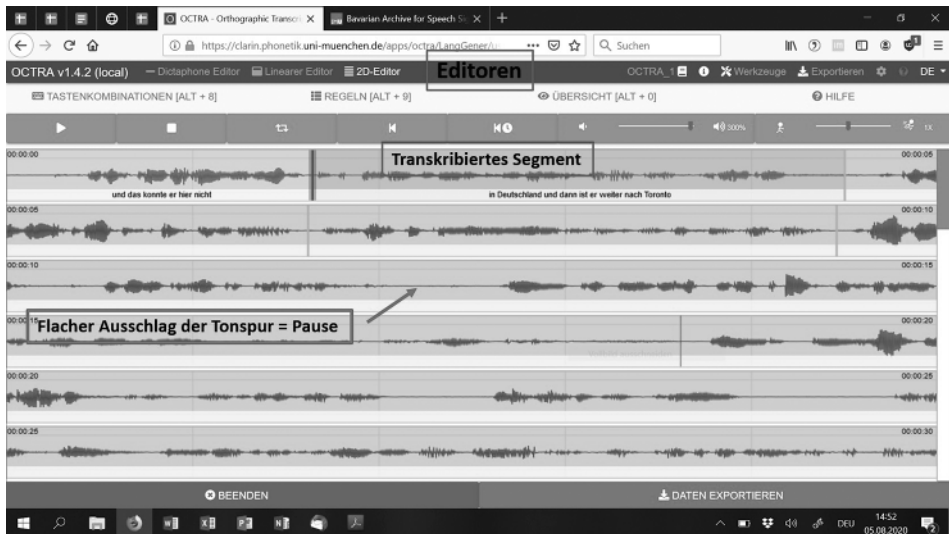


Abbildung 6: 2D-Editor in OCTRA mit bereits transkribierten Segmenten (grün)

Durch die Implementierung von OCTRA als Webapplikation muss der Fortschritt der Transkription nicht manuell gesichert werden. Wenn der Web-Dienst geschlossen wird, kann man durch erneutes Hochladen der Datei auf das bereits Transkribierte zurückgreifen. Erst bei vollständiger Transkription wurden die fertig transkribierten Interviews über die Schaltfläche *Daten exportieren* als TXT-Datei lokal gespeichert und zusätzlich in der Team-Cloud abgelegt.

Für die Zwecke des Projekts wurden die jeweiligen Sonderzeichen, die beispielsweise Hässitationen oder Pausen im Interview kennzeichnen (siehe Bučková & Prawdzic in diesem Band), entsprechend den Transkriptionsregeln angepasst und die Transkriptionsregeln wurden durch das Team des BAS an der LMU in das Tool eingebunden.

Durch das Schneiden der Interviews in kürzere Aufnahmen konnte gewährleistet werden, dass mehrere Transkribentinnen und Transkribenten gleichzeitig an einem Interview arbeiten konnten. Die Unterteilung in kürzere Abschnitte hatte zudem den Vorteil, dass die Transkribentinnen und Transkribenten, die diese Tätigkeit nicht als Hauptbeschäftigung hatten, auch kürzere Aufträge annehmen konnten. Dadurch war jedoch der koordinative Aufwand, also die Verteilung der Aufnahmen zwischen den Transkribentinnen und Transkribenten, arbeitsintensiver.

4.2 WebMAUS

Das nächste Tool, das in der Pipeline Verwendung findet und ebenfalls über BAS aufrufbar ist, ist WebMAUS. WebMAUS steht für die webbasierte Version von MAUS (**M**unich **a**utomatic **s**egmentation) und erstellt eine automatische Segmentierung transkribierter phonetischer Einheiten und deren Zuordnung zu der dazugehörigen Tonspur (Kisler et al. 2017: 331). Vereinfacht gesagt, ordnet WebMAUS den Lauten in der Tonspur die entsprechenden Grapheme in der Transkription zu. Obwohl im Projekt keine Untersuchung auf Phonem-Ebene vorgenommen wird, hilft die automatische Segmentierung, den Prozess der Transkription zu beschleunigen, da sonst eine Segmentierung von Hand vorgenommen werden muss. Dazu muss eine Aufnahme mehrfach abgehört werden, um die Grenzen genau festlegen zu können.

In WebMAUS werden die Segmentgrenzen an Wortgrenzen gesetzt, sodass für die Zwecke des Projekts diese im nächsten Schritt zu Klausen verbunden werden müssen. Dennoch hat die beschriebene Vorgehensweise mit WebMAUS den entscheidenden Vorteil, dass auch Wortgrenzen verfügbar sind. Auf die Segmentierung in Wortgrenzen wurde dann bei der Anonymisierung, bei der nur einzelne Worte und nicht ganze Klausen anonymisiert wurden, zurückgegriffen. Insgesamt konnte im Vergleich zu einer Pipeline, bei der die Segmentierung in ELAN auf Klausen-Ebene manuell vorgenommen wurde, viel Zeit eingespart werden. Dies zu quantifizieren fällt jedoch schwer, da bei den Testenden jeweils unterschiedliche Kenntnisse beider Pipelines berücksichtigt werden müssen, sodass der Unterschied in der Bearbeitungszeit sehr individuell war, jedoch stets zugunsten derjenigen Pipeline ausgefallen ist, in der WebMAUS integriert war.

WebMAUS unterstützt 24 Sprachen und Sprachvarianten, darunter Deutsch und Polnisch (Kisler et al. 2017: 336, 344f.). Tschechisch wird noch nicht unterstützt. Die für die tschechische Sprache zuständige Transkribentin verwendete deshalb WebMAUS nicht, sie bestimmte die Segmentgrenzen manuell. Da die Komplexität der im Hintergrund

laufenden Rechenoperationen, basierend auf dem Markov-Modell, mit der Signallänge exponentiell korreliert, empfiehlt es sich, Dateien mit maximal zehn Minuten Länge hochzuladen (Kisler et al. 2017: 313). Diese Beschränkung wurde im Projekt bereits im Stadium des Schneidens der Aufnahmen berücksichtigt.

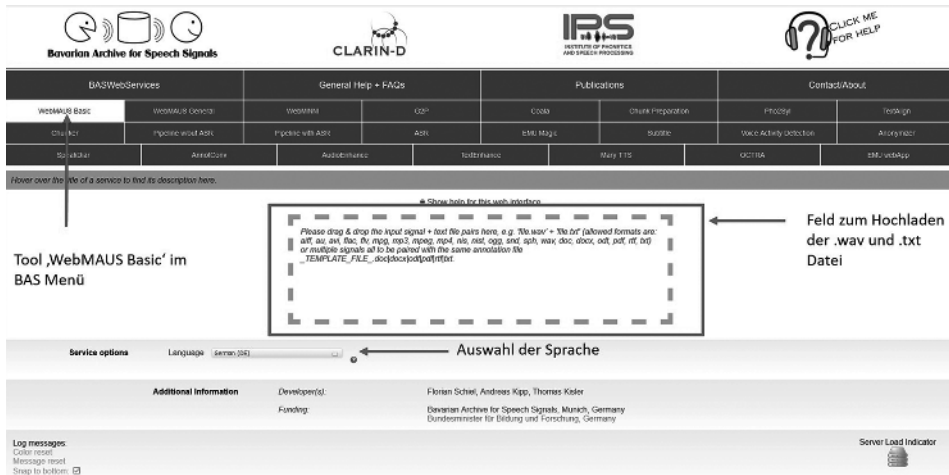


Abbildung 7: WebMAUS Basic Dialogfeld für den Upload

Nachdem das Rohtranskript als TXT-Datei abgespeichert worden war, wurden sowohl die WAV-Datei als auch die TXT-Datei in WebMAUS Basic geladen. In Abbildung 7 ist das Interface von WebMAUS zu sehen, in das sowohl die TXT- als auch die WAV-Datei hochgeladen wird, um eine TextGrid-Datei als Output zu bekommen. Dabei sollte darauf geachtet werden, dass beide Dateien bis auf die Endung gleich benannt werden, weil es sonst zu Problemen mit dem Tool kommt. Auch auf die Auswahl der Sprache muss geachtet werden, um Probleme bei der Alignierung (Verbinden der Tonspur mit dem Transkript) zu vermeiden. Nach erfolgreichem Upload kann das Ergebnis in Form einer TextGrid-Datei abgespeichert werden. Dieses Format dient der weiteren Verarbeitung in EXMARaLDA.

Insgesamt ist das Ergebnis der automatischen Segmentierung und Alignierung durch WebMAUS zufriedenstellend. Dennoch kommt es auch zu Verschiebungen der Tonspur im Vergleich zum Transkript. Diese Abweichungen wurden nicht quantifiziert und hängen stark von der Aufnahme ab, also davon, wie viele Störgeräusche auf der Aufnahme waren oder ob unterschiedliche Personen gleichzeitig gesprochen haben. Derartige Fehler wurden während der Annotation behoben.

4.3 EXMARaLDA

In EXMARaLDA (aus dem Englischen für **Extensible Markup Language for Discourse Annotation**) wurde das Rohtranskript weiterverarbeitet. EXMARaLDA wurde ursprünglich im Teilprojekt *Computergestützte Erfassungs- und Analysemethoden multilingua-*

ler Daten des Sonderforschungsbereichs *Mehrsprachigkeit* (SFB 538) der Universität Hamburg entwickelt. Nun wird es am Hamburger Zentrum für Sprachkorpora in Zusammenarbeit mit dem Archiv für Gesprochenes Deutsch am IDS Mannheim weitergeführt. EXMARaLDA ist „ein System für das computergestützte Arbeiten mit (vor allem) mündlichen Korpora“ (EXMARaLDA online).

Ziel des Programms innerhalb der Pipeline war es, das bisher linear transkribierte Rohtranskript auf die verschiedenen Sprecherinnen und Sprecher und Interviewerinnen und Interviewer aufzuteilen. Dazu mussten unterschiedliche Spuren in EXMARaLDA angelegt werden. Hierfür wurde der Partitur-Editor von EXMARaLDA verwendet, der es ermöglicht, Spuren anzulegen, umzuordnen und zu löschen (Schmidt & Wörner 2005: 174).

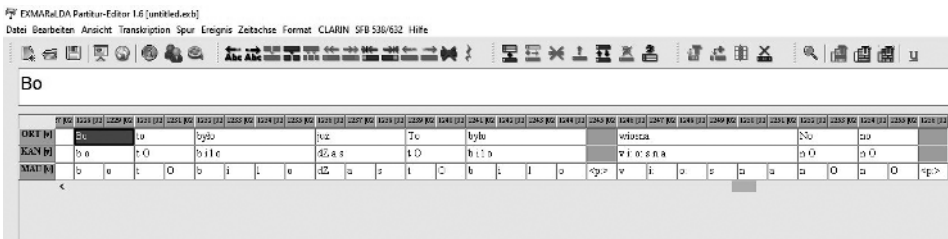


Abbildung 8: In EXMARaLDA importierte TextGrid-Datei aus WebMAUS mit Segmentierung

Bei der Arbeit mit EXMARaLDA wurden im Partitur-Editor zunächst die TextGrid-Datei und die dazugehörige WAV-Datei importiert bzw. hochgeladen. Dabei wird die Segmentierung, die in WebMAUS erstellt wurde, übernommen. Die Ansicht nach dem Öffnen der Datei im Programm ist in Abbildung 8 zu sehen. Die Spuren KAN-MAU (phonetische Transkription auf Wortebene) und MAU (phonetische Transkription der Laute) aus WebMAUS wurden gelöscht, da sie angesichts der Forschungsfragen des LangGener-Projekts nicht relevant waren. Neue Spuren für zwei Interviewerinnen und Interviewer wurden erstellt und Leerzeichen zwischen den Wörtern automatisch hinzugefügt. Im nächsten Schritt konnten die Wörter zu Klausen verbunden und unter den Sprecherinnen und Sprechern aufgeteilt werden. Dabei ist es in der gesprochenen Sprache nicht so einfach, Klausengrenzen festzusetzen. Als Orientierung wurde festgelegt, dass Klausen stets eine finite Verbform enthalten, außer es handelt sich um eine Ellipse (mehr dazu siehe Bučková & Prawdziej in diesem Band). Nach der Aufteilung und Segmentierung wurde die Datei im EAF-Dateiformat abgespeichert. Dieses Format ist das Standardformat des Programms ELAN, in dem die Datei weiterverarbeitet wird.

Grundsätzlich ist auch ELAN dafür geeignet, die eben beschriebenen Schritte durchzuführen. Jedoch hat ELAN eine komplexer aufgebaute Hierarchiestruktur, sodass das Anlegen der neuen Spuren komplizierter ist als in EXMARaLDA. Auch lassen sich die Klausen in EXMARaLDA leichter verbinden, da mehrere Segmente auf einmal markiert und zu einem verbunden werden können. Daher haben wir in dieser Arbeitsphase dieses Programm genutzt.

4.4 ELAN

Für die Annotation linguistischer Daten gibt es einige Programme (ELAN, EXMARaLDA, Praat), die jedoch für unterschiedliche Ziele bezüglich der Annotation geeignet sind. Während sich Praat beispielsweise auf die phonetische Analyse konzentriert, sind für die Analyse gesprochener Sprache sowohl EXMARaLDA als auch ELAN geeignet. Beide haben jedoch Vor- und Nachteile in Hinblick auf die Fragestellung des Projekts.

ELAN ist ein Programm zur komplexen Annotation von Audio- und Videomaterial, das am Max-Planck-Institut für Psycholinguistik in Nijmegen entwickelt wurde (ELAN online). Im Projekt wurde ELAN v. a. deshalb bevorzugt, weil dieses Tool die Möglichkeit bietet, Spuren hierarchisch aufzubauen. Der hierarchische Aufbau erhöht einerseits die Komplexität des Programms gegenüber EXMARaLDA, weswegen in der Pipeline der Zwischenschritt über EXMARaLDA eingebaut wurde. Andererseits erlaubt es diese Hierarchie aber auch, dass Spuren aufeinander Bezug nehmen und so beispielsweise nur diejenigen Segmente in der Spur Phrasentyp (*Sp_Phrasentyp*) annotiert werden können, die auch eine Annotation in der PAT-Spur (*Sp_PAT*) aufweisen.

Die Spuren in ELAN lassen sich in vier verschiedene Gruppen aufteilen:

- Spuren mit Transkriptionen der gesprochenen Sprache (*transcription*⁵):
Basierend auf der Spur mit der orthographischen Transkription (*Sp_Orth_Klausen*) sind die Spuren *Orth_Wörter* und *Dial_Wörter* von dieser orthographischen Spur abhängig. Die Spuren *Orth_Wörter* und *Dial_Wörter* dienen dazu, dialektale Merkmale auf Wortebene zu annotieren. Dies ist jedoch nur für das Deutsche in der „Generation Polen“ relevant. Zu den Spuren mit Transkriptionen gehören ebenfalls die Aussagen der Interviewenden, die in den Spuren *Int1* und *Int2* festgehalten wurden. In der „Generation Polen“ und im deutsch-tschechischen Korpus gibt es zudem eine Spur für die Transkription anderer, an dem Gespräch beteiligter Personen, welche *Dritte Person* genannt wurde.
- Spuren für die sprachformbezogene Annotation (*syntactic*):
Ebenfalls von der Spur *Sp_Orth_Klausen* ist die Spur *Sp_PAT* abhängig, in die Abweichungen von der Heimatland-Baseline annotiert werden (vgl. Bučková, Centner, Księżyk & Prawdziej in diesem Band). Wenn in dieser Spur etwas annotiert wird, ergibt sich die Möglichkeit, in der von der *Sp_PAT* abhängigen Spur das entsprechende Segment auch in der Spur *Sp_Phrasentyp* zu annotieren. Dort wird die Stelle der Abweichung, also der Phrasentyp, annotiert.
- Spuren für die soziolinguistische Annotation (*sociolinguistic*):
Das Besondere am LangGener-Projekt ist, dass neben einer kontaktlinguistischen Annotation auch soziolinguistisch relevante metasprachlichen Daten annotiert werden. Hierfür gibt es in ELAN die Spuren *Entwicklungsphase*, *Domäne* und *Konzeptualisierung der Sprache* (vgl. Jańczak, Nekula & Zielińska in diesem Band).
- Spuren für die technische Umsetzung (*anonymization*):

⁵ In Klammern werden die Namen der Spurentypen angegeben, wie sie in ELAN angelegt wurden.

In der ELAN-Struktur gibt es drei Spuren für die Anonymisierung, die den jeweiligen Personen (der oder die Interviewte, Interviewerin oder Interviewer 1 und Interviewerin oder Interviewer 2) zugeordnet sind und von den jeweiligen Spuren (*Sp_Orth_Klausen*, *Int 1* und *Int 2*) abhängig sind. Zusätzlich dazu ist im deutsch-tschechischen Teilprojekt eine Spur mit Metadaten zum Gesprächsverlauf zu finden.

Abbildung 9 aus ELAN zeigt die eben besprochene Hierarchie im deutsch-polnischen Teilprojekt in Form eines Baumdiagramms, in dem die Abhängigkeiten der einzelnen Spuren durch Linien angezeigt werden:

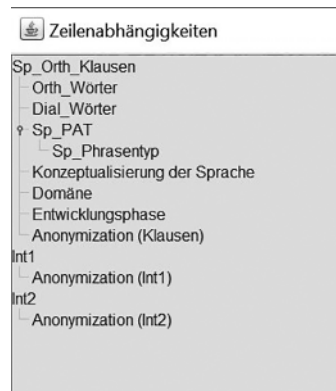


Abbildung 9: Baumdiagramm der Zeilenabhängigkeiten in der ELAN-Datei

Zur Annotation in ELAN wurden die Ergebnisse aus EXMARaLDA in ELAN geöffnet. Um nicht in jeder Datei die komplizierte Spurenhierarchie neu erstellen zu müssen, wurde eine Schablone mit den richtigen Spuren in ELAN in Form einer ETF-Datei importiert und die Transkriptionen der jeweiligen Spuren aus EXMARaLDA in die richtige Spur in ELAN kopiert.

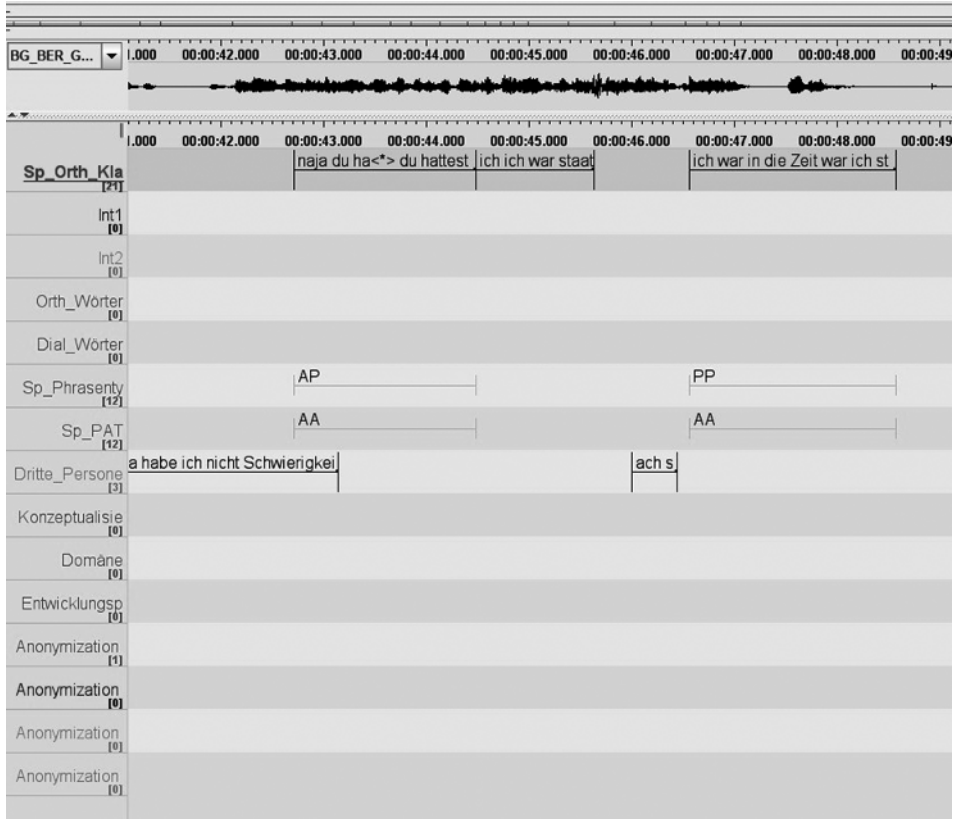


Abbildung 10: Ausschnitt aus einer ELAN-Datei mit den verschiedenen Spuren (links)

Damit ist die Transkription und die technische Vorbereitung auf die Annotation in der technisch beschriebenen Pipeline abgeschlossen, an diese schließt die Annotation an. Die Annotationsregeln und Beispiele für die jeweiligen Kategorien werden in Kapitel 4 (Bučková, Centner, Księżyk & Prawdzc in diesem Band) und in Kapitel 3 (Jańczak, Nekula & Zielińska in diesem Band) erläutert. Die in ELAN enthaltenen Informationen werden in einem XML-Format festgehalten, was ein gängiges Format bei der Speicherung von Korpusdateien ist, sodass die weitere Verarbeitung dieser (siehe Woźniak in diesem Band) erleichtert wird.

4.5 Vor- und Nachteile der Pipeline

Bei der Auswahl der Tools zur technischen Umsetzung im Projekt wurde darauf geachtet, dass die Tools die Arbeit unterstützen und nicht verkomplizieren. So konnte beispielsweise der Aufwand mit der manuellen Anpassung der Klausen an die Tonspur minimiert werden, indem WebMAUS eingesetzt und die automatische Segmentierung durch dieses Tool vorgenommen wurde. Das Feedback der Transkribentinnen und Transkribenten zeigte auch, dass OCTRA die Transkriptionsarbeit durch die funktionale Oberfläche

erleichterte. Während ELAN die Kriterien erfüllt, die in Hinblick auf das Projekt wichtig sind, wie beispielsweise die hierarchische Anordnung der Spuren, ist EXMARaLDA benutzerfreundlicher, sodass EXMARaLDA als Vorstufe zu ELAN in der Pipeline eingebaut wurde. Somit erfüllt jedes Programm seine Aufgabe in der Pipeline und verbessert die Effizienz in der technischen Umsetzung des Projekts auf dem Weg von der Tonaufnahme bis zur annotierten ELAN-Datei.

5 PAT-Kontrolle

Nachdem die Annotation in ELAN erfolgt war, galt es, diese zu überprüfen. Hierfür haben die Teams jeweils untereinander die bereits annotierten Dateien getauscht und die Annotationen verglichen, um eine größere Objektivität zu gewährleisten. Da die technischen Korrekturen, wie beispielsweise die Anpassung der Klausen in ELAN, bereits erfolgt waren, gab es keine Notwendigkeit, diese Kontrolle auch noch in ELAN durchzuführen. Dafür wurde für das Projekt eine schnellere Alternative in Form einer online verfügbaren Tabelle entwickelt, bei der alle Klausen mit den dazugehörigen sprachformbezogenen Annotationen (siehe Bučková, Centner, Księżyk & Prawdzcic in diesem Band) – sobald diese vorliegen – verfügbar sind. Diese Tabellen lassen sich online editieren und die Annotationen verändern.

Table: RK_OPO (German, Edited) Save Discard

Nr	Sp_Orth_Klausen	Sp_PAT	Sp_Phrasentyp	Anonymization	Comments	Segment start	Segment end	Duration
Source: RK_OPO_GP_DE02_pat.esf								
1	ich bin geboren neunzehnhundertneununddreißig hier in Oppeln					0:00:01.437	0:00:07.747	0:00:06.310
2	und <@> die ersten sieben <..> Jahre wurde ich nur Deutsch <@> habe ich nur Deutsch gesprochen					0:00:07.747	0:00:16.577	0:00:08.830
3	denn <..> wir sind von hier <..> aus geflüchtet					0:00:16.577	0:00:22.367	0:00:05.790
4	da war ich fünf Jahre alt					0:00:22.367	0:00:24.307	0:00:01.940
5	ja und <@> das war auch deswegen <@>					0:00:26.337	0:00:31.587	0:00:05.250
6	von meiner Mutter die Brüder waren alle in schon in Russland					0:00:31.587	0:00:37.917	0:00:06.330
7	kommen dann schon zurück auch verwandt					0:00:37.917	0:00:43.027	0:00:05.110
8	und sagten					0:00:43.027	0:00:44.921	0:00:01.894
9	ihr müsst von hier raus					0:00:44.921	0:00:47.097	0:00:02.176
10	denn mit dem Russen ist nicht zu Spaß	AA	VP			0:00:47.097	0:00:51.607	0:00:04.510
11	und so sind wir geflüchtet mit dem Ilus meine tante noch mit dem Sohn also eine Schwester von meiner Mutter die zweite Schwester schon mit vier Kindern und und Mann	PII	NP			0:00:51.607	0:01:08.417	0:00:16.810
12	der Mann hat das erledigt					0:01:08.417	0:01:10.377	0:00:01.960
13	denn er war in Breslau <@> tätig auf dem Flughafen					0:01:10.377	0:01:17.697	0:00:07.320
14	aber wie gesagt <@> gospodarczyz also so <@> Installateur oder so kann man sagen ja	CS	AP			0:01:17.697	0:01:27.447	0:00:09.750
15	und da sind wir gefahren					0:01:27.447	0:01:30.797	0:00:03.350

Abbildung 11: Tabelle zur PAT-Kontrolle mit der Möglichkeit, Annotationen zu ändern

Dazu trägt der oder die PAT-Kontrollierende aus dem jeweils anderen Team seine bzw. ihre Kommentare in die Kommentarspalte ein. Der/die ursprünglich Annotierende, der bzw. die für seine bzw. ihre Datei zuständig ist, überprüft die Kommentare und verändert ggf. die ursprüngliche Annotation. Streitfragen wurden in einer google-sheets-Tabelle zusammengetragen und regelmäßig besprochen. Durch Klicken auf *Save* und das Herunterladen der geänderten Datei konnte so eine neue Datei mit allen Änderungen im ELAN-Format generiert werden.

6 Metadaten

Bereits während der Feldforschung wurden von den Interviewleiterinnen und -leitern Informationen notiert, die sich auf die Aufnahmesituation und die Interviewten bezogen. Diese sind sowohl für die projektinterne Dokumentation und die anschließende soziolinguistische Analyse als auch für die späteren externen Korpusnutzerinnen und -nutzer bedeutsam. Im Bearbeitungsprozess haben sich aus den Transkripten weitere Informationen über die Interviewten und die berichteten Inhalte ergeben. Folglich wurden die Metadaten, die jeder Korpusdatei zugeordnet wurden, in vier Kategorien aufgeteilt: der oder die Interviewte, Interviewerin oder Interviewer, Interview und Dateien.

Die Kategorie *Speakers* hält personenbezogene Informationen fest, und zwar das Personenkürzel, den Namen⁶, die Zugehörigkeit zu einer der untersuchten Generationen oder Gruppen, das Geschlecht, das Geburtsjahr und das Jahr der Emigration, die Schulbildung, den Beruf, die Lokalisierung der Interviewten (Herkunfts- und Wohnregion), eine kurze Charakteristik der benutzten Sprachvarietäten und weitere Kommentare.

Für die Informationen zur Schulbildung wurden anhand der Skala der Europäischen Kommission (NaN 2020) sechs Angaben herausgearbeitet, für die jeweils die entsprechenden Übersetzungen nach dem deutschen, polnischen und tschechischen Bildungssystem gefunden wurden. Sie werden in Tabelle 1 vorgestellt. In den Metadaten werden sie durch die Nummern 0 bis 5 kodiert.

1	some classes of school	nedokončené základní vzdělání	einige Klassen der Grundschule	kilka klas szkoły podstawowej
2	primary education	nedokončené základní vzdělání	Grundschule	szkoła podstawowa
3	lower secondary education	základní vzdělání	Haupt-/Realschule	szkoła zawodowa
4	upper secondary education	středoškolské vzdělání, středoškolské vzdělání s maturitou	Gymnasium/ Berufsbildende Schule	technikum/ liceum
5	tertiary education	vysokoškolské vzdělání	Universität, Hochschule, weiterbildende Schule etc.	szkoła policealna, uniwersytet etc.
0	no information	žádný údaj	keine Angabe	brak danych

Tabelle 1: Angaben zur Schulbildung der Interviewten

⁶ Namen dienen der teaminternen Dokumentation und werden nicht veröffentlicht. Die Anonymisierung und der Datenschutz werden im Abschnitt 4.4 und in Kapitel 8 (Woźniak in diesem Band) besprochen.

Bei den Lokalisierungsangaben wurden Geburts- und Wohnregion vermerkt. Diese Angaben dienen dem Filtern bei der Korpusuche. Deshalb wird mit größeren Regionen gearbeitet, damit die durch das Filtern entstandenen Gruppen von Interviewten nicht zu klein ausfallen. Als Geburts- bzw. Wohnregion wurde in Deutschland das Bundesland, in Tschechien der Landesbezirk (kraj) und in Polen die Woiwodschaft angegeben. Bei der „Generation Polen“ (vgl. Centner in diesem Band) wurde bei der Geburtsregion der historische Name eingetragen, der in dem Geburtsjahr der oder des Interviewten die betreffende Region bezeichnete. Bei der Wohnregion wurde hingegen der aktuelle Name verwendet.

Der Geburts- und der Wohnort wurden jedoch noch genauer spezifiziert. Diese Information kann von großer Bedeutung sein, wenn es beispielsweise um die Untersuchung des Dialektes geht. Für Polen wurden daher auch die Woiwodschaft, der Kreis und ggf. die Stadt angegeben. Bei der „Generation Polen“ kommt im Falle des Geburtsortes zu dem aktuellen Namen auch der historische Name hinzu. Bei der „Generation Deutschland“ wurde das Bundesland, der Bezirk oder der Landkreis, die Stadt und beim Geburtsort ebenfalls das Dorf angegeben. Für Tschechien wurde der Bezirk (okres) eingetragen. Die entsprechenden Tabellen befinden sich am Ende dieses Bandes (Anhang 4).

Die Beschreibung der benutzten Sprachvarietäten ist eine allgemeine Charakteristik beider untersuchten Sprachen in Bezug auf die Fragestellung des Projektes. Bei der Interviewten LS_ZEI wurde im deutsch-tschechischen Teilprojekt beispielsweise Folgendes notiert: „CZ: Viele Abweichungen in der Grammatik und in der Lexik, phonetische Interferenzen aus dem Deutschen; DE: Dialektal (Bairisch) geprägte Umgangssprache, möglicherweise auch böhmisch-deutsche Merkmale.“ Bei der Interviewten HD_SWI aus der „Generation Polen“ wird das Deutsche so beschrieben: „Die Sprecherin realisiert zahlreiche Dialektmerkmale des Deutschen: die Palatalisierung von [g] zu [j] im Anlaut, die Spirantisierung von [g] zu [x] nach hinteren Vokalen, unverschobenes [t] oder [p], [k] als Folge der im Niederdeutschen unterbliebenen zweiten Lautverschiebung und Kontraktionen der Flexionsendungen an Verben mit auslautenden Verschlusslauten. Im morphologischen Bereich wird die Form *mir* als eine Abweichung von der deutschen Standardsprache verwendet.“

Das Kommentarfeld in der *Speakers*-Kategorie dient den frei formulierten Angaben der Interviewerinnen und Interviewer. Bei Bedarf wird hier die oder der Interviewte im Hinblick auf die Sprachbiographie (zum Beispiel Einstellungen zur Sprache, Eltern, Großeltern und deren Sprachgebrauch) kurz charakterisiert.

Die Kategorie *Interviewers* enthält das Kürzel der Interviewerinnen und Interviewer (zusammengesetzt aus Vor- und Nachnamen, zum Beispiel AB), den Namen, die E-Mail-Adresse und die Team-Zugehörigkeit (zum Beispiel CZ für das Team Tschechien).

In der Kategorie *Interviews* werden relevante Daten bezüglich der Aufnahmesituation angegeben. Jedes Interview erhält sein eigenes Kürzel, das der Bezeichnung entspricht, die der Ordner mit dem Interview im gemeinsamen Speicherplatz trägt. Des Weiteren werden die Generation oder Gruppe, das für das Interview zuständige Team und das Kürzel der Interviewten sowie der Interviewerinnen und Interviewer jeweils separat notiert. Es werden darin die Anzahl der Sitzungen, in denen das Interview aufgenommen wurde, das Aufnahmedatum, die Länge des gesamten Interviews, die Anzahl der daraus entstandenen Dateien, die im Interview verwendeten Sprachen, der Ort und das Land, wo das Interview aufgenommen wurde, sowie die Aufnahmebedingungen und die Technik erfasst.

Die Aufnahmebedingungen beziehen sich auf die Umgebung, in der das Interview aufgenommen wurde. Zur Auswahl stehen die Kategorien zu Hause, öffentlicher Raum (zum Beispiel im Café, Restaurant, Büro der deutschen Vereine o. ä.), Studio oder Büro. Die Umgebung hat Konsequenzen für die Aufnahme. Das narrative Interview wird typischerweise bei den Interviewten zu Hause durchgeführt, was zu einer vertraulichen, entspannten Stimmung beiträgt (siehe Jorroch & Prawdzcic in diesem Band). Aus Sicht der technischen Qualität bedeutet dies im Regelfall eine ruhige Aufnahme mit wenigen Störgeräuschen. Aus unterschiedlichen Gründen ist ein Hausbesuch bei den Interviewten jedoch nicht immer möglich gewesen, weswegen die Interviews auch an anderen Orten durchgeführt wurden. In solchen Aufnahmen kommt es häufiger zu Störgeräuschen oder Zwischenfragen durch dritte Personen. Diese werden zwar in den meisten Fällen herausgeschnitten (siehe Abschnitt 3), vereinzelt können sie aber dennoch in den Transkripten auftreten.

In der Spalte Technik werden Informationen über das verwendete Aufnahmegerät dokumentiert. Ähnlich wie die Aufnahmebedingungen deutet diese Angabe auf die technische Qualität der Aufnahme hin. Dies kann in Zukunft vor allem für Fragestellungen aus dem Bereich der Phonetik eine Rolle spielen (siehe Abschnitt 1).

Die letzte Angabe in der Kategorie Interviews ist Kommentar. Wie bei der Kategorie *Speakers* ist dieses Feld fakultativ und kann von den Interviewerinnen und Interviewern ausgefüllt werden. Der Kommentar informiert über den Interviewkontext, zum Beispiel darüber, ob das Gespräch oft unterbrochen wurde oder ob es technische Probleme gab.

Die letzte in den Metadaten enthaltene Kategorie ist *Files*. Hier werden alle zu einem Interview gehörenden Dateien aufgelistet – siehe Abbildung 12. Die Kategorie enthält die Namen der einzelnen Dateien, das Interview-Kürzel, die Nummer der Datei im Interview, die Sprache, die in der Datei überwiegt, das Kürzel der involvierten Sprecherinnen und Sprecher, die Dauer der zu der Datei zugehörigen Aufnahme, den Anfang und das Ende auf der Interviewzeitachse und die Information, ob der Interviewabschnitt ein Teil des Korpus ist (true) oder ob er ausgeschnitten wurde (false). Die letzte Angabe beinhaltet die Labels, die auf den Inhalt des Interviews in den einzelnen Dateien verweisen. Es handelt sich um dieselben Labels, die beim Schneiden der Aufnahmen eingetragen wurden (siehe Abschnitt 3).

Filename	Interview ID	Number	Session Number	Language	Speaker(s)	Duration hh:min:sec.ms	Interview Time Start	Interview Time End	In Corpus	Semantic Labels
LS_ZEI_SP_DE_0001	LS_ZEI	1		1 DE	LS_ZEI	00:10:01.191	0	00:10:01.190	true	Sprache; Familie
LS_ZEI_SP_DE_0002	LS_ZEI	2		1 DE	LS_ZEI	00:13:03.284	00:10:01.191	00:23:04.474	true	Sprache; Schule
LS_ZEI_SP_CZ_0003	LS_ZEI	3		1 DE	LS_ZEI	00:00:09.619	00:23:04.475	00:23:14.093	false	Übergangversuche ins Tschechische
LS_ZEI_SP_CZ_0004	LS_ZEI	4		1 CZ	LS_ZEI	00:09:53.637	00:23:14.094	00:33:07.730	true	Arbeit; Sprache; Bekannte
LS_ZEI_SP_CZ_0005	LS_ZEI	5		1 CZ	LS_ZEI	00:07:19.290	00:33:07.731	00:40:27.020	true	Familie; Vereine; Sprache
LS_ZEI_SP_CZ_0006	LS_ZEI	6		1 CZ	LS_ZEI	00:00:14.268	00:40:27.021	00:40:41.288	false	Störgeräusche
LS_ZEI_SP_CZ_0007	LS_ZEI	7		1 CZ	LS_ZEI	00:07:24.904	00:40:41.289	00:48:06.192	true	Sprache; Kontakte zu Heimat
LS_ZEI_SP_CZ_0008	LS_ZEI	8		1 CZ	LS_ZEI	00:13:03.174	00:48:06.193	01:01:09.366	true	Sprache; Familie; Kontakte zu Heimat
LS_ZEI_SP_CZ_0009	LS_ZEI	9		1 CZ	LS_ZEI	00:00:29.228	01:01:09.367	01:01:38.594	false	andere Themen

Abbildung 12: Metadaten – Kategorie Files

Literatur

Bučková, Aneta (in Vorb.): *Syntaktische Musterentlehnungen bei deutsch-tschechischen Bilingualen*. Inauguraldissertation zur Erlangung des Doktorgrades, Universität Regensburg.

- Draxler, Christoph/Schiel, Florian (2018): Moderne phonetische Datenbanken. In: Marc Kupietz, Thomas Schmidt (Hgg.), *Korpuslinguistik*. Berlin, Boston: De Gruyter, 179–208.
- Kisler, Thomas/Reichel, Uwe/Schiel, Florian (2017): Multilingual processing of speech via web services. *Computer Speech & Language*, 326–347.
- NaN (2020): National Education Systems. Online verfügbar unter: https://eacea.ec.europa.eu/national-policies/eurydice/national-description_en, letzter Zugriff am 23.12.2020.
- Schmidt, Thomas (2018). Gesprächskorpora. Aktuelle Herausforderungen für einen besonderen Korpusstyp. In: Marc Kupietz, Thomas Schmidt (Hgg.), *Korpuslinguistik*. Berlin: de Gruyter, 209–230.
- Schmidt, Thomas/Wörner, Kai (2005): Erstellen und Analysieren von Gesprächskorpora mit EXMARALDA. *Gesprächsforschung* 6, 171–195.

Tools

- Audacity ® (2020): Home. Online verfügbar unter: <https://www.audacityteam.org/>, letzter Zugriff am 11.08.2020.
- ELAN (Version 6.0) [Computer software]. (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Online verfügbar unter: <https://archive.mpi.nl/tla/elan>, letzter Zugriff am 28.04.2021.
- EXMARALDA. Online verfügbar unter: <https://exmaralda.org/de/>, letzter Zugriff am 28.04.2021.
- Jänsch, Klaus/Draxler, Christoph (2020): BAS – SpeechRecorder Software. Online verfügbar unter: <https://www.bas.uni-muenchen.de/Bas/software/speechrecorder/>, letzter Zugriff am 25.09.2020.
- OCTRA – modern web-application for the orthographic transcription (2020). Online verfügbar unter: <https://clarin.phonetik.uni-muenchen.de/apps/octra/octra/login>, letzter Zugriff am 13.08.2020.
- WebMAUS basic (2021): Online verfügbar unter: <https://clarin.phonetik.uni-muenchen.de/BAS-WebServices/interface/WebMAUSBasic>, letzter Zugriff am 28.04.2021.

Von der annotierten Datei zum Korpus

Michał Woźniak

Abstract:

The data processing: from the annotated data to the corpus

This chapter is dedicated to the technical side of the construction and use of the LangGener corpus. The first section deals with the preparation of the data, the automated processing and the building of an annotated multilingual corpus. We present the advantages of Git, a software for tracking changes in sets of files we used for the management of the transcription process and we present a special tool developed for the manual correction of the transcription and annotation. We discuss the automated part of the working process like data extraction, annotation of language contact results and data formats. Further, we present the corpus management system CWB and the data encoding pipeline. The second section deals with the web interface SpoCo used as basis for the interface of the LangGener corpus. The architecture of the interface, its main components are presented and the corpus query language CQL is explained. The final section shows how the interface can be used and how corpus queries can be formulated.

Hinführung

In diesem Beitrag geht es um die technische Verarbeitung einer annotierten Datei bis hin zu einer Korpusoberfläche. Dabei wird auf Git, ein Versionsverwaltungssystem, das im Projekt verwendet wurde, eingegangen, sowie ein eigens erstelltes Tool vorgestellt, das die Korrektur der sprachformbezogenen Annotation erleichtert. Anschließend werden die Schritte, die zur Erstellung einer Corpus Workbench (CWB) nötig sind, erläutert. Es wird das Korpus, das dabei entsteht, vorgestellt sowie Wege aufgezeigt, wie dieses zu benutzen ist. Die Schritte von der Aufnahme eines Gesprächs bis zur annotierten ELAN-Datei werden im Kapitel (Bučková & Centner in diesem Band) beschrieben.

1 Vom Text zum Korpus – technische Aspekte

1.1 Versionsverwaltungssystem: Git

Im LangGener-Projekt sind an der Transkription und Annotation zahlreiche Personen beteiligt, die mehr als 2 000 Dateien bearbeiten, von denen jede innerhalb der einzelnen Phasen mehrere Varianten aufweisen kann. Somit ist eine definierte Methode für die

Kontrolle und Koordination dieses Vorhabens unerlässlich. Aus diesem Grund haben wir uns für die Nutzung der Versionsverwaltungs-Software Git entschieden.

Versionsverwaltungssysteme ermöglichen die Verfolgung von Ordnerinhalten mit den zugehörigen Daten und jeglichen Änderungen, die daran vorgenommen werden. Obwohl diese Systeme im Hinblick auf die Anforderungen von Programmierern und Programmierern entwickelt wurden, die Änderungen an Codes (die sich ständig ändern und weiterentwickelt werden) überwachen möchten, spricht nichts dagegen, sie auch für die Arbeit mit Textdaten für wissenschaftliche Zwecke zu nutzen. Derzeit werden viele Systeme dieser Art angeboten. Aufgrund der Verbreitung und Stabilität der Anwendung fiel unsere Wahl auf Git.

Git verwaltet Dateien in einem entsprechend vorbereiteten Ordner (Repository). Jede Änderung in einer Datei des Repository wird sofort vermerkt. Die Änderungen können überprüft und bestätigt (bzw. abgelehnt) werden. Nach der Synchronisation können alle Benutzerinnen und Benutzer des betreffenden Repository auf die neue Version der Datei (bzw. Dateien) zugreifen. Dieses System wartet mit zahlreichen Vorteilen auf:

- Jede Datei existiert, unabhängig von der Phase, nur in einer einzigen, aktuellen Version, sodass nicht die Gefahr besteht, dass an einer veralteten Dateiversion gearbeitet wird.
- Der Zugriff auf den vollständigen Änderungsverlauf einer bestimmten Datei ist möglich.
- Einzelne Versionen lassen sich problemlos miteinander vergleichen.
- Es wird automatisch eine Sicherungskopie erstellt. Darüber hinaus ist ein Datenverlust sehr unwahrscheinlich (jede Benutzerin und jeder Benutzer verfügt über eine lokale Kopie des Repository, während eine zusätzliche Kopie auf einem für die Synchronisation zuständigen Remote-Server abgelegt ist).

Für die Nutzung von Git ist eine gewisse Kenntnis seiner Struktur und der Befehle, die per Texteingabe über die Benutzeroberfläche erteilt werden, erforderlich. Für Benutzerinnen und Benutzer ohne Erfahrung im Bereich der Informatik kann sich dies als wesentliche Hürde erweisen. Aus diesem Grund kommt im Rahmen des Projekts der sogenannte Git-Client zur Anwendung – ein Programm mit grafischer Benutzeroberfläche, genannt Fork (Pristupov 2022), das die Arbeit mit Git auf wenige einfache Schritte reduziert. Wie die Benutzeroberfläche von Fork aussieht, ist der folgenden Abbildung zu entnehmen.

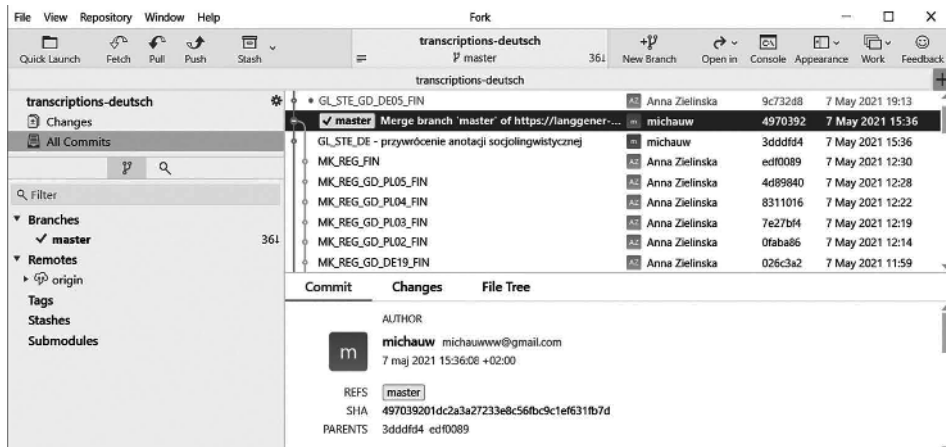


Abbildung 1: Fork Interface

Trotz der Vielzahl von Vorteilen und Funktionen der Versionskontrollprogramme können mitunter Probleme auftreten, zum Beispiel durch Nichtbeachtung der Synchronisationsverfahren oder Konflikte, die darauf zurückzuführen sind, dass mehrere Personen lokal an derselben Datei arbeiten. Daher wurde eine detaillierte Anleitung mit einer Beschreibung der Architektur von Git und der notwendigen Schritte erstellt. Darüber hinaus wird die Verteilung der zu bearbeitenden Dateien von benannten Personen überwacht und in einer speziellen Tabelle für diesen Zweck schriftlich festgehalten.

1.2 Tools für die Korrektur der syntaktischen Annotation

Obwohl ELAN ein nützliches Werkzeug für die ebenenbasierte Annotation von Daten ist, wie in dem Artikel von Bučková & Centner in diesem Band gezeigt wurde, ist seine Verwendung für die Annotation von abhängigen Ebenen (zum Beispiel von syntaktischen Ebenen) nicht besonders komfortabel. Ferner gestaltet sich die Annotation bzw. ihre Kontrolle/Korrektur durch mehr als eine Person zum selben Zeitpunkt äußerst schwierig. Beim LangGener-Projekt wird dieses Problem dadurch verstärkt, dass die Aufzeichnung in mehrere Komponentendateien fragmentiert ist, was die Durchsicht des vollständigen Interviews unmöglich macht. Angesichts dieser Herausforderungen wurde ein spezielles System entwickelt, um die syntaktischen Ebenen vollständiger Interviews effizient zu annotieren. Hierbei handelt es sich um eine Browser-Anwendung (geschrieben in Django, einem Python-basierten Web-Framework), die online verfügbar ist und mehreren Benutzerinnen und Benutzern gleichzeitig die Möglichkeit bietet, den Inhalt anzuzeigen und zu ändern.

Das System führt sämtliche Dateien eines Interviews zusammen (unter Beibehaltung der Unterteilung in Sprachversionen) und stellt sie in Form einer Tabelle dar, deren einzelne Zeilen den Segmenten der Ebene *Sp_Orth_Klausen* entsprechen. In den Spalten werden die Segmentnummer, der Segmentinhalt, die PAT-Annotation und der Phrasentyp sowie Felder für Anmerkungen und Anonymisierung angezeigt. Die Felder können

bearbeitet und Änderungen in einer EAF-Datei gespeichert werden (ELAN-spezifisches Format). Anschließend wird diese Datei in das Git-Repository kopiert und ersetzt dort die Vorgängerversion (wie bei Bučková & Centner in diesem Band dargestellt).

1.3 Datenextraktion, automatisches Tagging

Nachdem die mithilfe von Tabellen überprüften Dateien im EAF-Format abgespeichert und im Git-Repository abgelegt worden sind, ist damit der manuelle Annotationsprozess abgeschlossen und von hier an beginnt der eigentliche Korpusaufbau. Die nun fertigen Transkriptionen im ELAN-Format enthalten Daten, die zunächst extrahiert und anschließend in ein Format übertragen werden müssen, das zur Erstellung einer Korpusbasis verwendet werden kann. Bei den von ELAN erzeugten EAF-Dateien handelt es sich im Grunde um XML-Dateien mit einer definierten Struktur. Die Korpusdatenbank im CWB-Standard (nachfolgend beschrieben) erfordert Quelldateien im VRT-Format (Verticalized Text). Hierbei handelt es sich um ein Format, das für die Speicherung von Daten mit linearer Struktur und unterschiedlichen Annotationsebenen optimiert ist und sich somit gut für die Darstellung von Korpusdaten eignet. Zudem zeichnet sich dieses Format durch eine bessere Lesbarkeit aus und belegt weniger Speicherplatz als XML. Bei VRT-Dateien wird jeder Begriff in eine eigene Zeile geschrieben, und eine Zeile wiederum kann in Spalten unterteilt werden. Jede Spalte entspricht einer Annotationsebene, zum Beispiel Wortform, Lemma, grammatischer Marker, emotionale Färbung eines Wortes, Dialektmerkmale usw. (Die Bedeutung und Reihenfolge der einzelnen Spalten können beliebig festgelegt werden.) In der CWB-Terminologie entspricht jede Spalte einem sogenannten Positionsattribut. Zum zweiten Attributtyp gehören Strukturattribute, die durch XML-Tags¹ dargestellt werden. Diese ermöglichen die Unterscheidung linearer Strukturen innerhalb des Texts, zum Beispiel Sätze, Abschnitte bzw. Segmente im mündlichen Korpus. Auf der folgenden Abbildung wird beispielhaft die Struktur einer Musterdatei im VRT-Format dargestellt.

¹ Es ist zu beachten, dass das VRT-Format, obwohl es XML-Tags verwendet, nicht mit dem XML-Format kompatibel ist. Zum Beispiel können sich die durch die unterschiedlichen Tag-Typen abgegrenzten Bereiche überschneiden, die Struktur der Tags ist nicht hierarchisch.

```

<utterance id="2" spkr="AJ_AUG" from="3877" to="11517" domain="M" develop-
ment="AA" lang="de">
Ja                ADV ja
Ja                ADV ja
Natürlich        ADV natürlich
Ja                ADV ja
Natürlich        ADV natürlich
sym3             VVFIN <unknown>
wir              PPER wir
schauen         VVFIN schauen
jeden           PIAT jede
Tag             NN Tag
Die            ART die
polnische ADJA  polnisch
Nachrichten    NN Nachricht
</utterance>

```

Abbildung 2: Struktur einer VRT-Datei

Die Konvertierung der Dateien vom EAF-Format in das VRT-Format erfolgt automatisch mithilfe eines Python-Skripts. Dieses Skript ist zudem für das automatische Tagging von Daten zuständig. Für jede Sprache wird ein entsprechender Tagger ausgeführt – ein Tool, mit dem Wörter automatisch lemmatisiert und diesen grammatische Eigenschaften zugeordnet werden. Derzeit wird zum Taggen des deutschen Texts TreeTagger (Schmid 1995) verwendet, während für den polnischen Text TaKIPI (Piasecki 2007) zum Einsatz kommt.

Wie in Abbildung 2 ersichtlich, umfasst das Korpus zwei Basiseinheiten: Wort und Segment (in der VRT-Datei als *utterance* gekennzeichnet). Hier zeichnet sich ein grundlegendes Merkmal der Korpusarchitektur ab, das sich auch auf die Vorgehensweise beim Durchsuchen des Korpus auswirkt. Ein einzelnes Ergebnis der Benutzerabfrage bildet ein Segment, das eine Übereinstimmung mit der formulierten Abfrage aufweist. Die Abfrage jedoch wird auf der Grundlage eines Wortes oder einer Wortfolge konstruiert (siehe Abschnitt 2). Wie bereits oben erwähnt, wurde für das Management der Korpusdaten im LangGener-Projekt der CWB-Standard ausgewählt, dessen genauere Beschreibung im folgenden Abschnitt vorgenommen wird.

1.4 CWB-Korpus

Für die effektive Speicherung von Korpusdaten und die Abfrage des Korpus sind spezielle Tools unerlässlich. Unter den verfügbaren Lösungen fiel die Wahl auf den bereits angeführten CWB-Standard (IMS Open Corpus Workbench, siehe Evert & Hardie 2011). Hierbei handelt es sich um einen seit vielen Jahren entwickelten Standard, der die gewünschte Stabilität und Geschwindigkeit beim Durchsuchen großer Korpora bietet. In der Korpuslinguistik ist er weit verbreitet. Die von den vom CWB bereitgestellten Tools gebildete Datenbank stellt das Backend des Korpus dar, das die Browser-Schnittstelle nutzt. In diesem Zusammenhang ist anzumerken, dass dieses Modul ausschließlich die Textebene des Korpus betrifft. Damit die Benutzerinnen und Benutzer

auch bei der endgültigen Korpusversion Zugriff auf die Aufnahmen haben, ist ein separater Mechanismus (siehe Abschnitt 2) notwendig.

2 Schnittstelle

Das Korpus des LangGener-Projekts ist ein Korpus der gesprochenen Sprache, das für wissenschaftliche Zwecke gedacht ist und künftig hauptsächlich von Forscherinnen und Forschern genutzt werden wird. Aus diesen Gründen sollte die Benutzeroberfläche des Korpus bestimmte Voraussetzungen erfüllen:

- Möglichkeit zur gleichzeitigen Darstellung von Text- und Tonebene
- Struktur, die Benutzerfreundlichkeit mit elaborierten Suchfunktionen kombiniert (siehe Abschnitt 2.2)
- Stabilität beim Betrieb
- einfacher Zugang

Diese Eigenschaften treffen auf die Schnittstelle SpoCo (Waldenfels & Woźniak 2017) zu, die als Basis für die endgültige Form der Schnittstelle dient, die sich zum Entstehungszeitpunkt dieses Texts noch in Vorbereitung befindet.

Die SpoCo-Schnittstelle (die Abkürzung steht für **Spoken Corpus** und entspricht außerdem dem polnischen Begriff *spoko*, der in der Umgangssprache für „in Ordnung“ bzw. „ist gut“ verwendet wird) wurde im Hinblick auf die Anforderungen von Korpora gesprochener Sprache entwickelt und kommt (in unterschiedlichen Entwicklungsphasen) für folgende verwandten Projekte zur Anwendung: Korpus Gwary Spiskiej (Korpus des Zipser Dialekts (Grochola-Szczepanek et al. 2019)), Korpus Języka Rusińskiego (Korpus der russinischen Sprache (Rabus 2015)), Korpus Ustya (Ustya-Korpus (Daniel et al. 2014)). Sie zeichnet sich durch folgende Hauptmerkmale aus:

1. Browser-Anwendung: Die browserbasierte Schnittstelle SpoCo ermöglicht die Erstellung von Korpora über das Internet.
2. In AngularJS geschrieben: AngularJS basiert auf moderner Technologie, welche die Erstellung dynamischer Internetseiten ermöglicht, die bei Änderungen nicht erneut aktualisiert werden müssen.
3. Flexibel: Ermöglicht die Anpassung der Oberfläche im Hinblick auf die spezifischen Anforderungen eines Korpus (zum Beispiel zusätzliche Annotationsebenen, dialektale Form eines Wortes, syntaktische Beziehungen usw.).
4. Modular: Im Hinblick auf das relativ einfache Hinzufügen neuer Funktionen bzw. Ändern vorhandener Funktionen entwickelt (zum Beispiel zusätzliche statistische Funktionen).

Wie genau das Zusammenspiel zwischen der ausgewählten SpoCo-Schnittstelle und den projektspezifischen textuellen und auditiven Korpusdaten funktioniert und auf welche Art und Weise es zu einem Austausch von Informationen auf allen diesen Ebenen kommt, wird ausführlicher im folgenden Abschnitt behandelt.

2.1 Architektur der Schnittstelle

Der Korpusaufbau in SpoCo beruht auf drei zusammenhängenden Komponenten: Oberfläche, Ordner mit dem Korpus und Ordner mit den Aufnahmen. Die erste Komponente umfasst sämtliche Elemente, die für die Bereitstellung des Korpus in Form einer Browser-Anwendung notwendig sind, während die zweite Komponente die erforderlichen Textdaten und die dritte die Audiodaten bereitstellen (Abbildung 3).

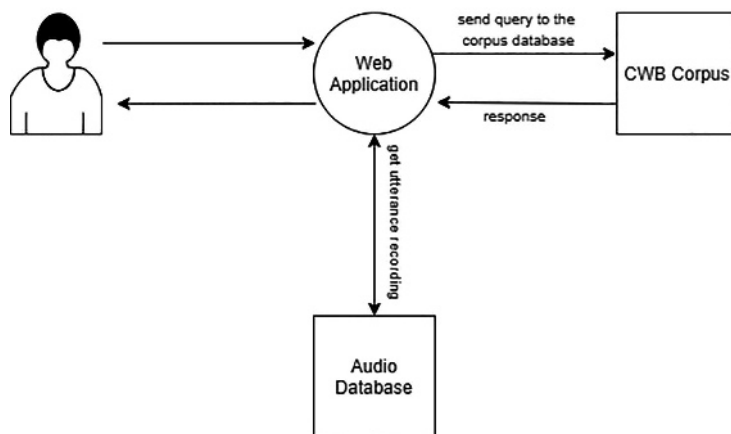


Abbildung 3: Die Architektur der SpoCo-Schnittstelle

2.1.1 Korpus

SpoCo ist derzeit für die Arbeit mit Korpora im CWB-Format ausgelegt. Ein entsprechend vorbereitetes Korpus wird an die Schnittstelle angebunden. Zu diesem Zweck werden PHP-Skripte verwendet, die über die Benutzerabfragen an die Suchmaschine CQP (Corpus Query Processor) übermittelt werden. Anschließend werden die Ergebnisse abgerufen und an die Schnittstelle übertragen.

Bei CWB (IMS Open Corpus Workbench, Evert & Hardie 2011) handelt es sich um einen Satz von Tools zum Erstellen und Durchsuchen großer Korpora. Dieser seit vielen Jahren entwickelte Standard bietet die gewünschte Stabilität und Geschwindigkeit beim Durchsuchen von Daten und ist in der Korpuslinguistik weit verbreitet.

2.1.2 Datenbank für Aufnahmen

Das CWB-Korpus enthält ausschließlich Textdaten (zusammen mit Annotationen), ist jedoch nicht für die Speicherung von Audiodaten ausgelegt. Dieser Bereich des Korpus stellt eine separate Komponente des Gesamtsystems dar. Aus technischer Sicht handelt es sich hierbei um einen separaten Ordner, der Aufnahmefragmente beinhaltet (in Form von WAV-/MP3-Dateien), von denen jedes einem Transkriptionssegment entspricht.

2.1.3 Browser-Anwendung

Die Anwendung spielt eine entscheidende Rolle bei der Vermittlung zwischen der Benutzerin oder dem Benutzer und dem eigentlichen Korpus (und der Aufnahmedatenbank) und verknüpft alle Elemente miteinander. Wie bereits oben angemerkt, wurde das Korpus in AngularJS geschrieben: Diese Technologie basiert auf der Programmiersprache JavaScript und wurde unter der Leitung von Google entwickelt. Sie ermöglicht die Erstellung sogenannter Single-Page-Anwendungen, d. h. von Internetseiten, die mit der Benutzerin oder dem Benutzer und dem Server kommunizieren und sich dynamisch an die Situation anpassen, ohne dass eine Aktualisierung der Seite erforderlich ist. Dies wirkt sich vorteilhaft auf die Reaktionsschnelligkeit, die Einfachheit der Erstellung erweiterter Seitenelemente und nicht zuletzt auf die Benutzerfreundlichkeit aus.

2.2 Korpusuche

Nachdem die grundlegenden Elemente des Korpusaufbaus präsentiert wurden, ist es auch wichtig, auf die Gestaltung der Interaktion mit den künftigen Korpusbenutzerinnen und -benutzern einzugehen, die sich mittels einschlägiger Eingabe linguistischer Abfragen gestalten wird. An dieser Stelle wird also ein kleiner Einblick in die Struktur der CQL-Abfragesprache gegeben, die dem Prozess der Korpusuche zugrunde liegt.

Das LangGener-Korpus setzt sich aus zwei grundlegenden Einheiten zusammen: Wörtern und Segmenten (die in ELAN in der Transkriptionsphase unterschieden werden, siehe dazu Bučkova, Centner, Księżyk & Prawdzcic sowie Bučkova & Centner in diesem Band). Die Abfrage wird auf der Grundlage einzelner Wörter oder anhand von Wortsequenzen gebildet. Als Ergebnis einer Abfrage wird eine Liste sämtlicher Stellen im Korpus ausgegeben, die der Abfrage entsprechen (diese Liste kann in bestimmten Fällen leer sein oder aber das gesamte Korpus umfassen). Dabei wird jedes Ergebnis im Kontext eines vollständigen Segments angezeigt, in dem der Suchbegriff (Wort) enthalten ist. Bei Bedarf kann der Kontext um mehrere benachbarte Segmente erweitert werden.

Die Suche im Korpus erfolgt auf der Grundlage der ebenfalls weit verbreiteten Abfragesprache CQL (Corpus Query Language). Einen gewissen Nachteil für Einsteiger stellt die Tatsache dar, dass für die Nutzung des Korpus die Kenntnis einer bestimmten Syntax erforderlich ist. Allerdings sind ihre Grundlagen relativ leicht zu erlernen. Mit zunehmender Erfahrung eignen sich die Benutzerinnen und Benutzer zahlreiche nützliche Mechanismen an, welche die Formulierung von erweiterten Abfragen ermöglichen. Zudem stellt SpoCo einen Mechanismus bereit, mit dem die Benutzerinnen und Benutzer ohne CQL-Kenntnisse Abfragen erstellen können.

2.2.1 CQL-Grundlagen

Die einfachste Abfrage bezieht sich auf die orthographische Form eines Wortes:

- [word="myslał"]: Ermittelt alle Vorkommen von *myslał* im Korpus.

Die obige Abfrage setzt sich aus mehreren Elementen zusammen:

- []: Ein Paar eckiger Klammern zeigt ein einzelnes Wort an.

- *word*: Bezeichnung des sogenannten Positionsattributes, mit dem ein gesuchter Begriff definiert werden soll. In diesem Fall betrifft dies die orthographische Form des Wortes.
- „*myślal*“: Erwarteter Wert des Attributs. Hierbei ist zu beachten, dass der Wert in Anführungszeichen eingeschlossen sein muss (einzelne Anführungszeichen sind ebenfalls zulässig: ‚*myślal*‘).

Der Attributwert wird stets als regulärer Ausdruck betrachtet, sodass auch allgemeinere Abfragen gebildet werden können, zum Beispiel findet

- [word=“myślala?”] Vorkommen der Form *myślal* und *myślala*, und die Abfrage
- [word=“myśl.*”] findet alle Wörter, die mit *myśl* beginnen.

In der Regel wird im CWB-Korpus jedes Wort auf mehr als einer Ebene beschrieben (d. h. mithilfe mehrerer Positionsattribute). Die Abfrage

- [lemma=“myśleć”] findet sämtliche Wörter mit dem Lemma *myśleć* unabhängig von der tatsächlichen grammatischen Form (*myślę*, *myślala*, *myślaloby* usw.).
- [tag=“NN”] findet alle Substantive (im deutschsprachigen Teil des Korpus).
- [tag=“subst.*”] ermittelt alle Substantive (im polnischsprachigen Teil des Korpus).

Folgendes ist noch zu den zwei letzten Beispielen anzumerken. Die Abfrage bezüglich der grammatischen Form (durch Nutzung des Attributs *tag*) ist sprachabhängig. Die grammatischen Kategorien der einzelnen Sprachen weichen mehr oder weniger stark voneinander ab. Darüber hinaus kommen bei der grammatischen Annotation Programme, sogenannte Tagger, zur Anwendung, die abhängig von der Sprache sind, auf deren Grundlage sie funktionieren. Bei der Klassifizierung von Wörtern wird ein sogenanntes Tag-Set verwendet, d. h. ein Satz von Kategorien und Werten, die den einzelnen Kategorien zugeordnet werden können. Die Tag-Sets wiederum unterscheiden sich in der Regel hinsichtlich der Namenskonventionen. Aus diesem Grund ist für Abfragen bezüglich grammatischer Formen die Kenntnis der Tag-Sets erforderlich. Das Tag-Set für die deutsche Sprache ist hier beschrieben: <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/STTS-Tagset.pdf>. Das Tag-Set für die polnische Sprache findet sich unter: <http://nkjpl.pl/poliqarp/help/en.html>. Die Daten in tschechischer Sprache sind in der aktuellen Phase des Projekts noch nicht mit Tags versehen.

Die unterschiedlichen Kriterien für die Suche nach Wörtern können mittels logischer Operationen miteinander verknüpft werden:

- [word=“A.*” & tag=“NN”]: Alle Substantive, die mit *A* beginnen (für Deutsch).
- [word=“pies” | word=“kot”]: Alle Vorkommen der Formen *pies* oder *kot*.

Auch eine Negation kann verwendet werden:

- [lemma=“być” & word!=“byłem”]: Vorkommen des Lexems *być* mit Ausnahme der Form *byłem*.
- [word!=lemma]: Alle Wörter, deren orthographische Form vom Lemma abweicht.

Die Benutzerinnen und Benutzer müssen sich nicht auf die Suche nach einzelnen Wörtern beschränken. Bei Bedarf können sie eine Sequenz für die Suche definieren:

- [word=“Ala”][word=“ma”][word=“kota”]: Vorkommen der drei aufeinanderfolgenden Wörter werden gesucht.
- [lemma=“mieć”][word=“ochotę”][word=“na”][]: Alle Vorkommen des Ausdruckes *mieć ochotę na* + beliebiges Lexem werden gesucht.
- [lemma=“przyjmować”][]{0,2}[word=“do”][word=“wiadomości”]: Zwischen dem Wort des Lemmas *przyjmować* und dem Wort *do* können zwei beliebige Wörter

auftreten (mögliche Treffer wären also beispielsweise *przyjmować do wiadomości*, *przyjmować to do wiadomości*, *przyjmować jego odmowę do wiadomości* usw.).

2.3 Die Arbeit mit dem Korpus

2.3.1 Suchseite

Die Arbeit mit den Korpusdaten beginnt mit der Formulierung der entsprechenden Abfrage. Zu diesem Zweck wird die Suchseite verwendet. Die Abfrage muss (in ihrer endgültigen Form) in der Sprache CQL ausgedrückt werden. Die Benutzerinnen und Benutzer können sie direkt eingeben oder den Abfrage-Generator nutzen, der die Erstellung einfacher Fragen wesentlich erleichtert. Zu diesem Zweck verfügt die Suchseite über zwei Hauptbereiche, und zwar die Einfache Suche und die CQL-Suche.

2.3.1.1 Einfache Suche

The image shows a web interface titled "Basic Search". It has three input fields: "Word", "Lemma", and "Tag". Below the "Word" field are three checkboxes: "begins with", "ends with", and "case sensitive". To the right of these checkboxes is a button labeled "Add/Delete Row" with "+" and "-" symbols. Below the input fields is a "Filters" section with four tabs: "main", "syntactic", "sociolinguistic", and "metadata". Under the "main" tab, there are two dropdown menus: "Generation" and "Language". At the bottom of the interface, there are two checkboxes: "one result per utterance" and "count each PAT/Phrasentyp occurrence".

Abbildung 4: Einfache Suche

Über diesen Bereich (Abbildung 4) können Benutzerinnen und Benutzer ohne CQL-Kenntnisse Abfragen erstellen. Im Generator-Fenster bezieht sich jede Zeile auf ein einzelnes Wort und enthält mehrere Felder, in denen die Werte der Positionsattribute für das betreffende Wort angegeben werden können. Im Fall des LangGener-Korpus sind dies *word* (grammatische Form), *lemma* (Lemma) und *tag* (grammatische Form). Wenn mehr als ein Feld ausgefüllt wird, werden diese Felder in der CQL-Abfrage durch eine logische Operation vom Typ „Konjunktion“ verknüpft (alle Bedingungen müssen erfüllt sein). Folgende zusätzliche Optionen sind verfügbar:

- *begins with*: Gleiches alles ab, was mit dem angegebenen Wert beginnt.
- *ends with*: Gleiches alles ab, was mit dem angegebenen Wert endet.
- *case sensitive*: Unterscheidet zwischen Groß- und Kleinbuchstaben (standardmäßig wird Groß-/Kleinschreibung nicht beachtet).

Falls die Definition einer weiteren Wortsequenz erforderlich ist, kann über „+“ eine weitere Zeile hinzugefügt werden. In diesem Fall kann auch definiert werden, in welchem Abstand zueinander die einzelnen Teile der Sequenz auftreten sollen.

Ein zusätzliches Element des Bereichs ist ein Set von Filtern (korpusabhängig), die angewendet werden können, um die Ergebnisse einer Abfrage einzuschränken. Im Lang-Gener-Projekt gibt es vier Typen von Filtern: Hauptfilter, die es erlauben, die Ergebnisse auf eine ausgewählte Sprache und Generation zu beschränken, syntaktische Filter, bei denen PAT- und Phrasentyp-Werte angegeben werden können sowie soziolinguistische Filter und Metadaten (bezogen auf die Probandinnen und Probanden – Wohnort, Geburtsdatum usw.). Standardmäßig sind die Suchergebnisse in keiner Weise eingeschränkt. Durch das Festlegen eines Wertes für einen beliebigen Filter oder für mehrere Filter – denn die Einschränkungen lassen sich kombinieren – werden Ergebnisse, die die angegebenen Kriterien nicht erfüllen, nicht berücksichtigt. Obwohl die Grundeinheit der Suche ein Wort (oder eine Sequenz) ist, kann es vorkommen, dass der Benutzer nach allen Segmenten (nicht Wörtern) suchen möchte, die die angegebenen Kriterien erfüllen. Dies ist möglich, indem man die Option *one result per utterance* auswählt. Die Option *count each PAT/Phrasentyp occurrence* (siehe dazu die Beschreibung von PAT-Kategorien und Phrasentypen in Bučková, Centner, Księżyk & Prawdziej in diesem Band) ist nützlich, wenn die Benutzerinnen und Benutzer nicht an bestimmten Wörtern oder Segmenten interessiert sind, sondern nur am Vorkommen von bestimmten PAT oder syntaktischen Phrasentypen im Korpus. Diese Konstruktionen stellen ein Zwischenelement zwischen einem Wort und einem Segment dar – sie können aus vielen Wörtern bestehen oder in einem bestimmten Segment mehrfach vorkommen. Die PAT/Phrasentyp-Annotation hat daher die Form einer Folge von kommagetrennten Symbolen, zum Beispiel ‚PAT, AA, PAT, AA‘ bedeutet, dass in dem Segment abwechselnd jeweils zwei PAT- und zwei AA-Konstruktionen auftreten. Bei Auswahl dieser Option wird jedes Segment so oft als separates Ergebnis behandelt, wie es die angegebenen PAT/Phrasentyp-Werte enthält, was die Analyse erleichtert und dafür sorgt, dass die Anzahl der erhaltenen Ergebnisse mit der Anzahl der gesuchten Strukturen übereinstimmt. Wenn der Benutzer zum Beispiel nach einer AA-Konstruktion sucht und die Option *count each PAT/Phrasentyp occurrence* auswählt, wird ein Segment mit der Annotation ‚PAT, AA, PAT, AA‘ zweimal erscheinen (und zweimal gezählt werden). Diese Option sollte in der Regel in Verbindung mit der Option *one result per utterance* verwendet werden.

Jede Aktion der Benutzerinnen und Benutzer in diesem Bereich wird kontinuierlich in die entsprechenden CQL-Strukturen übersetzt. Die auf diese Weise generierte Abfrage wird im Bereich *CQL Search* (CQL-Suche) angezeigt. Dies erfüllt folgende Zwecke: Erstens machen sich die Benutzerinnen und Benutzer so mit der Syntax von CQL vertraut, wodurch sie direkten Kontakt zur Sprache CQL haben und ihre Grundlagen in der Praxis anwenden können. Zweitens erhalten sie die Möglichkeit, die bei diesem Prozess erzeugte Abfrage ihren Bedürfnissen entsprechend anzupassen.

2.3.1.2 CQL-Suche

Abbildung 5: CQL-Suchfeld

In diesem Bereich (Abbildung 5) können die Benutzerinnen und Benutzer eigenständig eine gültige CQL-Abfrage eingeben. Dies ermöglicht die Formulierung erweiterter Abfragen, deren Erstellung mit dem Suchassistenten für die einfache Suche nicht möglich wäre (zum Beispiel Verwendung des Operators für Negation oder Alternative). Im Abschnitt *Basic Search* (einfache Suche) erstellte Abfragen werden hier angezeigt und können beliebig geändert werden.

2.3.2 Ergebnisseite

Abbildung 6: Ergebnisseite – Informationen über die Abfrage und Symbole zur Kontrolle der Ergebnisdarstellung

Die Ergebnisse der Abfrage werden auf der Ergebnisseite angezeigt (Abbildung 6). Diese ist in zwei Bereiche unterteilt. Im oberen Bereich werden die Basisdaten der Abfrage angezeigt: ihre CQL-Form, die Anzahl der Ergebnisse sowie Symbole, über welche die Ergebnisanzeige angepasst werden kann:

- Ansicht ändern (siehe unten)
- Metadaten ein-/ausblenden
- Markierte Ergebnisse speichern
- Alle Ergebnisse speichern

Recording	Transcription	
▶ 0:00 / 0:00	wynikało to naturalnie z filmu Czterech Pancernych i Pies tak fakt hmm	▼ <input type="checkbox"/>
▶ 0:00 / 0:00	zresztą na każdy dzwonięcie tak nasz Leon dlatego znaczcy to ten pies	▼ <input type="checkbox"/>
▶ 0:00 / 0:00	jakiś pies z kulawą nogą gdzieś tam za winklem przejdzie	▼ <input type="checkbox"/>
▶ 0:00 / 0:00	a pies naturalnie nie szczekol na mnie	▼ <input type="checkbox"/>
▶ 0:00 / 0:00	bo to bo pies leży	▼ <input type="checkbox"/>

Abbildung 7: Ergebnisseite – Basisansicht

	Left Context	Match	Right Context	
▶ +	wynikało to naturalnie z filmu Czterech Pancernych i	Pies	tak fakt hmm	<input type="checkbox"/>
▶ +	na każdy dzwonięcie tak nasz Leon dlatego znaczcy to ten	pies		<input type="checkbox"/>
▶ +		jakiś	z kulawą nogą gdzieś tam za winklem przejdzie	<input type="checkbox"/>
▶ +		a	naturalnie nie szczekol na mnie	<input type="checkbox"/>
▶ +		bo to bo	leży	<input type="checkbox"/>
▶ +		że	sie łasi	<input type="checkbox"/>

Abbildung 8: Ergebnisseite – KWIC-Ansicht

Der zweite Bereich – der eigentliche Hauptbereich – zeigt die Ergebnisse in Form einer Tabelle an.


Jeder Treffer wird als separate Zeile angezeigt, die folgende Elemente enthält:

- 1) Audiofragment des Segments, in dem der Treffer aufgetreten ist.
- 2) Transkription des Segments, aufgeteilt in: linker Kontext, Treffer, rechter Kontext. Für jedes Wort können die Benutzerinnen und Benutzer zusätzliche Informationen zum Lemma und zum grammatischen Merkmal anzeigen lassen, indem er den Mauszeiger auf das betreffende Wort hält. Zudem können die Benutzerinnen und Benutzer den Kontext um mehrere vorangehende und nachfolgende Segmente erweitern. In diesem Fall wird das zentrale Segment speziell hervorgehoben und jedes der benachbarten Segmente kann nach einem Klick auf das entsprechende Symbol abgespielt werden.
- 3) Kontrollkästchen zum Speichern der Ergebnisse (siehe unten).



Die Daten können auf zwei Arten angezeigt werden:

- 1) Basisansicht: Das Audioelement ist aufgeklappt und alle Elemente der Transkription des Segments (linker Kontext, Treffer und rechter Kontext) werden als vollständiger Text gespeichert. Hierbei wird der Treffer hervorgehoben. Diese Ansicht schafft eine gute Übersicht über den Inhalt der Treffer (siehe Abbildung 7).

- 2) KWIC-Ansicht (**Keyword in Context**): Hier ist der Audio-Player auf ein kleines Symbol reduziert. Die Kontexte und der Treffer werden in separaten Spalten angezeigt. Dies ermöglicht eine einfachere Analyse der Treffer im Kontext. In dieser Ansicht können die Ergebnisse auf der Grundlage des Treffers bzw. anhand des Kontexts alphabetisch sortiert werden (siehe Abbildung 8).

Über das Symbol  kann zwischen den einzelnen Ansichten gewechselt werden.

2.2.3 Speichern von Ergebnissen

Die Abfrageergebnisse können in einer Datei vom Typ **TSV** (*tab-separated values*, dt.: tabulatorgetrennte Werte) gespeichert werden. Bei diesem Format werden alle Ergebnisse in einer separaten Zeile gespeichert, während die zugehörigen einzelnen Elemente durch Tabulator voneinander getrennt werden. Dieser Dateityp kann zur weiteren Analyse in Tabellenkalkulationen (zum Beispiel Excel, LibreOffice, Google Sheet) geöffnet werden. Über  können alle Ergebnisse gespeichert werden, während mit  nur die markierten Ergebnisse gespeichert werden.

In der Basisansicht kann auch das Audiofragment gespeichert werden, das dem Segment zugeordnet ist (über das Symbol  im Audio-Player).

Literatur

- Daniel, Michael/Dobrushina, Nina/Waldenfels, Ruprecht von (2014): The language of the Ustja river basin. A corpus of North Russian dialectal speech. 2013. Bern, Moscow. Online verfügbar unter: <http://www.parasolcorpus.org/Pushkino/login.php>, letzter Zugriff am 08.03.2022.
- Evert, Stefan/Hardie, Andrew (2011): Twenty-first century corpus workbench: updating a query architecture for the new millennium. In: *Proceedings of the Corpus Linguistics 2011 Conference*. Birmingham, UK, Paper 153. Online verfügbar unter: <https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-153.pdf>, letzter Zugriff am 08.03.2022.
- Grochola-Szczepanek, Helena/Górski, Rafał L./Waldenfels, Ruprecht von/Woźniak, Michał (2019): Korpus języka mówionego mieszkańców Spisza. *LingVaria* 14.27, 165–180. Online verfügbar unter: spisz.ijp.pan.pl, letzter Zugriff am 12.03.2022.
- Piasecki, Maciej (2007): Polish Tagger TaKIPI: Rule Based Construction and Optimisation. *Task Quarterly* 11, 151–167. Online verfügbar unter: <http://www.nlp.pwr.wroc.pl/en/tools-and-resources/tools/takipi>, letzter Zugriff am 08.03.2022.
- Pristupov, Danil (2022): Fork. A fast and friendly Git client. In: <https://git-fork.com/>, letzter Zugriff am 28.04.2021.
- Rabus, Achim (2015): Current developments in Carpatho-Rusyn speech – preliminary observations. In: Patricia Krafcik, Valerii Padiaki (Hgg.), *Juvilejnyj zbirnyk na čest' profesora Pavla-Roberta Magočija*. Užhorod, 489–496. Online verfügbar unter: <http://www.russinisch.uni-freiburg.de/>, letzter Zugriff am 08.03.2022.
- Schmid, Helmut (1995): Improvements in part-of-speech tagging with an application to German. In: S. Armstrong, K. Church, P. Isabelle, S. Manzi, E. Tzoukermann, D. Yarowsky (Hgg.), *Natural Language Processing Using Very Large Corpora. Text, Speech and Language Technol-*

- ogy, 11. Dordrecht: Springer, 13–25. Online verfügbar unter: <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>, letzter Zugriff am 08.03.2022.
- Waldenfels, Ruprecht von/Woźniak, Michał (2017): SpoCo – a simple and adaptable web interface for dialect corpora. In: M. Kupietz, A. Geyken (Hgg.), *Special Issue on Corpus Linguistic Software Tools*, *Journal for Language Technology and Computational Linguistics* 31.1, 155–170.

Dokumentation und Analysen

Dialektale Merkmale in den sprachbiographischen Interviews der älteren Generation in Polen

Anna Jorroch – Felicja Księżyk – Anna Zielińska

Abstract:

Dialectal features in biographical interviews with the older generation living in Poland

One of the goals of the specialized corpus of Polish-German bilingualism, created as part of the project *Language across Generations: Contact Induced Change in Morphosyntax in German-Polish Bilingual Speech*, is to document the endangered German dialects in western and northern Poland. This chapter illustrates the regional diversity of German dialectal features in terms of phonetics and morphology, which are a continuation of the Silesian and Low German dialects. The analysis is based on thirty biographical interviews with representatives of the older generation inhabiting the areas incorporated into Poland in 1945. Further, the most important features of Silesian in Polish are presented.

1 Einführung

Eines der wichtigsten Ziele des Projektes „Language across Generations: Contact Induced Change in Morphosyntax in German-Polish Bilingual Speech“ ist die Dokumentation der dialektalen Erscheinungsformen im Gebrauch des Deutschen der „Generation Polen“. Darüber hinaus werden auch schlesische polnische Merkmale im Gebrauch des Polnischen der Interviewten in den Blick genommen.

2 Historische deutsche Dialektgebiete im heutigen Polen

Die deutschen dialektalen Merkmale, die in den durchgeführten Interviews dokumentiert werden konnten, bedürfen einer detaillierteren und differenzierteren Betrachtung in Anbetracht dessen, dass die Sprecherinnen und Sprecher aus Regionen stammen, die sprachgeographisch unterschiedlichen historischen Dialektgebieten zuzuordnen sind. Die Herkunftsorte der Interviewten liegen auf Gebieten des heutigen Polens, die vor 1945 zum Deutschen Reich gehörten. Verwaltungsmäßig lassen sich die Geburtsorte der Sprecherinnen und Sprecher zum Zeitpunkt ihrer Geburt den Provinzen Schlesien (zeitweise untergliedert in die Provinz Nieder- und Oberschlesien), Pommern (woraus 1922 die bis

1938 bestehende Provinz Grenzmark Posen-Westpreußen¹ ausgegliedert wurde), sowie West- und Ostpreußen zuordnen.

Dialektgeographisch sind diese Herkunftsorte im Rahmen des Schlesischen, Ostpommerschen, Nieder- und Hochpreussischen zu verorten. Die historische Verteilung der deutschen Dialekte auf den Gebieten des heutigen Polens veranschaulicht folgende Karte aus Wiesinger (1983a: 830).



Karte 47.4: Einteilung der deutschen Dialekte: Die Gliederung der deutschen Dialekte (in den ersten Jahrzehnten des 20. Jhs.)

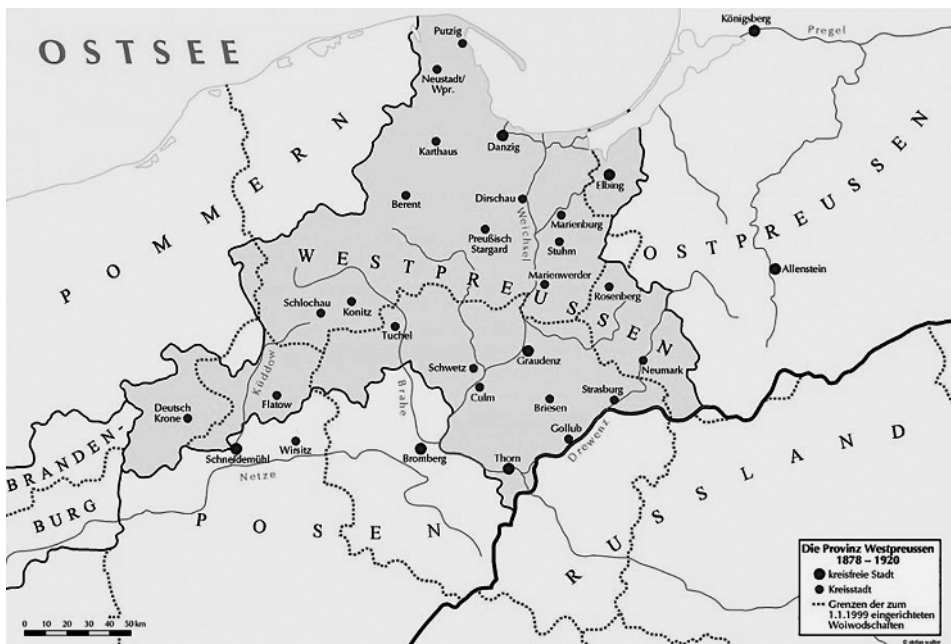
Karte 1: Gliederung der deutschen Dialekte in den ersten Jahrzehnten des 20. Jahrhunderts (Wiesinger 1983a: 830, Karte 47.4)

¹ Die Grenzmark Posen/Westpreußen war deckungsgleich mit dem Regierungsbezirk Schneidemühl/Piła. Ab 1938 bis 1945 bildete sie als Regierungsbezirk erneut einen Teil der Provinz Pommern. Siehe dazu Porada (2015).

Die im Rahmen des Projekts LangGener interviewten schlesischen Sprecherinnen und Sprecher stammen überwiegend aus Oberschlesien, ihre Geburtsorte liegen größtenteils in den Grenzen der heutigen Woiwodschaft Oppeln. Eine solche Wahl der Informantinnen und Informanten rührt daher, dass in der Oppelner Woiwodschaft heute fast die Hälfte der Angehörigen der deutschen Minderheit in Polen lebt, wohingegen der Anteil der deutschstämmigen Bevölkerung in Niederschlesien heute gering ist (vgl. Gudaszewski 2015: 47). Auch stammen einige der deutschstämmigen Bewohnerinnen und Bewohner Niederschlesiens aus Oberschlesien.

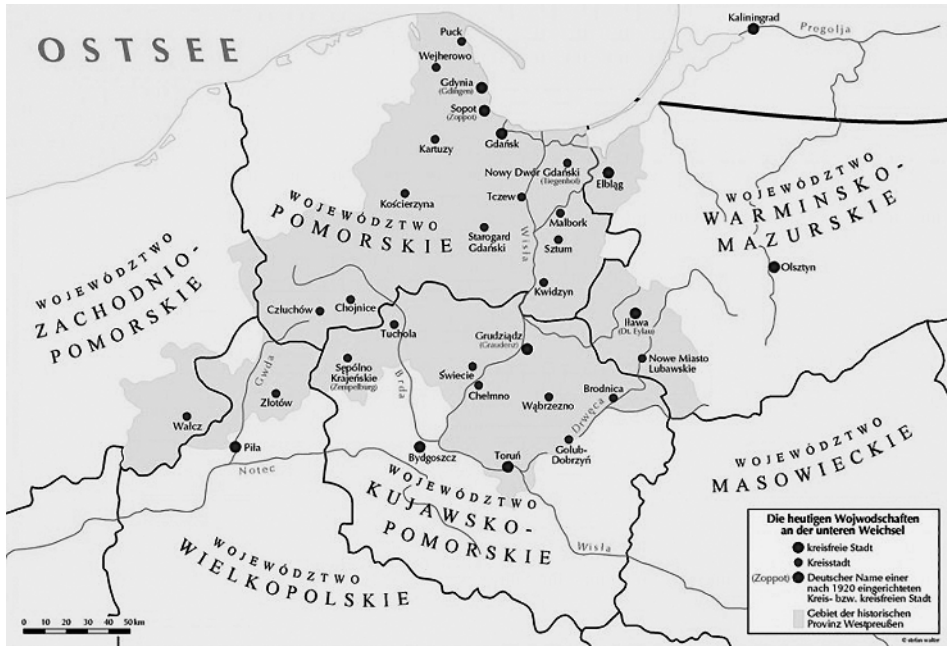
2.2 Niederdeutsche Dialekte im heutigen Polen

Neben Sprecherinnen und Sprechern aus Schlesien wurden die Interviews auch mit Informantinnen und Informanten durchgeführt, die in allen anderen ehemals deutschen Gebieten des heutigen Polens leben. Die Erhebungsorte in Pommern gehören dialektgeographisch dem Ostpommerschen an, d. h. einem niederdeutschen Sprachraum, der, so Wiesinger (1983a: 886), nur sehr unzulänglich erforscht wurde. Ebenfalls innerhalb der niederdeutschen Dialekte sind die Ortschaften zu verorten, die sich auf dem Gebiet der früheren Provinz Westpreußen befinden: Gdańsk (Danzig) und Elbląg (Elbing). Genauer genommen handelt es sich dabei um das sog. Niederpreußische, wobei ein enger Kontakt mit dem Hochpreußischen berücksichtigt werden muss.



Karte 3: Provinz Westpreußen 1878–1920 (Westpreußische Gesellschaft 2022)

Das Territorium der ehemaligen Provinz Westpreußen gehört heute zu fünf Woiwodschaften, wie es die Karte 4 veranschaulicht.

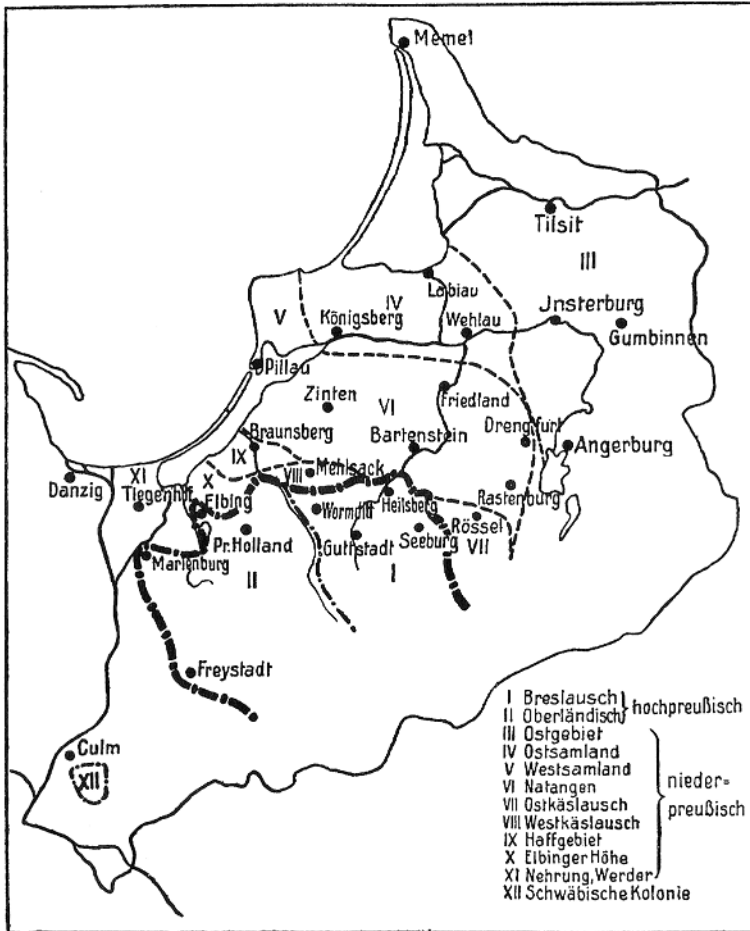


Karte 4: Fünf Woiwodschaften auf dem Gebiet des ehemaligen Westpreußens (Westpreußische Gesellschaft 2022)

Die im Rahmen des Projektes durchgeführte Feldforschung umfasst ebenso den südlichen Teil Ostpreußens, d. h. Ermland und Masuren, das 1945 an Polen angeschlossen worden ist. Der nördliche Teil, der Königsberger Kreis, der heute zu Russland gehört, wurde nicht berücksichtigt, da Sprachkontaktphänomene in der Rede der deutsch-polnischen Bilingualen erforscht werden sollten. In der ehemaligen Provinz Ostpreußen stammen die interviewten Bilingualen direkt aus den Ortschaften bzw. aus der Umgebung von Giżycko (Lötzen), Kętrzyn (Rastenburg), die im nördlichen Teil der Region liegen, aus der Gegend von Elk (Lyck) im Osten sowie von Ruciane-Nida (Rudczanny) und Ukta in Südwestmasuren. Die Verortung der durchgeführten Aufnahmen ergibt gewisse Unterschiede bezüglich der dialektalen Eigenschaften in der Rede der zweisprachigen Informantinnen und Informanten.

Die Analyse der phonetischen Merkmale des Deutschen zeigt zahlreiche Einflüsse verschiedener Sprachsysteme, neuhochdeutsche Elemente und auch ältere lautliche Formen aus den früheren Entwicklungsetappen der Sprache. Dies ist mit der Tatsache verbunden, dass Masuren ein Grenzgebiet ist, in dem Vertreter mehrerer Kultur- und Sprachgemeinschaften zusammenlebten und bis heute leben. Die älteren Sprachformen hängen mit der Einwanderung der Siedler aus verschiedenen Teilen des heutigen Deutschlands zusammen. Die Siedler kamen sowohl aus dem mittel- und niederdeutschen Sprachgebiet wie

Magdeburg, Halberstadt und Hessen-Nassau als auch aus der heutigen Schweiz und Österreich, vor allem aus Salzburg (Kossert 2001: 91f.). Walther Ziesemer (2009: 119–136) hat die ostpreußischen Dialekte in die hoch- und niederpreußische Gruppe eingeteilt und dann näher charakterisiert. Die Verteilung der Mundarten hat der Forscher auf einer Karte dargestellt, die zeigt, dass sich die Wohngegenden unserer Sprecherinnen und Sprecher im südlichen Teil *des Ostgebiets*, im östlichen Teil *des Natangischen* und im Süden *des Ostkäslauschen* im Bereich der niederpreußischen Mundarten, allerdings in der Nähe der hochpreußischen Gruppe befinden (ebd.: 137).



Die ostpreußischen Mundarten.

Karte 5: Die ostpreußischen Mundarten (Ziesemer 2009: 137)

3 Restdialektale Eigenschaften des Deutschen im heutigen Polen

Die deutschsprachigen Aufnahmen der interviewten Sprecherinnen und Sprecher stellen keine Mundartproben dar, wie wir sie vom Anfang des 20. Jahrhunderts etwa für das Schlesische von Graebisch (1929) oder Mak (1929) kennen. Aufgrund des Bevölkerungsaustausches nach 1945 kommen deutsche Dialekte in Polen heute nur rudimentär vor. In einigen Äußerungen deuten zwar die Sprecherinnen und Sprecher an, ihre Vorfahren hätten sich eines deutschen Dialekts bedient, der Dialektabbau sei aber bereits in ihrer Kindheit so fortgeschritten gewesen, dass sie diese dialektalen Äußerungen selbst nicht verstanden hätten. So berichtet die Informantin IT_GLO_GP, dass ihre Großmutter aus Kranzdorf (heute Nowy Browiniec, ein Dorf nahe Oberglogau, heute Głogówek, das historisch zum Gebirgsschlesischen gehörte) in einem deutschen Dialekt sprach, den sie jedoch nicht verstanden habe:

- (1) Interv. (und wissen Sie vielleicht wie die Großeltern zu Hause gesprochen haben <?>)

Plattdeutsch was ich überhaupt nicht kann hab auch nicht verstanden meine Oma hat Plattdeutsch gesprochen sie konnte auch Deutsch

Interv. (Aber die von jetzt von?)

die aus aus Kranzdorf die hat die Oma die aus Oberglogau die hat Deutsch gesprochen aber die Oma aus Kranzdorf hat Plattdeutsch gesprochen also da kam ich nicht mit (IT_GLO_GP)

Ähnlich verhält es sich in den anderen Gebieten Polens. So waren unsere Informantinnen und Informanten aus Pommern, der ehemaligen Grenzmark Posen-Westpreußen und Masuren, ebenfalls keine aktiven Dialektsprecherinnen und -sprecher mehr. Sie verwiesen entweder darauf, dass in ihrer Umgebung Standarddeutsch gebraucht worden sei:

- (2) ja aber das mehr in den Dörfern in der Stadt nicht ich hab nie Plattdeutsch gehört obwohl meine Großeltern auf dem Dorf wohnten [...] aber sie haben immer Hochdeutsch gesprochen nicht Platt denn ich hab Platt Platt hab ich nie gehört (QK_BOB_GP)

Oder sie beteuerten, dass zwar jemand aus ihrer Umgebung noch Dialektsprecher gewesen sei, sie selbst hätten jedoch keinen Dialekt gesprochen:

- (3) die konnte Deutsch und Platt sogar nicht <?> Platt ja (HD_SWI_GP)

- (4) a moja babcia i dziadek to między sobą mówili językiem Platt językiem który no podobny jest do jakiegoś dialektu jak kaszubski tylko że zbliżony raczej do angielskiego języka ja kiedyś jak byłem mały to wiedziałem o czym mówią ale nie powtórzę teraz (XL_PIL_GP)

„meine Großmutter und mein Großvater sprachen untereinander die Platt-Sprache eine Sprache die einem Dialekt ähnlich ist wie zum Beispiel Kaschubisch aber näher am Englischen liegt als ich klein war wusste ich schon wovon sie sprachen aber ich kann es jetzt nicht mehr wiederholen“

Die heute nur rudimentär auftretenden deutschen Dialekte in Polen korrespondieren mit Untersuchungsbefunden, die bereits in der ersten Hälfte des 20. Jahrhunderts ansatzweise

formuliert wurden. Darin wird hinsichtlich der hier analysierten Dialekte auf einen weitgehenden Dialektrückgang hingewiesen, wie das beispielsweise in dem nachfolgenden Zitat aus einer Erhebung von 1937 explizit für den ostpommerschen Dialektraum diagnostiziert wird:

„Bei der Untersuchung mußte ich die Feststellung machen, daß die Mundart hier immer mehr an Bedeutung zu verlieren droht. An vielen Orten beherrscht sie nur noch ein kleiner Hundertsatz der Kinder. Sie wird oft als minderwertig gegenüber der Schriftsprache empfunden. Die Mundart gilt als nicht ‚fein‘ genug, und in vielen Familien wird deshalb mit den Kindern, besonders mit den Mädchen, nur noch hochdeutsch gesprochen. So scheint das Platt in dieser Landschaft allmählich auszusterben, falls die Bemühungen, die Mundart wieder zu beleben, die heute von vielen Seiten her einsetzen, scheitern.“ (Stritzel 1937: 2)

Die im heutigen Polen vorkommenden Varietäten des Deutschen stellen folglich gemessen an der Triade *Dialekt-Umgangssprache-Standardsprache* eine relativ standardnahe Umgangssprache dar. Somit wird in der einschlägigen Literatur etwa zum Deutschen in Oberschlesien konstatiert, dass es sich vom Standarddeutschen insbesondere durch die darin vorkommenden Sprachkontaktphänomene abhebt (vgl. Pelka 2006: 126f.; Matuschek 1998: 195; Bellmann 1971: 10, 16; Rohfleisch 2001: 165f.; Reiter 1960: 64; Lasatowicz 2002: 780; Lasatowicz & Weger 2008: 160).

Neben sprachkontaktinduzierten Standardabweichungen lassen sich in den durchgeführten sprachbiographischen Interviews allerdings auch einige remanente dialektale Merkmale aufzeigen, die unsere Sprecherinnen und Sprecher ungeachtet des fortgeschrittenen Dialektabbaus in ihren deutschsprachigen Äußerungen realisieren und die als Anzeichen von Dialektkontinuität gelten können.

Als konstitutives dialektale Merkmal bei den schlesischen Sprecherinnen und Sprechern (vgl. (5) und (6)), aber auch bei Informantinnen und Informanten in Ostpreußen (vgl. (7) und (8)) lässt sich die Entrundung deutscher Vordervokale [y:] > [i:] bzw. [ʏ] > [ɪ] sowie [ø:] > [e:] bzw. [œ] > [ɛ] werten.

- (5) die haben das immer *ausgespirt* [ausgə'ʃpi:rt] [‚die haben das immer ausgespürt‘] (SL_OPO_GP)
- (6) ich konnte mich doch nicht *bicken* ['bikən] [‚ich konnte mich doch nicht bücken‘] (IT_GLO_GP)
- (7) haben sie umgedreht und sind *zurückgefahren* [tsʊrikgə'fɑ:rən] [‚zurückgefahren‘] nach Falkenau (AL_PAR_GP)
- (8) in die Wohnung die haben das alles jelassen... hier die *Mebel* ['me:bəl] [‚Möbel‘] auch das was alles war und so (AL_PAR_GP)

In Oberschlesien kann dieses restdialektale Merkmal insbesondere bei einigen der vor 1945 monolingual aufgewachsenen Sprecherinnen und Sprechern relativ konsequent beobachtet werden. In der Rede der Sprecherinnen und Sprecher in Ostpreußen kann es auch zur Variation kommen, wenn die gerundeten Vordervokale sowohl standardmäßig realisiert als auch entrundet werden:

- (9) wieder *kije* ['ki:jə] [„Kühe“] ... zusammenjachen und melken da die Frauen haben gemelken gefüttert und so was da (AL_PAR_GP)

Zwar wird bei dieser Erscheinung zum Teil auch auf einen potenziellen Einfluss der polnischen Kontaktsprache hingewiesen (vgl. Pelka 2007: 135), allerdings dürfte in dem hier untersuchten Korpus die Annahme einer Dialektkontinuität plausibler erscheinen. Zum einen deshalb, weil bei den interviewten schlesischen Sprecherinnen und Sprechern, die zweisprachig aufgewachsen sind (d. h. in der Domäne „Familie“ den schlesisch-polnischen Ethnolekt gebrauchten, das Deutsche wiederum in der Schule und in der offiziellen Domäne), derartige Entrundungen, wenn sie überhaupt vorgekommen sind, eher vereinzelt Charakter hatten. Zum anderen darum, weil die Entrundung gerundeter Vokale zu den sog. gemeinschlesischen Merkmalen gerechnet wird (vgl. Graebisch 1929: 121; Friemel 1937: 104). In Ostpreußen dürfte die Vokalentrundung hingegen auf den Einfluss des Hochpreußischen zurückzuführen sein.

Insbesondere bei den monolingual deutsch sozialisierten Sprecherinnen und Sprechern, die auch bei der Schilderung ihres Sprachgebrauchs den Stammvokal im Verb *können* [kənən] entrunden (oft mit einhergehender en-Apokope), wäre somit nicht pauschal auf eine *Pattern replication* (PAT) in Anlehnung an die polnische Kollokation *znać język* („eine Sprache kennen“) zu schließen, etwa:

- (10) die *kenn* [kən] Deutsch [„die können Deutsch“] (SL_OPO_GP)

Es zeigt sich nämlich, dass diese Probandinnen und Probanden im Präteritum in ähnlichen Kontexten eindeutig das Verb *können* gebrauchten:

- (11) die Lehrerin war eine ältere Frau, die Deutsch konnte (SL_OPO_GP)
 (12) sie konnte auch Deutsch (IT_GLO_GP)

Darüber hinaus wird das Verb *können* bei diesen Informantinnen und Informanten auch in anderen Verwendungskontexten entrundet, wie die folgenden Beispiele veranschaulichen:

- (13) *kenn* [kən] [„können“] wir nischt machen (BT_KLO_GP)
 (14) die sitzen in Oppeln auf dem Bahnhof und *kenn* [kən] [„können“] nicht hin und nicht zurück (KQ_GOG_GP)
 (15) heutzutage *kenn* [kən] [„können“] wir sprechen so wie wir wollen nicht <?> (KQ_GOG_GP)

Da das Schlesische sich dialektgeographisch von anderen untersuchten Regionen abhebt, offenbart es einige Merkmale, die anderswo in Polen nicht anzutreffen sind. Anders als im ehemals niederdeutschen Pommern, West- und Ostpreußen war nämlich das aus mehreren Untermundarten bestehende Schlesische teils stärker mitteldeutsch, teils eher oberdeutsch-ostfränkisch geprägt. Es sei daher zunächst auf einige Merkmale eingegangen, die wir nur bei unseren schlesischen Informantinnen und Informanten angetroffen haben.

Relativ verbreitet ist in Schlesien die Wiedergabe des Suffixes *-er* nicht in Form des standardgerechten Reduktionsvokals [ɐ], sondern als offenes [ɛ]:

- (16) mit de *Mutte* ['mɔtɛ] [,Mutter⁴] (SL_OPO_GP)
- (17) der *Eigentime* ['aegənti:mɛ] war nicht da [,der Eigentümer war nicht da⁴] (IT_GLO_GP)
- (18) ich war so *ibemidet* [i:bɛ'mi:dət] [,übermüdet⁴] (KQ_GOG_GP)
- (19) die *Tochte* ['tɔxtɛ] ist dann fünf Jahre *jinge* ['ji:ŋɛ] nicht <?> [,Die Tochter ist dann fünf Jahre jünger⁴] (IT_GLO_GP)

Vor diesem Hintergrund sind auch folgende Redeauszüge nicht vorschnell als Musterreplikationen (PAT) zu deuten. Die nachfolgenden Belege exemplifizieren nämlich keine Subjektauslassung in Anlehnung an das Polnische, vielmehr wird hier das Personalpronomen *er* als offenes [ɛ] realisiert, was irrtümlicherweise als Flexionsendung herausgehört werden könnte:

- (20) bei de Tante war *e* [ɛ] gewesen dort. Hat *e* [ɛ] war *e* [ɛ] dann angemeldet (KQ_GOG_GP)
- (21) die Schuhe da hat *e* [ɛ] die Sohle so (BQ_RAC_GP)

Als Dialektkontinuität in Schlesien ist auch die bei einigen Sprecherinnen und Sprechern vorkommende s-Palatalisierung bei vorausgehendem [r] zu werten, etwa:

- (22) *zuallerersch* hat sie mir die Puppe weggeschmissen (IT_GLO_GP)
- (23) die *erschte* Zeit (SL_OPO_GP)
- (24) Ich habe musste mir einmal *ersch* ein Kursus machen (BQ_RAC_GP)

Restdialektalen Charakter hat ebenso die vereinzelt vorkommende stimmhafte Aussprache des intervokalischen [f], die Reiter (1960: 64) auf niederschlesischen Einfluss zurückführt, etwa:

- (25) die Jungs die wohnen in Libeck ['li:bək] [,Lübeck⁴] da ist eine *Hawenstadt* ['ha:vənʃtat] [,Hafenstadt⁴] (BT_KLO_GP)
- (26) auf *Bauernhewen* ['baʊənhe:vən] [,Bauernhöfen⁴] (IT_GLO_GP)

Relativ verbreitet ist in den deutschsprachigen Äußerungen der Oberschlesier die en-Apokope nach Nasallauten, etwa:

- (27) und da hat er uns *mitgenomm* [,mitgenommen⁴] (KQ_GOG_GP)
- (28) wir *könn* [,können⁴] hier nicht bleiben (KT_DOB_GP)

Was in den deutschsprachigen Teilen der Interviews mit den Oberschlesiern auffällt, ist ebenso die Tatsache, dass hier einige heute veraltet klingende sprachliche Besonderheiten auftreten, die dem Sprachstand vor 1945 entsprechen, etwa die teilweise vorkommende singularische Dativmarkierung mit Hilfe der Endung *-e*:

(29) und sie hatte vom vierten *Kinde* hatte sie Löcher in den Beinen (IT_GLO_GP)

(30) und er hat im *Kopfe* gehabt (QN_BRO_GP)

Ebenso zählt dazu die teilweise vorkommende Abhebung der prädikativen Verwendung der Zahladjektive vom attributiven Gebrauch durch Schwa-Epithese, etwa:

(31) da waren *zweie* aus Hindenburg (IT_GLO_GP)

vs.

(32) die eine Schwester hatte *zwei* kleine Kinder (IT_GLO_GP)

Diese Erscheinung begegnet auch bei unseren Sprecherinnen und Sprechern, die in der Provinz Westpreußen geboren wurden:

(33) und die fahren mit dem Auto morgens um *sechse* und um *viere* kommt er zurück (SO_NIE_GP)

Die deutschen Varietäten in den ehemaligen deutschen Gebieten Polens außerhalb Schlesiens teilen miteinander mehr sprachliche Gemeinsamkeiten, worauf nachfolgend eingegangen wird.

Besonders auffallend ist hier die Realisierung des anlautenden [g] vor den vorderen Vokalen als [j] (vgl. Stritzel 1937: 56), wobei dies im Korpus nicht durchgehend so ist:

(34) man hat's *jelernt* ruck zuck (SB_KLA_GP)

(35) meine Eltern sind von sechs Jahre in die Schule *jegangen* (HD_SWI_GP)

(36) ja da haben wir dann *jewohnt* (SO_NIE_GP)

(37) kannte auch nicht da die *Jejend* [',Gegend'] (AL_PAR_GP)

(38) na dann wurde immer Polnisch gesprochen na dann hat er *verjessen* (AL_PAR_GP)

Bei einigen Probandinnen und Probanden in Ostpreußen tritt darüber hinaus variabel eine Art Zwischenlaut [g̊] auf:

(39) na später alles ist die eine Schwester hat *gjeheiratet* [...] die haben da gewohnt, nachdem haben wir da *gjewohnt* in die Wohnung, die haben das alles *jelassen* (AL_PAR_GP)

(40) na das ist schon so *gjewehnt* [qjə've:nt] [',gewöhnt'] (UM_JAN_GP)

(41) meine Schwester die haben die Russen *weggjenommen* [',weggenommen']... eine auch die andere (UM_JAN_GP)

Nach hinteren Vokalen ist demgegenüber die Spirantisierung des [g] zu [x] gut vertreten (vgl. Stritzel 1937: 55):

(42) so wie Papa *sachte* [',sagte'] (IL_PIL_GP)

- (43) ein Blitz *ingeschlachen* [‚eingeschlagen‘], nicht <?> (HD_SWI_GP)
- (44) na was so zwei *Tache* [‚Tage‘] an dem zweiten Tach [‚Tag‘] (AL_PAR_GP)
- (45) mitm Bruder haben wir sie noch zum *Zuch* [‚Zug‘] nach Kętrzyn gebracht mitn *Handwachen* [‚Handwagen‘] (AL_PAR_GP)

Nur sporadisch treten dagegen Belege für die im Niederdeutschen unterbliebene zweite Lautverschiebung auf (vgl. Wolf 1983: 1116):

- (46) no aber gewohnt hab *ik* [‚ich‘] alleine (HD_SWI_GP)
- (47) wie heißt *dat* [‚dass‘] (SO_NIE_GP)
- (48) die haben die Schul<*> die Schulen immer ham waren Soldaten in die Schule da<*> *dat* [‚das/dass‘] Lehrer so etwas draußen hat er uns da so (AL_PAR_GP)
- (49) da jekocht haben sie da an der um die Kirche Ziegel zwei Ziegel hinjestellt *kochtop* [‚Kochtopf‘] und da (AL_PAR_GP)

Auf der morphosyntaktischen Ebene gilt der Verlust von Kasusdifferenzierungen als Besonderheit der deutschen Dialekte (vgl. Barbour & Stevenson 2012: 91). Sehr auffällig bei unseren Informantinnen und Informanten in Pommern, West- und Ostpreußen ist das Vorkommen des niederdeutschen obliquen Einheitskasus in Form des Akkusativs (vgl. Wiesinger 1983a: 868), was sich am deutlichsten in Präpositionalphrasen äußert und noch am konsequentesten in Pommern realisiert wird:

- (50) *in die Zeit* war ich staatenlos (QK_BOB_GP)
- (51) bloß zu Hause wenn ich nach Hause kam *mit die Mutter* nicht da sprach ich Deutsch (IL_PIL_GP)
- (52) gehe ich *mit die mit die Nachbarin* ihre Enkelin die kleine kommt *mit die* spreche ich denn (SO_NIE_GP)
- (53) und *in die Scheune* hatten die Soldaten Küche (AL_PAR_GP)
- (54) *in die Kirche* im Januar Frost damals ja war jewesen (AL_PAR_GP)

Als Merkmal der niederdeutschen Dialekte gilt auch die fehlende Dativ-/Akkusativ-Unterscheidung der Personalpronomina (vgl. Barbour & Stevenson 2012: 91; Lübben 1882: 106). In unserem Korpus tritt insbesondere in West- und Ostpreußen die vereinheitlichte Form *mi* auf:

- (55) gebt *mi* Ruh (SO_NIE_GP)
- (56) dann sie fruch *mi* [‚dann fragte sie mich‘] was ich essen will hier (UM_JAN_GP)

Darüber hinaus kommen in unserem Korpus jedoch beide Formen vor, vielfach treten jedoch diesbezüglich Abweichungen von der deutschen Standardsprache auf:

- (57) aus Stettin haben sie *mir* denn mitgenommen (HD_SWI_GP)
- (58) einen Tag ruft *mir* mein Chef (HD_SWI_GP)
- (59) und *mir* haben sie immer friher Friedel jenannt (AL_PAR_GP)
- (60) und einmal traf ich *ihr* (UM_JAN_GP)

Erwähnenswert erscheint, dass in diesem Kontext Verzögerungssignale auftreten, die davon zeugen dürften, dass die Sprecherinnen und Sprecher unsicher bei der Unterscheidung dieser Kasusformen sind:

- (61) keiner konnte *mi*<*> *mich* erklären auch nicht die Polen auch nicht die Deutschen (IL_PIL_GP)

Gelegentlich ist der Gebrauch auch bei derselben Sprecherin schwankend:

- (62) *für* *ihr* war die Sprache sehr schwierig (NH_PIL_GP)
- vs.
- (63) das war schwer *für* *mich* (NH_PIL_GP)

Sehr oft ist in den Äußerungen der Sprecherinnen und Sprecher aus Regionen, die historisch zum niederdeutschen Dialektraum gehörten, das Fehlen einer Unterscheidung zwischen der koordinativen Konjunktion *denn* und dem Adverb *dann* festzustellen. Dies rührt daher, dass sich eine Unterscheidung beider Formen erst in der Mitte des 18. Jahrhunderts festgesetzt hat und somit reflektieren die folgenden Belege eine frühere Entwicklungsperiode des Deutschen (vgl. Deutsches Wörterbuch (DWB) (2021) Wiktorowicz 2001: 112; Jorroch 2015: 215f.; Zielińska 2019: 111). Diese Bewahrung des älteren Sprachzustandes ist in den Sprachdaten sehr verbreitet:

- (64) ja anfangs wenn das die Kinder nicht sollten verstehen *denn* haben wir Deutsch jesprochen (IL_PIL_GP)
- (65) wenn ich jetzt noch mal jung wäre *denn* möchte ich bestimmt zehn Sprachen lernen (SB_KLA_GP)
- (66) und *denn* hatte mich hingesezt mit einem Junge aus der Ukraine der auch kein Polnisch kannte der sprach bloß Ukrainisch und konnte ganz gut Deutsch (KL_GDA_GP)
- (67) Friedel *dann* du geh mit mit die mit ganze Gruppe wo sie werden ibenachten *denn* (AL_PAR_GP)
- (68) waren wir fünf Jungens und drei Mädchen und *denn* bloß die zwei Schwestern hab ich noch in Deutschland da (AL_PAR_GP)

Auf der syntaktischen Ebene macht sich insbesondere im Bereich der Hypotaxe eine Abweichung vom Standarddeutschen bemerkbar, die Wegera (1983: 1485) per se als sprachliche Eigentümlichkeit von Dialektsprechern anführt, und die auf der Verwendung der Konjunktion *wie* in Temporalsätzen anstatt der standardsprachlichen Konjunktion *als* beruht.⁴ In dem zusammengestellten Korpus verwenden unsere Sprecherinnen und Sprecher aus dem ehemals niederdeutschen Dialektraum die Konjunktion *wie* in Nebensätzen temporaler Art sehr konsequent:

- (69) mein Vater war zweiundvierzig Jahre wie er gestorben ist (JT_BAR_GP)
- (70) früher haben wir noch Deutsch gesprochen wie die Deutschen noch alle hier waren (FL_BON_GP)
- (71) wie ich die Schmerzen schon bekommen habe ist der Arzt gekommen (LA_MIK_GP)
- (72) die Kinder machen das bisschen alles anders ganz anders ich hab geweint, wie wie meine Mutti gestorben war wie sie Begräbnis hat gehabt da habe ich geweint (FT_ELB_GP)
- (73) na gleich na wie die Polen na hier waren vor zwei Jahre die Russen <.> (AL_PAR_GP)

Neben den ansatzweise geschilderten Gemeinsamkeiten lassen sich in den gesammelten sprachlichen Daten ebenso regionalspezifische Unterschiede beobachten. Außer den bereits für Schlesien geschilderten Eigentümlichkeiten weisen auch unsere Sprecherinnen und Sprecher aus Pommern, West- und Ostpreußen einige Besonderheiten auf, die in anderen, in unserem Korpus repräsentierten Regionen nicht verzeichnet wurden.

Beispielsweise treten in den Sprachbiographien aus Pommern häufig Kontraktionen der Flexionsendungen an Verben mit auslautenden Verschlusslauten auf, etwa:

- (74) bis neunzehnhundertsiebenundvierzig haben wir alle bei den Russen *gearbeit* [‚gearbeitet‘] (HD_SWI_GP)
- (75) und denn hab ich noch eine Enkeltochter die wohnt hier jetzt in Posen sie *arbeit* [‚arbeitet‘] schon (NH_PIL_GP)
- (76) Weil er nicht *verheirat* [‚verheiratet‘] war und ich war *verheirat* [‚verheiratet‘] hatte Kinder (IL_PIL_GP)

Rein (1983: 1150) zufolge handelt es sich dabei um eine weit verbreitete dialektale Erscheinung. Ein solcher Lautzusammenfall ist beispielsweise auch für die oberschlesische Sprachinsel Kostenthal charakteristisch (vgl. Księżyk 2008: 134). Im Rahmen unseres Korpus tritt allerdings diese Besonderheit lediglich bei Sprecherinnen und Sprechern aus Pommern auf.

Dagegen konnte in Westpreußen beobachtet werden, dass bei der Konjugation der starken Verben der Ablaut nicht konsequent durchgeführt wird:

⁴ Insbesondere in Verbindung mit dem historischen Präsens leitet die Konjunktion *wie* auch in der Umgangssprache Temporalsätze ein.

(77) no eine Tochter die *sprecht* [ˌsprɪchtˈ] Deutsch (FT_ELB_GP)

(78) Oma *helf* [ˌhɪlfˈ] mich bisschen Oma *helf* [ˌhɪlfˈ] mich bisschen (FT_ELB_GP)

4 Polnische dialektale Merkmale in Schlesien

Die Interviews mit allen unseren Informantinnen und Informanten wurden auf Polnisch und Deutsch durchgeführt. Wir machten keine Aufnahmen des Schlesischen und verwenden daher nicht die für das Standardschlesische ausgearbeitete Orthographie, wie sie in „Kanōna szrajbowanio ślōnskij godki“ (2009) eingeführt wird.

Die polnischen Interviews der Informanten aus Schlesien wurden gemäß der polnischen Standardorthographie unter Beibehaltung der schlesischen Merkmale transkribiert. Die schlesischen Merkmale wurden auf der Grundlage der Studie von Halina Karaś (2010) ermittelt. In diesem Abschnitt stellen wir die wichtigsten dialektalen Merkmale in den Texten vor.

1. Die für das Schlesische typische Unterscheidung zwischen geneigten und hellen Vokalen⁵. Dieses Merkmal findet man in unseren Texten.

1.1. gedehntes [ä] erscheint als [ɔ] oder [ɔw̃] in den Formen:

– des Verbs für 3. Pers. Sg. mask. Präteritum:

(79) Jak dzień *mijol* [mʲ iɔw̃]; Standardpolnisch [mʲ ijaw̃] (KQ_GOG_GP)

– des Verbs 1. Pers. Sg. Präsens und Futur

(80) „Gościa“ *czytom* [ʧʲɪtɔm] [1. Sg. Präs.]; Standardpolnisch [ʧʲɪtam] (KQ_GOG_GP)

(81) A potym „Gościa“ *przeczytom* [pʃɛˈʧʲɪtɔm] [1. Sg. Fut.]; Standardpolnisch [pʃɛˈʧʲɪtam] (KQ_GOG_GP)

– des Verbs 2. Pers. Sg. im Präsens:

(82) Ale jeszcze *ogłondolsz* [ɔglɔndɔw̃ʃ]; Standardpolnisch [ɔglɔndaʃ] msza w języku niemieckim (KQ_GOG_GP)

⁵ Die geneigten Vokale [ä] [ɛ̃] [ɔ̃] haben eine engere Aussprache als die entsprechenden hellen Vokale. Es handelt sich dabei um eine Kontinuität der alten Langvokale. Nach dem Schwund der Vokalquantität, d. h. der Unterscheidung von langen und kurzen Vokalen je nach der Artikulationszeit, wurden die langen Vokale enger bei gleichzeitiger Verkürzung der Artikulationszeit.

1.2. weitere Beispiele für die Aussprache des geneigten [ä] als [ɔ]:

- (83) Zamek *godali* [gɔdali]; Standardpolnisch [gadal'i] na to *zomek* [zɔmek]; Standardpolnisch [zamek] (KQ_GOG_GP)
- (84) *Joł* [jɔw]; Standardpolnisch [ja] nie moga narzykać (KQ_GOG_GP)

1.3. gedehntes [ê] erscheint als [i], [ĩ]:

- (85) *Bo tyż* [tĩʃ]; Standardpolnisch [tɛʃ] kiedyś (KQ_GOG_GP)
- (86) *Musza pedzieć tako kobyta* [kɔb'ita]; Standardpolnisch [kɔb'ieta] rzeczywiście (SV_NAK_GP)

2. Die Entwicklung der alten Nasalvokale im Schlesischen ist anders als im Standardpolnischen.

2.1. Der auslautende Nasalvokal [ɔ̃] erscheint als [ɔw]:

- (87) *Jak malol* [mawɔw]; Standardpolnisch [mawɔw̃] była, *joł* miała bardzo dobre życie jak byłach dzieckiem (KQ_GOG_GP)
- (88) *No na świętol* [ɛv'ientɔw]; Standardpolnisch [ɛv'ientɔw̃] Anna jak odpust buł, ja była tyż *takol* [takɔw]; Standardpolnisch [takɔw̃] tradycja z Gogolina (KQ_GOG_GP)

2.2. Der Nasalvokal [ê] erscheint als [i], [ĩ]:

- (89) *ciyżko* [tɛĩkɔ]; Standardpolnisch [tɛɛw̃kɔ] się narobił (KQ_GOG_GP)
- (90) *Tyn kościół zastympczy* [zastɪmpɕĩ]; Standardpolnisch [zastɛmpɕĩ] to była staro szkoła (BT_KLO_GP)

2.3. Der Nasalvokal [ê] wird vollständig zu [a] entnasaliert in den Formen des Verbs 1. Sg. Präsens:

- (91) *Ale na ogół lubia* [lub' a]; Standardpolnisch [lub'ɛw̃] gotować (KQ_GOG_GP)
- (92) *Jo cie naucza* [nauɕa]; Standardpolnisch [nauɕɛw̃] (LU_KLO_GP)

3. Charakteristisch für das Schlesische ist die Aussprache von *trzi*, *trzimać* usw. – wie im Altpolnischen –, d. h. mit Beibehaltung des Vokals *i* nach dem aus palatalisiertem *r* entstandenen *rz*. Belege dafür sind in unseren Texten:

- (93) *A normalnie jak się warzi* [vazi]; Standardpolnisch [vazi] hmm, ale jakieś dwudaniowe czy jednodaniowe (KQ_GOG_GP)
- (94) *Tyn starszy zawsze się omy trzimoł* [tʃimɔw]; Standardpolnisch [tʃimaw] kiecki (BT_KLO_GP)

4. Aussprache *ul / oł* [uw/ɔw] statt *il* [iw] oder *yl* [iɥ]:
- (95) Jak odpust *bul* [buw]; Standardpolnisch [biw] (BT_KLO_GP)
- (96) Brał filiżankę wody łyknoł i już się *umol* [umɔw]; Standardpolnisch [umiw] (SV_NAK_GP)
5. Phonetische Vereinfachungen, zum Beispiel *pedział* [pɛdʑiaw] statt *powiedział* [pɔvʲɛdʑiaw]:
- (97) I ci *pedzieli* [pɛdʑɛli], co bydziesz ze szwabami mieszkał przyjedź tutaj (SV_NAK_GP)
6. Bildung des Präteritums von *iść*: *szoł* [ʃɔw] statt standardsprachlich *szedł* [ʃɛd]:
- (98) To *prziszol* [priʃɔw] stary Drozd (BT_KLO_GP)
7. Ein grammatisches Merkmal in den Texten ist die Bildung des Präteritums mit der alten Aoristendung *-ch*. Die Endung wird nicht an das Partizip Präteritum angefügt, sondern bleibt beweglich und wird mit der Partikel *że* verbunden:
- (99) A jak *zech* do szkoły *chodzila* [1. Sg. Prät.] (KQ_GOG_GP)
,Und als ich zur Schule ging.‘
- (100) Tam *zech* się *nauczyla* pływać [1. Sg. Prät.] (KQ_GOG_GP)
,Dort lernte ich schwimmen.‘
- (101) Nopród *zech musiol* [1. Sg. Prät.] dziesiynć <ni> dziesiynć albo sto razy napisac (SV_NAK_GP)
,Zuerst musste ich zehn oder hundertmal schreiben.‘
- (102) No to jo *zech byl* [1. Sg. Prät.] do wszystkiego (BT_KLO_GP)
,Nun war ich für alles da.‘
8. Formen des Präteritums mit Aoristendung *-ch*:
- (103) W każdym bądź razie nie *miolech* [1. Sg. Prät.] lekko (BT_KLO_GP)
,Jedenfalls hatte ich es nicht leicht.‘
9. Bildung des Konjunktivs mit Aoristendung *-ch* und der Partikel *że*:
- (104) Mówię jo *bych* se galoty *uszol* [1. Sg. Konj.] (LU_KLO_GP)
,Ich sage, ich würde mir eine Hose nähen.‘

Die schlesischen Merkmale treten selbst innerhalb eines Idiolekts nicht konsequent und mit unterschiedlicher Frequenz bei den einzelnen Sprechern auf.

5 Fazit

Angesichts der Tatsache, dass infolge des Zweiten Weltkriegs die deutschen Bewohnerinnen und Bewohner der an Polen gefallen Gebiete weitgehend neuen Bewohnerinnen und Bewohnern aus Zentralpolen oder den sog. *Kresy*, d. h. aus den ehemaligen ostpolnischen Gebieten, den Platz räumen mussten, wird die Beschäftigung mit deutschen Dialekten in Polen nunmehr als ein historischer Gegenstand betrachtet (vgl. Wiesinger 1988: 645). Die von uns interviewten Informantinnen und Informanten, mit denen wir auf sehr unterschiedliche Weise Kontakt aufgenommen haben, wie bei der praktischen Durchführung der Feldforschung von Joroch & Prawdzic in diesem Band beschrieben wurde, stellen somit die letzten Zeugen der historischen Sprachwirklichkeit dar. Die sprachbiographischen Interviews mit der älteren, in Polen sesshaft gebliebenen Generation, die im Rahmen des LangGener-Projekts durchgeführt wurden, dokumentieren Varietäten des Deutschen, die vom Aussterben bedroht sind. In die multimodalen LangGener-Korpora wurden insgesamt 30 Vertreterinnen und Vertreter der „Generation Polen“ aufgenommen: Neun von ihnen stammen aus den historischen Provinzen Ostpreußen, acht aus Schlesien, sieben aus Pommern und je drei aus Westpreußen bzw. der Grenzmark Posen-Westpreußen. Eine Zusammenstellung der beschriebenen Dialektmerkmale mit Angabe, wie viele Sprecherinnen und Sprecher aus den genannten Regionen diese Merkmale aufweisen, liefert die nachfolgende Tabelle.

Identifizierte Dialektmerkmale	Anzahl der Sprecherinnen und Sprecher im LangGener-Korpus, die das Merkmal realisieren
Entrundung von [y:] zu [i:] bzw. [ʏ] zu [ɪ]	7 (4 aus Schlesien, 3 aus Ostpreußen)
Entrundung von [ø:], [œ] zu [ɛ]	8 (5 aus Schlesien, 2 aus Ostpreußen, 1 aus Pommern)
Spirantisierung von [g] zu [x] nach hinteren Vokalen	5 (3 aus Ostpreußen, 1 aus Pommern, 1 aus Grenzmark Posen-Westpreußen)
Palatalisierung von [g] zu [j] im Anlaut	6 (3 aus Pommern, 2 aus Ostpreußen, 1 aus Grenzmark Posen-Westpreußen)
Zwischenlaut [gj]	2 (2 aus Ostpreußen)
unverschobenes [t], [k] oder [p] als Folge der im Niederdeutschen unterbliebenen zweiten Lautverschiebung	2 (2 aus Ostpreußen, 2 aus Pommern)
Realisierung des reduzierten Vokals [ɐ] als [ɛ]	7 (6 aus Schlesien, 1 aus Ostpreußen)
s-Palatalisierung nach [r]	4 (4 aus Schlesien)

Identifizierte Dialektmerkmale	Anzahl der Sprecherinnen und Sprecher im LangGener-Korpus, die das Merkmal realisieren
stimmhafte Realisierung des intervokalischen [f]	2 (2 aus Schlesien)
en-Apokope nach Nasallauten	5 (5 aus Schlesien)
Markierung des Dativs singularischer Substantive mit der Endung -e	5 (4 aus Schlesien, 1 aus Ostpreußen)
Schwa-Epithese bei prädikativ gebrauchten Zahladjektiven	3 (2 aus Schlesien, 1 aus Pommern)
Gebrauch des Akkusativs als Einheitskasus	9 (5 aus Pommern, 3 aus Ostpreußen, 1 aus Grenzmark Posen-Westpreußen)
Einheitsform <i>mi</i> infolge der fehlenden Dativ-/Akkusativ-Unterscheidung der Personalpronomina	2 (1 aus Pommern, 1 aus Ostpreußen)
standardabweichende Form <i>mir</i>	3 (2 aus Ostpreußen, 1 aus Pommern)
Schwankender Gebrauch von Akkusativ- und Dativformen bei Personalpronomina	5 (2 aus Pommern, 2 aus Grenzmark Posen-Westpreußen, 1 aus Ostpreußen)
Nicht-Unterscheidung zwischen <i>denn</i> und <i>dann</i>	13 (7 aus Ostpreußen, 3 aus Pommern, 2 aus Grenzmark Posen-Westpreußen, 1 aus Westpreußen)
Gebrauch der Konjunktion <i>wie</i> in Temporalsätzen anstelle der Konjunktion <i>als</i>	8 (4 aus Ostpreußen, 2 aus Pommern, 1 aus Schlesien, 1 aus Westpreußen)
Kontraktionen der Flexionsendungen an Verben mit auslautenden Verschlusslauten	5 (2 aus Schlesien, 1 aus Pommern, 1 aus Westpreußen, 1 aus Grenzmark Posen-Westpreußen)
Variable Ablaut-Realisierung	1 (1 aus Westpreußen)

Tabelle 1: Verbreitung der analysierten Dialektmerkmale bei den Sprecherinnen und Sprechern im LangGener-Korpus

Auch wenn die Vertreterinnen und Vertreter der älteren Generation („Generation Polen“) ihre Äußerungen heute relativ standardnah realisieren, offenbaren diese Interviews eine Reihe von restdialektalen Merkmalen, die als Dialektkontinuität zu betrachten sind und als immaterielles Kulturerbe einer besonderen Pflege bedürfen. Im vorliegenden Kapitel wurden lediglich die auffälligsten Merkmale besprochen, sie können allerdings einen

Anstoß für weitere Untersuchungen bilden, die das multimodale LangGener-Korpus dank der Zusammenstellung soziolinguistischer und korpuslinguistischer Fakten ermöglicht.

Literatur

- Barbour, Stephen/Stevenson, Patrick (2012): *Variation im Deutschen: Soziolinguistische Perspektiven*. Berlin; New York: De Gruyter.
- Bellmann, Günter (1971): *Slavoteutonica. Lexikalische Untersuchungen zum slawisch-deutschen Sprachkontakt im Ostmitteldeutschen*. Berlin/New York: De Gruyter.
- Friemel, Wilhelm (1937): Die Mundart von Oberglogau. In: *Schlesische Geschichtsblätter* 3. Breslau, 104–108.
- Graebisch, Friedrich (1929): Proben der deutschen Mundarten Oberschlesiens. In: *Der Oberschlesier*, 120–128.
- Gudaszewski, Grzegorz (2015): *Struktura narodowościowo-etniczna, językowa i wyznaniowa ludności Polski. Narodowy Spis Powszechny Ludności i Mieszkań 2011*. Warszawa: GUS. Online verfügbar unter: <https://stat.gov.pl/spisy-powszechny/nsp-2011/nsp-2011-wyniki/struktura-narodowo-etniczna-jezykowa-i-wyznaniowa-ludnosci-polski-nsp-2011,22,1.html>, letzter Zugriff am 10.12.2020.
- Jorroch, Anna (2015): *Die deutsche Sprache der dreisprachigen Altgläubigen in Masuren*. Warszawa: Warszawa: Instytut Germanistyki Uniwersytetu Warszawskiego.
- Jungandreas, Wolfgang (1937): *Zur Geschichte der schlesischen Mundart im Mittelalter*. Breslau: Universität zu Breslau.
- Kanōna szrajbowania ślōnskij godki (2009). In: *Ślůnsko Nacyjo* 3.8, 1–9.
- Karaś, Halina (Hg.) (2010): *Dialekty i gwary polskie. Kompendium internetowe*. Online verfügbar unter: <http://www.dialektologia.uw.edu.pl>, letzter Zugriff am 17.01.2022.
- Klapper, Joseph (1925): *Schlesische Volkskunde auf kulturgeschichtlicher Grundlage*. In: Theodor Siebs (Hg.), *Schlesisches Volkstum – Quellen und Arbeiten der Schlesischen Gesellschaft für Volkskunde*. Bd. 1. Breslau: Hirt, 56–59.
- Klein, Karl Kurt (1956): Hochsprache und Mundart in den deutschen Sprachinseln. *Zeitschrift für Mundartforschung* 24, 193–229.
- Kossert, Andreas (2001): *Masuren. Ostpreußens vergessener Süden*. 2. Aufl. München: Siedler.
- Księżyk, Felicja (2008): *Die deutsche Sprachinsel Kostenthal. Geschichte und Gegenwart*. Berlin: trafo Verlag.
- Lasatowicz, Maria Katarzyna (2002): Zur Dynamik sprachlicher Wirklichkeiten im Raum Oberschlesien. In: Edward Białek, Manfred Durzak, Marek Zybura (Hgg.), *Literatur im Zeugenstand: Beiträge zur deutschsprachigen Literatur- und Kulturgeschichte. Festschrift zum 65. Geburtstag von Hubert Orłowski*. Frankfurt/Main: Peter Lang, 773–788.
- Lasatowicz, Maria Katarzyna/Weger, Tobias (2008): Polen. In: Ludwig M. Eichinger, Albrecht Plewnia, Claudia M. Riehl (Hgg.), *Handbuch der deutsche Sprachminderheiten in Mittel- und Osteuropa*. Tübingen: Narr, 145–169.
- Lübben, August (1882): *Mittelniederdeutsche Grammatik: nebst Chrestomathie und Glossar*. Leipzig: Weigel.
- Mak, Wilhelm (1929): Die schlesischen Mundarten: Deutsche Mundarten in Oberschlesien. In: *Der Oberschlesier* 11(2), 1–86.
- Matuschek, Herbert (1998): Das Polnisch der Oberschlesier. Zu den Kontroversen um ein Idiom (Fortsetzung). In: Hans Ludwig Abmeier, Peter Chmiel, Nikolaus Gussone, Waldemar Zylla (Hgg.), *Oberschlesisches Jahrbuch*. Dülmen: Laumann, 193–213.
- Pelka, Daniela (2006): *Der deutsch-polnische Sprachkontakt in Oberschlesien am Beispiel der Gegend von Oberglogau*. Berlin: trafo Verlag.

- Pelka, Daniela (2007): Methode und Ergebnisse einer kontaktlinguistischen Untersuchung in Oberschlesien. *Germanoslavica. Zeitschrift für germano-slawische Studien*. 18. 1–2, 125–146.
- Porada, Haik Thomas (2015): Grenzmark Posen-Westpreußen. In: Online-Lexikon zur Kultur und Geschichte der Deutschen im östlichen Europa. Online verfügbar unter: ome-lexikon.uni-oldenburg.de/p32545, letzter Zugriff am 12.12.2021.
- Rein, Kurt (1983): Kontraktion in den deutschen Dialekten. In: Werner Besch, Ulrich Knoop, Wolfgang Putschke, Herbert Ernst Wiegand (Hgg.), *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Bd. 2. Berlin/New York, 1148–1121.
- Reiter, Norbert (1960): *Die polnisch-deutschen Sprachbeziehungen in Oberschlesien*. Wiesbaden: Harrassowitz.
- Rohfleisch, Irene (2001): *Sprachsituation und Sprachverhalten in Teilen des heutigen Oberschlesiens*. Heidelberg.
- Schwarz, Ernst (1935): *Sudetendeutsche Sprachräume*. München: Reinhardt.
- Stritzel, Herbert (1937): *Die Gliederung der Mundarten um Lauenburg in Pommern*. Marburg: Elwert.
- Tessmann, Wilhelm (1969): *Hochpreußisch und Schlesisch-Böhmisch-Mährisch mit den Sprachinseln des Südostens*. Uelzen.
- Unwerth, Wolfgang von (1908): Die schlesische Mundart in ihren Lautverhältnissen grammatisch und geographisch dargestellt. In: *Wort und Brauch*. 3. Heft. Breslau 1908.
- Unwerth, Wolfgang von (1911): Das Entwicklungsgebiet der schlesischen Mundart. In: Theodor Siebs (Hg.), *Festschrift zur Jahrhundertfeier der Universität Breslau, im Namen der Schlesischen Gesellschaft für Volkskunde*. Breslau: Marcus, 155–176.
- Unwerth, Wolfgang von (1913): Die schlesische Mundart. In: *Schlesisches Jahrbuch*, 99–104.
- Wegera, Klaus-Peter (1983): Probleme des Dialektsprechers beim Erwerb der deutschen Standardsprache. In: Werner Besch, Ulrich Knoop, Wolfgang Putschke, Herbert Ernst Wiegand (Hgg.), *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Bd. 2. Berlin/New York: De Gruyter, 1474–1492.
- Westpreußische Gesellschaft e.V. (2022): <https://westpreussische-gesellschaft.de/wirueber-uns/was-ist-westpreussen/>, letzter Zugriff am 12.12.2021. Westpreußisches Landesmuseum: Online verfügbar unter: <http://westpreussisches-landesmuseum.de/pl/muzeum/czym-sa-prusy-zachodnie/>, letzter Zugriff am 15.12.2020.
- Wiesinger, Peter (1983a): Die Einteilung der deutschen Dialekte. In: Werner Besch, Ulrich Knoop, Wolfgang Putschke, Herbert Ernst Wiegand (Hgg.), *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Bd. 2. Berlin/New York: De Gruyter, 801–900.
- Wiesinger, Peter (1983b): Deutsche Dialektgebiete außerhalb des deutschen Sprachgebiets: Mittel-, Südost- und Osteuropa. In: Werner Besch, Ulrich Knoop, Wolfgang Putschke, Herbert Ernst Wiegand (Hgg.), *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Bd. 2. Berlin/New York: De Gruyter, 900–927.
- Wiesinger, Peter (1988): Die Vokalentwicklungen des Niederländisch-Schlesischen in strukturgeographisch-strukturhistorischer Sicht. Eine Studie zur historischen Dialektologie. In: Jacek Fisiak (Hg.): *Historical Dialectology – Regional and Social*. Berlin u. a.: Mouton de Gruyter (Trends in Linguistics. Studies and Monographs. 37), 645–678.
- Wiktorowicz, Józef (2001): *Die Temporaladverbien im Frühneuhochdeutschen*. Teil 1: 1350–1500. Warszawa.
- Wolf, Norbert Richard (1983): Durchführung und Verbreitung der zweiten Lautverschiebung in den deutschen Dialekten. In: Werner Besch, Ulrich Knoop, Wolfgang Putschke, Herbert Ernst Wiegand (Hgg.): *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Bd. 2. Berlin: De Gruyter, 1116–1121.
- Zielińska, Anna (2019): *Grenzlandsprache. Untersuchung der Sprachen und Identitäten in der Region Lebus*. Berlin: Peter Lang.

Ziesemer, Walter (2009): *Die Ostpreußischen Mundarten*. Nachdr. d. Ausg. 1924, Vaduz: Sändig. URL 1: <https://woerterbuchnetz.de/?sigle=DWB&lemma=denn#0>, letzter Zugriff am 10.12.2020 Bitte nicht eintragen, sondern ganz unten im Literaturverzeichnis „URL 1: <https://woerterbuchnetz.de/?sigle=DWB&lemma=denn#0>, letzter Zugriff am 10.12.2020

Die LangGener-Korpora als multifunktionale Ressourcen der Mehrsprachigkeitsforschung zwischen Sozio- und Korpuslinguistik

Björn Hansen – Marek Nekula

Abstract:

Sociolinguistics meets corpus linguistics: The LangGener corpora as multifunctional resources for studies on multilingualism

This chapter introduces the binational research project „Language across Generations: Contact Induced Change in Morphosyntax in German-Polish Bilingual Speech“ (LangGener), funded by the German Research Foundation and the Polish National Science Centre. The aim of the project is an integrative account of the Polish-German bilingualism in Poland and Germany including linguistic as well as sociolinguistic aspects. The research focuses on the interplay between language biographies and speech of two generations of bilingual communities having experienced important historical events. The project includes a sub-project on Czech-German bilingualism in Germany, which concerns two groups with different profiles in terms of language ideologies.

Building on Baker (2010) we show that corpus linguistics and sociolinguistics in general share a number of fundamental tenets. Finally, we introduce and summarize potential research questions relating to the use of the LangGener corpora such the study of attitudes, superdiversity, input vs intake, and typology of pattern replications, as well as invite to share the data in the German-Polish and the Czech-German corpus in a wider research context.

Language contact remains conceptualized primarily between languages or schematic neighboring groups. The roles of bilingual (and even more so multilingual) speakers as individual agents and of the dynamic configurations of the multilingual societies in which they are actors have not been systematically taken into account. [...] Therefore, it is time to turn to the more concrete instantiations of language contact as created in the language use of bi- and multilingual individuals. (Lüpke 2016: 38)

1 Einführung

Das Projekt „Language across Generations: Contact Induced Change in Morphosyntax in German-Polish Bilingual Speech“ (LangGener), das ein deutsch-tschechisches Projekt

einschließt, ist an der Schnittstelle von zwei Disziplinen angesiedelt: Zum einen liegt der Fokus auf dem Sprachgebrauch bilingualer Individuen und den damit zusammenhängenden Effekten auf semantische und syntaktische Strukturen. Damit leistet das LangGener-Projekt einen spezifischen Beitrag zur Sprachkontaktforschung im Sinne der Spracherwerbs- und Spracherosionsforschung in Bezug auf Bilinguale. Zum anderen arbeitet das Projekt mit sprachbiographischen Daten, anhand derer soziale und sozialpsychologische Faktoren und damit auch Faktoren des Spracherwerbs und des Spracherhalts analysiert werden. Hier bewegt sich das LangGener-Projekt auf dem Gebiet der Soziolinguistik. Im Rahmen des Projekts entsteht eines der ersten Korpora, das konsequent das bilinguale Individuum in den Fokus stellt, indem es den Sprachgebrauch in beiden Sprachen, also Polnisch und Deutsch bzw. Tschechisch und Deutsch, erhebt, transkribiert, annotiert und die Ergebnisse mit weiteren Forscherinnen und Forschern teilt.

Wie bereits im Vorwort dieses Bandes erwähnt, ist die zentrale Leitfrage des gesamten Projekts, ob es im Bereich von kontaktbedingten Musterreplikationen zum einen Unterschiede zwischen den beiden jeweils untersuchten Generationen (mehr dazu siehe Centner in diesem Band) oder Gruppen (mehr dazu siehe Bučková & Nekula in diesem Band) und zum anderen Unterschiede zwischen den jeweiligen Sprachen gibt. Gibt es in der Sprache der Bilingualen die gleichen Typen von Musterreplikationen im Deutschen und im Polnischen bzw. Tschechischen? Damit hängt auch die Frage der Direktionalität der Replikationen (Deutsch ↔ Slavisch) zusammen. Ferner strebt das Projekt an, eine exhaustive Typologie von Musterreplikationen zu entwickeln. Diese Leitfragen haben die Art und Weise der Datenerhebung, der Datenverarbeitung und des Korpusbaus bestimmt. In diesem Kapitel wollen wir aufzeigen, dass die Korpora als multifunktionale Ressourcen für ein breites Anwendungsfeld nutzbar gemacht werden können.

Das Besondere des Projekts ist vor allem in der Methodologie zu sehen. Während viele Studien zu Effekten des Sprachkontakts unter Bedingungen der Mehrsprachigkeit mit experimentellen Methoden wie Tests und Datenelizitierung arbeiten, setzt das LangGener-Projekt auf natürliche sprachliche Daten aus den narrativen, sprachbiographisch orientierten Interviews und auf ihre korpuslinguistische Bearbeitung. Die transkribierten Interviews werden doppelt annotiert: sprachformbezogen in Bezug auf die sprachlichen Daten und soziolinguistisch in Bezug auf die metasprachlichen Daten, d.h. Aussagen zu Spracherwerb, Sprachgebrauch in unterschiedlichen Domänen oder Spracherleben, Sprachideologien und Sprachmanagement. Unsere eigens entwickelte Annotation von Sprachkontaktphänomenen zielt auf Replikationen semantischer und syntaktischer Muster. Die soziolinguistische Annotation, die aus Kapazitätsgründen nur im deutsch-polnischen Korpus durchgeführt wurde, fokussiert auf den individuell ausgeprägten, wenn auch sozial verankerten Spracherwerb, -gebrauch, -erhalt und -verlust. Die soziolinguistischen Variablen, die Auswirkungen auf die sprachkontaktbedingte Variation haben, sind aber sowohl im deutsch-polnischen als auch im deutsch-tschechischen Projekt in den Metadaten zum jeweiligen Transkript im Korpus enthalten. Dies erlaubt, dass beide Bereiche – sprachliche und metasprachliche Daten sowie soziolinguistische Metadaten – aufeinander bezogen und Ergebnisse verknüpft ausgewertet und interpretiert werden können.

Damit stellt sich auch die Frage nach der Verknüpfung der soziolinguistischen und korpuslinguistischen Forschung, der wir hier nachgehen werden. Diese Frage wollen wir programmatisch anhand ihrer Verknüpfung im LangGener-Projekt beantworten und uns darüber hinaus die Frage stellen, wie sich die soziolinguistischen Metadaten an sich und/

oder in Kombination mit sprachlichen Daten, die jeweils annotiert sind, korpuslinguistisch auswerten lassen und den Einblick in die Kovarianz zwischen den sozialen und sprachlichen Strukturen möglich machen. Die Korpora LangGener PL und CZ zählen dabei zu den sog. „small specialised“ (Koester 2010) bzw. „unconventional corpora“ (Kendall 2010: 363) mit einer manuellen tiefen Annotation, in denen sich die quantitative(n) Methode(n) nur begrenzt anwenden lassen.

2 Korpuslinguistik trifft Soziolinguistik

Die Parallelen zwischen korpuslinguistischen und soziolinguistischen Ansätzen beschreibt Baker (2010) folgendermaßen:

„First, corpus linguists and sociolinguists already share a number of fundamental tenets of best practice when it comes to linguistic analysis. Both approaches entail the collection and analysis of naturally occurring language data (as opposed to making introspective judgements about language use). [...] Secondly, both sociolinguistics and corpus linguistics make use of quantitative methodologies in order to carry out comparisons of different populations, focusing on differences and similarities, which can be facilitated with statistical tests. Thirdly, both approaches often use sampling techniques in order to be able to extrapolate claims to a wider population. Fourthly, both examine variation and change, and both consider a wide range of linguistic features (phonetics, morphology, lexis, grammar, discourse and pragmatics). And finally, sociolinguists and corpus linguists both attempt to provide explanations, where possible, for the findings that their research produces. In a sense, then, corpus linguistics and sociolinguistics overlap in terms of their epistemology, focus and scope.“ (Baker 2010: 8 f.)

2.1 Natürliche Daten

Die Korpora enthalten transkribierte Audioaufnahmen von bilingualen Personen, die Ergebnis von „freien“, genauer gesagt narrativen Interviews sind, die Sprachbiographien der Interviewten zum Gegenstand hatten. Die Aufnahmen enthalten auch spontane Gespräche, wie Tischgespräche über Themen wie Gesundheit, Kinder oder Kochen. Sie sind sehr gute Quellen für authentisch gesprochene Daten (Riehl 2014: 52). Bei den narrativen Interviews ist nach Bernamoun, Montrul & Polinsky (2013) oder Montrul (2008) – anders als bei den textbasierten Tests oder der Datenelizitierung – von der Absenkung der kognitiven Kontrolle bei der Sprachproduktion auszugehen. Durch die emotionale Involviertheit in Bezug auf das Thema (vgl. Busch 2017: 18f.) ist zudem zu erwarten, dass die kognitive Kontrolle weiter abgesenkt wird und authentische sprachliche Daten erzeugt werden. Aus demselben Grund wurde bei den Aufnahmen auf eine möglichst ungezwungene Atmosphäre geachtet; viele Interviews wurden bei den Interviewten zu Hause durchgeführt. Im Vorgespräch wurde das Format des sprachbiographischen Interviews erklärt und mit der Aufforderung verbunden, möglichst von sich aus die eigene Sprachbiographie zu erzählen, wonach der Erzählstimulus erfolgte und die Narration ausgelöst wurde (zum narrativen Interview und seiner Strukturierung vgl. etwa Küsters 2009 oder Lamnek & Krell 2016). Bei der Haupterzählung hielten sich die Interviewerinnen und Interviewer zurück. In der Nachfragephase wurden nach den immanenten

Fragen auch exmanente Fragen gestellt, die die Interviewer im Vorfeld in Form einer Liste vorbereitet hatten und deren Inhalte durch die Narration nicht behandelt wurden (zur Durchführung der Interviews siehe Joroch & Prawdzic in diesem Band). Damit steht das LangGener-Projekt im Kontrast zu streng kontrollierten elizitierten Korpusdaten zu Herkunftssprachen wie sie zum Beispiel in der „Research Unit Emerging Grammars in Language Contact Situations“ RUEG (Sprecherin Heike Wiese, HU zu Berlin) erhoben werden.¹ Bei dieser Forschungsgruppe werden die Respondentinnen und Respondenten gebeten, eine Filmszene eines Unfalls mehrmals wiederzugeben, einmal als Information an die Polizei und einmal an einen Freund. Dieses Verfahren soll Registerkompetenz erfassen, entspricht aber nicht dem Setting zur Erhebung authentischer Daten.

Durch die Sammlung und Analyse von natürlichen sprachlichen Daten unterscheiden sich die Sozio- und Korpuslinguistik von den Zugängen, die auf introspektive Urteile setzen. Sie betonen den natürlichen Sprachgebrauch im sozialen Kontext, den die Korpuslinguistik in der möglichst ausgewogenen Architektur allgemeiner Korpora reflektiert und in den Metadaten der Quellentexte dokumentiert sowie in spezialisierten Korpora fokussiert. In diesem Sinne verknüpfen sich die Soziolinguistik und die Korpuslinguistik auch in unseren spezialisierten deutsch-polnischen und deutsch-tschechischen Korpora, indem die auf sprachformbezogene und soziolinguistische Fragestellungen ausgerichteten Aufnahmen korpuslinguistisch aufbereitet werden. Die Metadaten enthalten soziolinguistische Informationen, im deutsch-tschechischen Korpus etwa die Unterscheidung *Spätaussiedlerinnen und Spätaussiedler* (SP) vs. *Migrantinnen und Migranten* (MI) (mehr dazu siehe Bučková & Nekula und zu den soziolinguistischen Metadaten generell siehe Bučková & Centner in diesem Band). Die sprachlichen Daten in den Transkripten werden mit Blick auf die Formen der sprachkontaktbedingten Variation manuell annotiert (zur sogenannten PAT-Annotation siehe Bučková, Centner, Księżyk & Prawdzic in diesem Band). Dadurch sind die für die Fragestellung des Projekts relevanten Phänomene im narrativen „Fluss“ natürlicher sprachlicher Daten fokussierbar und auch quantitativ auszuwerten und mit den soziolinguistischen Metadaten in Beziehung zu setzen. Die spezifisch ausgerichtete Annotation erlaubt jedoch nicht, die sprachkontaktbedingte Variation automatisch zu Formen, die davon unbetroffen sind, in Relation zu setzen; d.h. es sind nur ‚auffällige‘ Strukturen annotiert.

2.2 Datenquellen und sprachformbezogene quantitative Studien

Auch wenn es sich bei den LangGener-Korpora um kleine Korpora mit natürlichen Daten handelt, sind quantitative Analysen durchaus möglich. Bei der Erstellung der Subkorpora wurde stets auf Balanciertheit hinsichtlich der Anzahl der Respondentinnen und Respondenten und der Aufnahmedauer geachtet. Hier ist der Umfang der polnisch-deutschen Subkorpora:

¹ <http://doi.org/10.5281/zenodo.3236069>

	Anzahl der Respondentinnen und Respondenten	Dauer der Aufnahmen auf Polnisch	Dauer der Aufnahmen auf Deutsch
Generation Deutschland (GD)	29	17:19:39	17:55:51
Generation Polen (GP)	31	18:55:42	23:56:08

Tabelle 1: Umfang der deutsch-polnischen Subkorpora

Das Subkorpus GD ist mit seinen ca. 35 Stunden etwas symmetrischer, aber auch kleiner als das Subkorpus GP mit fast 43 Stunden.

Beim deutsch-tschechischen Projekt umfasst das Korpus neun Audioaufnahmen von Spätaussiedlerinnen und -aussiedlern (davon wurde ein Interview mit zwei Sprecherinnen und Sprechern durchgeführt) und sechs Audioaufnahmen mit tschechischen Migrantinnen und Migranten (davon wurde ein Interview mit zwei und eins mit vier Sprechern bzw. Sprecherinnen durchgeführt). So wurden insgesamt zwanzig Bilinguale ins Korpus aufgenommen:

	Anzahl der Respondentinnen und Respondenten	Dauer der Aufnahmen auf Tschechisch	Dauer der Aufnahmen auf Deutsch
Spätaussiedler/in (SP)	10	06:10:43	06:35:28
Migrant/in (MI)	10	07:52:15	06:53:28

Tabelle 2: Umfang der deutsch-tschechischen Subkorpora

Das deutsch-polnische Korpus ist nach drei Annotationsebenen durchsuchbar: Es gibt die syntaktische Ebene des Phrasentyps, die Ebene des Abweichtungstyps sowie die Ebene der soziolinguistischen Annotation; letztere enthält sprachbiographisch relevante Informationen zu Lebensphase, Domäne, Sprachideologien, Sprachmanagement und Sprachereleben (Näheres siehe Jańczak, Nekula & Zielińska in diesem Band). Damit erlaubt das Korpus quantitative Studien zu den beiden Sprecherpopulationen GD und GP bzw. der Gruppen von SP und MI im deutsch-tschechischen Korpus, wobei das deutsch-tschechische Korpus nicht sprachbiographisch annotiert ist. Die soziolinguistisch relevanten Angaben, wie Herkunfts- und Wohnort, Alter, Gender oder Bildung, sind im deutsch-tschechischen Korpus lediglich in den soziolinguistischen Metadaten der Transkripte der sprachbiographischen Interviews enthalten.

2.3 Externe Validität

Wie Baker (2010: 8) hervorhebt, streben beide Disziplinen nach einer gewissen Verallgemeinerungsfähigkeit; d. h. die Schlussfolgerungen, die auf der Basis der Daten aus dem LangGener-Korpus gewonnen werden, sollen sich auf die Gesamtheit der zwei Typen von Sprechergruppen, also die zwei Generationen beim deutsch-polnischen Teilprojekt und die Gruppen von SP und MI beim deutsch-tschechischen Projekt, verallgemeinern

lassen. Im deutsch-tschechischen Projekt unterscheiden sich die SP von den MI sowohl durch den unterschiedlichen sozialen Status vor und nach der Aussiedlung als auch durch Sprachideologien, Sprachmanagement und soziale Netzwerke (zu Sprachideologien und Sprachmanagement siehe auch Jańczak, Nekula & Zielińska in diesem Band), in denen der Sprachgebrauch von Deutsch und Tschechisch vor und nach der Aussiedlung im jeweiligen Sprachregime verankert ist, wobei bei einem Teil der SP der Spracherwerb des Deutschen in der Familie gegeben war (mehr dazu siehe Bučková & Nekula in diesem Band). Beim deutsch-polnischen Teilprojekt wurden bei der Auswahl der Respondentinnen und Respondenten feste Kriterien eines linguistisch verstandenen Generationsbegriffs angelegt (mehr dazu siehe Centner in diesem Band).

GP umfasst demnach Respondentinnen und Respondenten, die auf dem Gebiet des Deutschen Reichs geboren wurden, welches infolge des Zweiten Weltkriegs an Polen fiel. Definitorische Merkmale sind daher einerseits die Zweisprachigkeit der Informantinnen und Informanten und andererseits die Herkunftsregion. Die zwischen 1933 und 1945 geborenen und von diesem Projekt erfassten Respondentinnen und Respondenten wachsen auch nach 1945 zunächst in einer deutschsprachigen familiären Umgebung auf, ihre L1 ist Deutsch. Durch den politischen Wechsel um 1945 und die veränderten Rahmenbedingungen, die die Sprachpolitik und Sprachplanung auf der Makroebene verändern, erlernen auch sie noch vor Erreichen der sensitiven Periode, dem 12. Lebensjahr, das Polnische. Sie gelten damit als Sprecherinnen und Sprecher mit sukzessivem kindlichen (und schulischen) Zweitspracherwerb (cL2 Sprecherinnen und Sprecher). Sie haben in beiden Sprachen (Deutsch und Polnisch) kommunikative Kompetenzen, wobei sich diese in verschiedenen Domänen unterscheiden können. Die Respondentinnen und Respondenten weisen keine längeren Aufenthalte in Deutschland auf, da ein langer Aufenthalt in Deutschland die Variable des Polnischen als Umgebungssprache stark beeinflusst hätte.

GD umfasst Respondentinnen und Respondenten, die bereits nach dem Krieg in Polen geboren sind. Diese Personen sind Nachfahren der GP in dem Sinne, dass mindestens ein Elternteil dieser Personen bilingual (deutsch-polnisch) ist und aus den Gebieten stammt, welche seit 1945 zu Polen neu gehören. Eine Verwandtschaft zur im Projekt untersuchten GP muss nicht gegeben sein. Im Gegensatz zur GP ist die L1 der Respondentinnen und Respondenten aus der GD Polnisch. Obwohl die Eltern Deutsch können, erwerben die Respondentinnen und Respondenten der GP das Deutsche erst nach der Auswanderung nach Deutschland, die zumeist zwischen 1960 und 1990 erfolgt. Zum Zeitpunkt des Zweitspracherwerbs sind die Respondentinnen und Respondenten bereits erwachsen, sodass sie als Sprecherinnen und Sprecher mit Zweitspracherwerb im Erwachsenenalter (aL2 Sprecherinnen und Sprecher) gelten. Die Auswanderung aus Polen liegt zum Zeitpunkt der Untersuchung mindestens 10 Jahre zurück, sodass die Respondentinnen und Respondenten kommunikative Kompetenzen in beiden Sprachen aufweisen können. Diese Generation lebt über das gesamte Gebiet der Bundesrepublik verstreut, weswegen keine kleineren Gebiete wie bei Bezug auf das polnische Territorium für die Datenerhebung ausgemacht werden können. Durch die Abstammung von der GP, die territorial bestimmten Gebieten angehört, ergibt sich hier eine Parallele zwischen den Generationen. Dank der deutschen Abstammung durften die Respondentinnen und Respondenten der GD die deutsche Staatsbürgerschaft erlangen und weisen damit auch Parallelen zu den SP im deutsch-tschechischen Projekt auf.

Der Spracherwerb und die Sprachkompetenz, die Dominanz sowie die Distribution von Sprachen auf verschiedene Funktionsbereiche und Domänen oder die Intensität des Sprachkontakts sind dann Parameter bei der Herausarbeitung von Sprachbiographien. Die Validität der in den Sprachbiographien verankerten metasprachlichen Daten, die in soziolinguistischen Metadaten verdichtet werden, sowie ihre Erhärtung zu prototypischen Sprachbiographien der GP und der GD wird im Kapitel zur soziolinguistischen Annotation des deutsch-polnischen Korpus diskutiert (siehe Jańczak, Nekula & Zielińska in diesem Band).

Dabei orientieren sich die Respondentinnen und Respondenten sowie die Rekonstruktion ihrer deutsch oder polnisch dominierten Sprachregime an historischen Eckdaten. Gemeinsame Eckpfeiler sind auch die Zweisprachigkeit sowie eine territoriale Zugehörigkeit zu den Gebieten, die einst deutsch waren und nach 1945 an Polen fielen. So haben beide Generationen gemeinsam, dass es sich um deutsch-polnisch bilinguale Respondentinnen und Respondenten handelt, die kommunikative Kompetenzen in beiden Sprachen aufweisen. Die Erwerbsreihenfolge und das Erwerbsalter unterscheiden sich jedoch grundlegend. Im deutsch-tschechischen Teilprojekt bilden die Zweisprachigkeit und Sprachideologien die Eckpfeiler.

2.4 Ebenen der Variation: diatopisch, diastratisch und diaphasisch

Bestehend aus natürlichen Daten mit Äußerungen bilingualer Personen deckt das LangGener-Korpus alle drei Dimensionen sprachlicher Variation ab (im Sinne von Coseriu 1980). Die diatopische, also räumliche Dimension, ist in dem Korpus des deutsch-polnischen Projekts zu den beiden Generationen deutsch-polnischer Bilingualer dadurch erfasst, dass die Sprachdaten aus Regionen stammen, die vor 1945 zu Deutschland gehörten. Für das Projekt wurden Respondentinnen und Respondenten befragt, die aus Pommern, Masuren, Schlesien sowie den Gebieten um Piła und Stettin stammen. So enthalten die Metadaten zu jeder Respondentin bzw. jedem Respondenten Angaben zu Herkunfts- und Wohnregion. Bei der sprachformbezogenen Annotation und der späteren Analyse wurde ein besonderes Augenmerk auf dialektale Merkmale sowohl im Deutschen als auch im Polnischen gelegt. Dies war notwendig, um die Abweichungen von der Baseline angemessen zu annotieren. Als *baseline of comparison* dient die überregionale Umgangssprache des Deutschen bzw. Polnischen. Eine ähnliche Rolle spielt die diastratische, also soziale Dimension. So wurden zusätzlich zentrale biographische Variablen wie Geschlecht, Alter, Ausbildung und Beruf erhoben und in den soziolinguistischen Metadaten der Transkripte dokumentiert. Wie in dem Kapitel „Vom Gespräch zum Korpus“ dargelegt (siehe Bučková & Centner in diesem Band), enthalten die Metadaten Angaben zur Ausbildung, die sich an der Klassifikation von Ausbildungssystemen der International Standard Classification of Education orientieren. Zusätzlich annotieren wir außerhalb von ISCED als eigene Kategorie den auf einige Jahre beschränkten Schulbesuch der 1940er bis 1960er Jahre. Alle diese Kategorien sind neben dem deutsch-polnischen auch in dem deutsch-tschechischen Korpus enthalten. Damit sind in Zukunft Studien möglich, die bestimmte diastratische Variablen in den Fokus nehmen. Denkbar wäre ein Vergleich des Sprachgebrauchs von Männern und Frauen und Sprecherinnen und Sprechern mit unterschiedlichem Bildungsgrad.

Was die diaphasische Dimension betrifft, so geht es um einheitlich mündliche Sprachdaten. Bei der Transkription wurde darauf geachtet, die Daten möglichst exakt zu erfassen. Zwar wurde auf eine phonetische Umschrift verzichtet und die Standardorthographie als Basis der Transkription verwendet; es wurden jedoch alle Hässitationssignale und andere ‚Disfluencies‘ wie Abbrüche und Wiederholungen erfasst. Von einer ‚Korrektur‘ der Daten wurde weitgehend abgesehen. Aus unserer Sicht liegt die Stärke der LangGener-Korpora gerade in der adäquaten Abbildung der gesprochenen Sprache (Genaueres zur Transkription siehe Bučková & Prawdzc in diesem Band).

Das LangGener-Projekt sucht nicht nach mehr oder weniger homogenen Sprechergruppen, sondern bildet die Diversität in der Sprachkompetenz in ihrer ganzen Spannweite ab. Dabei handelt es sich bei GP um Herkunftssprecherinnen und -sprecher des Deutschen und bei GD um Spätbilinguale mit deutschem Hintergrund. Im deutsch-tschechischen Korpus geht es um Spätbilinguale mit und ohne deutschen Hintergrund. Polinsky (2018: 5) attestiert Herkunftssprecherinnen und -sprechern „bewildering variation in proficiency“. Das gleiche lässt sich auch über Spätbilinguale sagen.

2.5 Reihe an linguistischen Merkmalen

Das LangGener-Projekt legt einen Fokus auf die Morphosyntax des Deutschen und des Polnischen bzw. des Tschechischen. Während andere Studien sich auf einzelne Kategorien konzentrieren (zum Beispiel nur Verbalaspekt), enthält das LangGener-Korpus Annotationen für Phänomene auf unterschiedlichen syntaktischen Ebenen, d. h. die „auffälligen“ Strukturen sind hinsichtlich ihrer Distribution auf Phrasentypen annotiert (mehr dazu siehe Bučková, Centner, Księżyk & Prawdzc in diesem Band). Es werden je nach auftretendem Phänomen folgende sechs Phrasentypen (hier in eckigen Klammern markiert) annotiert (glossierte Beispiele aus dem Transkript BK_HAM_GD):

Nominalphrase: fehlender Artikel

(1) das war für mich [ganz fremde Sprache]

das	war	für	mich	[ganz	fremde	Sprache] _{NP}
das	sein.PST	für	mich	ganz	fremd.NOM.F	Sprache

das war für mich eine ganz fremde Sprache

Verbalphrase: Verbbedeutung

(2) obwohl sie [sehr gut Deutsch kannte] <.>

obwohl	sie	[sehr	gut	Deutsch	kannte] _{VP} <.>
obwohl	sie	sehr	gut	Deutsch	kenn.PST.SG

obwohl sie sehr gut Deutsch konnte

Adjektivphrase: Kongruenz

(3) aber die [andere beide] Kinder haben ke<*> kein Kontakt

aber	die	[andere	beide] _{AP}	Kinder	haben	ke<*> kein	Kontakt
aber	die	ander.PL	beid.PL	Kind.PL	hab.PRS	keinen	Kontakt

aber die anderen Kinder haben keinen Kontakt

Präpositionalphrase: Rektion

- (4) <@> bo <.> tacy znajomi odradzili mi żebym poszedł na kurs zorganizowany [z arbeitsamtu]

<@> bo	<.> tacy	znajomi	odradzili	mi	żebym	poszedł
denn	solch.NOM.PL	bekannt.M.PL	abrat.PST.M.PL	mir	COMP.COND.1.SG	geh.PST

na	kurs	zorganizowany	[z	arbeitsamtu] _{pp}
auf	Kurs	organisier.PTC.PASS.M.SG	von	Arbeitsamt.GEN.SG

denn diese Bekannten haben mir davon abgeraten, einen vom Arbeitsamt organisierten Kurs zu besuchen

Adverbialphrase: Wortform des Adverbs

- (5) ich wollte damals <@> Geld verdienen [hauptsätzlich]

ich	wollte	damals	<@> Geld	verdienen	[hauptsätzlich] _{AdvP}
ich	will.PST.1SG	damals	Geld	verdien.INF	hauptsächlich

ich wollte damals hauptsächlich verdienen

Satz: Wortfolge

- (6) <@> [nach dem Studium ich wollte einfach keine keine Zeit in der Armee verbringen]

<@> [nach	dem	Studium	ich	wollte	einfach
nach	das.DAT	Studium	ich	will.PST.1SG	einfach

kein keine	Zeit	in	der	Armee	verbringen] _s
kein	Zeit	in	die.DAT	Armee	verbringen

nach dem Studium wollte ich einfach keine Zeit in der Armee verbringen

Auf einer weiteren Annotationsebene wurden die Daten hinsichtlich sog. Abweichungen von den Varianten annotiert, die wir als Heimatland-Baseline des Deutschen, Polnischen und Tschechischen definiert haben (auch zu dieser Annotation siehe im Detail Bučková, Centner, Książek & Prawdzic in diesem Band). Auf der Ebene der sog. Abweichungen spielen sowohl formale Abweichungen wie die Auslassung des Artikels in (1) als auch rein semantische Musterreplikationen wie die Polysemiekopierung eine Rolle wie in (2) (PAT). Neben den Materialreplikationen wie in (4) (MAT) werden allerdings auch andere Abweichungen von der Heimatland-Baseline annotiert, die nicht als Kontaktergebnis interpretiert werden können: die Adjektivendung in (3) (AA), die im Baseline-Deutschen nicht-existente lexikalische Form *hauptsätzlich* in (5) (AA) und die Stellung des finiten Verbs an dritter Stelle in (6) (WO). Von besonderem Interesse sind Fälle, bei denen sowohl PAT als auch MAT vorliegt wie in der Verbalphrase *przjdzie dran* in (7):

- (7) kiedy on przjdzie dran <.> nie <?> (BT_KOL_GP)

kiedy	on	[przjdzie	dran] _{vp} <.>	nie <?>
wenn	er	komm.PRS.3SG	dran	nicht

wenn er drankommt, ne?

Hierbei handelt es sich zum einen um eine Musterreplikation nach dem deutschen Vorbild des deutschen Verbs mit trennbarem Präfix *drankommen* und zum anderen um eine Replikation der Form *dran*. Zudem enthält der Annotationsapparat das Merkmal Code-switching. Mit diesem reichen Annotationsapparat kann das LangGener-Korpus als Basis für Studien einerseits zu verschiedenen morphosyntaktischen und lexikalischen Kategorien und andererseits zu Kontakt- und Isolationsphänomenen dienen, die dann in einem zweiten Schritt mit weiteren Datensets aus experimentellen Methoden gezielt erforscht werden können.

3 Andere kleine Korpora zur Zweisprachigkeit

Im Folgenden gehen wir kurz auf vorhandene spezialisierte Korpora ein, die in Bezug auf die Zweisprachigkeit und den Sprachkontakt von Interesse sind. Dabei beschränken wir uns auf Polnisch und Tschechisch. In dieser Kurzdarstellung geht uns in Bezug auf die bereits vorhandenen Korpora darum, die Spezifika der LangGener-Korpora herauszustellen und die für die weitere Forschung relevanten Schnittmengen aufzuzeigen.

3.1 Polnisch

Zur deutsch-polnischen Zweisprachigkeit gibt es lediglich ein weiteres Korpus: das „Hamburg Corpus of Polish in Germany“ HamCoPoliG (folgende Darstellung nach Czachór 2012). Das Korpus ist im Kontext des DFG-Projekts „Current Polish-German Bilingualism in Germany“ entstanden, das einen Fokus auf kontaktinduzierte Wandelprozesse in der polnischen Morphosyntax legte. Es enthält Audioaufnahmen von bilingualen Sprecherinnen und Sprechern des Deutschen und Polnischen sowie von polnisch monolingualen Erwachsenen im Alter von 16 bis 46 Jahren. Aufgenommen wurden Personen aus drei Gruppen: 1. Frühbilinguale, die Polnisch als L1 erworben, aber keinen formalen Unterricht genossen haben (*heritage speakers*); 2. Spätbilinguale, die erst in einem Alter von 16 Jahren oder später nach Deutschland emigriert sind (*attriters*) und 3. monolinguale Kontrollgruppe aus Polen. Dabei strebte man nach maximal homogenen Sprechergruppen. Es sind Aufnahmen von semi-spontanen Daten zu den drei Themen Urlaub, Tagesablauf mit Fahrt zur Arbeit und die Welt im Jahre 3000 sowie die Nacherzählung einer Bildergeschichte von E. O. Plauen. Das Korpus umfasst insgesamt 2 264 Minuten von 93 Sprecherinnen und Sprechern (64 Frauen, 29 Männer) im Umfang von 294 663 Wörtern.

Das Projekt „Current Polish-German Bilingualism in Germany“, das das Korpus HamCoPoliG hervorgebracht hat, hat LangGener in vielerlei Hinsicht inspiriert. In beiden Projekten wurden systematisch polnische Sprachdaten von Bilingualen in Deutschland erhoben und ein orthographiebasiertes Transkriptionssystem sowie EXMARaLDA verwendet. HamCoPoliG ist eine Quelle zusätzlicher Evidenz, um Typen von Musterreplikationen im Polnischen nachzuweisen. Das LangGener-Projekt baut auf diesen Daten auf und erweitert den Blickwinkel in Richtung zweisprachiger Sprecherinnen und Sprecher, indem es sowohl das Polnische als auch das Deutsche desselben Respondent bzw. derselben Respondentin aufnimmt und analysiert. Das LangGener-Projekt hat auch dadurch einen breiteren Fokus, da es Daten bilingualer Sprecherinnen und Sprecher sowohl in

Deutschland als auch in Polen enthält. Das LangGener-Korpus ist somit eine Ergänzung zu HamCoPoliG.

3.2 Tschechisch

Im Rahmen des Tschechischen Nationalkorpus gibt es neben den allgemeinen Korpora auch eine Reihe von spezialisierten Korpora. Zu diesen zählen auch drei Akquisitionskorpora, die Texte enthalten, die von Schülern produziert wurden, die Tschechisch als Zweitsprache erwerben, wobei der Zweitspracherwerb sowohl gesteuert als auch ungesteuert und innerhalb Tschechiens erfolgt. Im Korpus *CzeSL-Plain* (A Learner Corpus of Czech) werden „Fehler“ in dieser Lernervarietät differenziert manuell annotiert, im Korpus *CzeSL-SGT* (Czech as a Second Language with Spelling, Grammar and Tag) werden sie automatisch annotiert. Die Fragen der Normabweichung und der automatischen Annotation werden in Štindlová (2013), Rosen et al. (2020) und bei Kotková (2017) mit Blick auf die Fehlerlinguistik kritisch diskutiert. Für den Sprachkontakt ist auch ein spezialisiertes Korpus von spontan gesprochenem Englisch fortgeschrittener Englischlerner mit Erstsprache Tschechisch relevant, das ebenfalls im Tschechischen Nationalkorpus zu finden ist. Zu erwähnen sind auch das Schüler-Korpus AKCES, das in Kooperation der TU Liberec und FF Prag entstanden ist (akces.ff.cuni.cz/node/155) sowie das Korpus MERLIN (UK Prag, Europäische Akademie Bozen und TU Dresden), das Texte enthält, die im Rahmen von standardisierten Sprachtests entstanden sind (www.merlin-platform.eu), wobei die Lernervarietäten des Tschechischen etwa in Nekula & Šichová (2017) eingeordnet werden.

Das spezialisierte Korpus der Transkripte von deutsch-tschechischen sprachbiographischen Interviews, das aus dem LangGener-Projekt hervorgegangen ist, ist inhaltlich in anderer Weise ausgerichtet und räumlich in anderer Weise verankert. Es geht darin u.a. um den Spracherwerb von Deutsch als Zweitsprache und die Attrition von Tschechisch in Deutschland. Die manuelle Annotation von Abweichungen von der jeweiligen Baseline macht sie allerdings etwa für Vergleichsstudien in Bezug auf den Zweitspracherwerb des Deutschen und Tschechischen interessant, wobei die Sprachkontaktphänomene beim unvollständigen Spracherwerb und der Attrition zum Teil dieselben sind. Die Aufnahmen der Interviews wurden nach Regeln transkribiert, die im Rahmen des Tschechischen Nationalkorpus für die Transkription von Audioaufnahmen genutzt werden. Bei der Transkription wurde bei den tschechischen Aufnahmen sogar auf Transkribentinnen und Transkribenten zurückgegriffen, die für das Tschechische Nationalkorpus Audioaufnahmen transkribieren, und zwar mit dem Ziel, die aufbereiteten Daten in einer technisch weit entwickelten und rezipierten Umgebung mit anderen Forscherinnen und Forschern zu teilen. Den Transkripten sind die Audioaufnahmen in segmentierter Form zugeordnet. Bei den Transkripten gibt es die üblichen Suchoptionen (Lemma-Suche, Kollokationen usw.), die oben skizzierte Annotation ist eingeschlossen. Die Verfügbarmachung des deutsch-tschechischen Korpus halten wir nicht nur für einen wichtigen Beitrag zur Transparenz der Forschung, sondern auch für einen Beitrag zum Data Sharing, durch das die Daten auch von anderen Forschenden für unterschiedlich ausgerichtete komparative oder im Kontext der Spracherwerbsforschung sprachformbezogene Studien genutzt werden können.

4 Ausgewählte Forschungsfragen

Das LangGener-Korpus ist ein Datenbestand von sprachlichen und metasprachlichen Daten sowie soziolinguistischen Metadaten, der für eine ganze Reihe von Studien verwendet werden kann. Im Folgenden seien einige Forschungsperspektiven genannt.

4.1 Einstellungen

Von zentraler Bedeutung für die Soziolinguistik und Kontaktlinguistik ist der Begriff der Einstellung, „a learned disposition to think, feel and behave toward a person (or object) in a particular way“ (Allport 1954, zitiert nach Garret 2010: 19). Im linguistischen Kontext geht es um Spracheinstellungen (language attitudes), d. h. um Einstellungen sowohl gegenüber Varietäten und Sprachen als auch ihren Sprecherinnen und Sprechern, bzw. auch gegenüber Sprachgemeinschaften und Sprechergruppen. Diese konkretisieren sich im episodischen Erzählen der narrativen sprachbiographischen Interviews im Spracherleben und in explizit werdenden Sprachideologien (zur Unterscheidung vgl. Busch 2017), mit anderen Worten in der emotionalen und kognitiven Einstellung von Bilingualen gegenüber Varietäten und Sprachen und Sprechergruppen, die sich in Bezug auf konkrete Erlebnisse und Ereignisse manifestieren und durch die das individuelle Sprachmanagement bzw. das Familienmanagement in Bezug auf den Spracherwerb, -gebrauch, -erhalt und -verlust argumentiert wird (zum Sprachmanagement siehe mehr in Jańczak, Nekula & Zielińska in diesem Band).

Nach Garret (2010) unterscheidet man in Bezug auf die Spracheinstellungen zwischen direkten und indirekten Methoden; im ersten Falle werden Respondentinnen und Respondenten direkt nach ihren Bewertungen von Sprachen befragt, im zweiten Fall werden solche Einstellung über *matched-guise technique* erhoben. Dabei werden die Respondentinnen und Respondenten über die Forschungsfrage im Unklaren gelassen. In der unterschiedlich ausgeprägten Bewertung vorgespielter Leseproben in Bezug auf das Erscheinungsbild und Charakter der Sprecherinnen und Sprecher artikulieren sich die Sprachattitüden in Bezug auf die untersuchten Sprechergruppen und Sprachen. Indem die Leseproben durch Bilinguale eingelesen werden, ist die unterschiedliche Einschätzung der Respondentinnen und Respondenten in Bezug auf dieselben Sprecherinnen und Sprecher nicht etwa auf die Stimme, sondern auf die Sprache zurückzuführen.

Die sprachbiographische Forschung lässt die Respondentinnen und Respondenten in diesem Punkt ebenfalls im Unklaren, indem die Spracheinstellungen nicht explizit elizitiert und die Zielsetzung der sprachbiographischen Forschung den Respondentinnen und Respondenten nicht bewusst gemacht werden. Dennoch werden die Sprachattitüden durch das Thema und die persönliche Betroffenheit von Bilingualen in Bezug auf den Spracherwerb, -gebrauch, -erhalt und -verlust gerade in einem durch die Sprache der Majorität dominierten Sprachregime ausgelöst. Dabei wird die Perspektive gewissermaßen umgedreht. Nicht die Sprachattitüden einer Mehrheitsgesellschaft in Bezug auf die bilinguale Minorität, sondern die der Bilingualen rücken in Fokus. Dabei schließen die sprachbiographischen Interviews die Sprachattitüden der Majorität in Bezug auf die bilinguale Minorität mit ein. Zudem binden sie diese Attitüden an konkrete Gebrauchskontexte und schließen den Wandel von Einstellungen mit ein. Dadurch bieten sie eine von

den Akteuren selbst ausgehende Differenzierung an, die man bei ihrer von den Forschenden ausgehenden Elizitierung nicht erreichen kann, auch weil sie durch die Annahmen der Forschenden verstellt ist. Durch die soziolinguistische Annotation des deutsch-polnischen Korpus in Bezug auf Lebensphase, Domänen, Spracherleben, Sprachideologien und Sprachmanagement, wobei sich diese verschränken (siehe im Detail Jańczak, Nekula & Zielińska in diesem Band), stehen die in diesem Korpus erfassten sprachbiographischen Narrationen auch für die Erforschung von Sprachattitüden in ihrer Variation und ihrem Wandel zur Verfügung.

4.2 Superdiversity

Viele Ansätze zur Kontaktlinguistik streben nach homogenen Datensätzen und gehen quasi von homogenen Sprechergruppen aus. Dies trifft auf systemorientierte Ansätze wie zum Beispiel die Interface-Hypothese zu, die besagt, dass „language structures involving an interface between syntax and other cognitive domains are less likely to be acquired completely than structures that do not involve this interface“ (Sorace 2011: 1). Diese Hypothese ist auch auf den bilingualen Spracherwerb übertragen worden. Der Ansatz der Superdiversity hingegen wendet sich gegen das methodologische Gebot der Homogenität von Sprecherdaten:

„Over a period of several decades – and often emerging in response to issues predating superdiversity – there has been ongoing revision of fundamental ideas (a) about languages, (b) about language groups and speakers, and (c) about communication. Rather than working with homogeneity, stability and boundedness as the starting assumptions, mobility, mixing, political dynamics and historical embedding are now central concerns in the study of languages, language groups and communication.“ (Blommaert et al. 2011: 3)

Da unsere Daten – wie oben dargestellt – aus recht natürlichen Kontexten stammen und wir sowohl in der GP als auch in der GD sowie in den beiden untersuchten Gruppen im deutsch-tschechischen Korpus ein breites Spektrum an sprachlichen Kompetenzen vorgefunden haben, denken wir, dass das LangGener-Korpus auch für Studien im Rahmen des Ansatzes der Superdiversity Verwendung finden kann. Unsere soziolinguistische Annotation im deutsch-polnischen Korpus und die soziolinguistischen Metadaten für beide Korpora geben Aufschluss über die Mobilität der Respondentinnen und Respondenten; die Frage der ‚Ausreise‘ zieht sich wie ein roter Faden durch alle Interviews. Gleichermäßen wurde die Wahrnehmung politischer Wandelprozesse in ihrer historischen Einbettung thematisiert.

4.3 Input vs. Intake

Wie Polinsky (2018: 28ff.) hervorhebt, unterscheiden sich Herkunftssprecherinnen und -sprecher sehr stark im Grad der Beherrschung des Sprachsystems; diese Variation trifft man in gleichem Maße unter Zweitsprachesprecherinnen und -sprechern. Polinsky (2018: 29) macht für diese Diversität ganz wesentlich den Umgang mit dem sprachlichen Input und die Art des sprachlichen Inputs in der Erwerbsphase verantwortlich. Sie identifiziert

gewisse Schwelleneffekte (threshold effects), wonach ab einem bestimmten Punkt der Input keine Rolle mehr spielt. Wir sind davon überzeugt, dass sich in den LangGener-Daten solche Effekte über die Analyse von sprachbiografischen Daten aufspüren lassen, in denen der individuelle Spracherwerb und Sprachgebrauch in seinem sozialen Umfeld in einzelnen Lebensphasen erfasst wird.

Polinsky (2018: 31) unterscheidet ferner zwischen Input und Intake. Während ersterer jegliche Form von sprachlichen Äußerungen, denen ein Bilingualer ausgesetzt ist, bezeichnet, geht es beim Intake um „the language that is actually internalized by a learner and is then used by that learner to form hypotheses about structure“. Polinsky weist darauf hin, dass hochfrequente Strukturen nicht unbedingt zum Intake zählen, da sie der Aufmerksamkeit der bilingualen Sprecherinnen und Sprecher beim Zuhören entgehen. Solche Effekte haben wir vor allem in den deutschen sprachlichen Daten der Spätbilingualen der GD gefunden: So finden sich dort sehr viele Abweichungen in der Deklination der Adjektive wie zum Beispiel in:

(8) weil mein Vater <@> ist mit der deutsche Sprache groß geworden (OD_REG_GD)

weil	mein	Vater	<@> ist	mit	der
weil	mein	Vater	sei.PRS.3SG	mit	die.DAT

deutsche	Sprache	groß	geworden
deutsch.NOM	Sprache	groß	werd.PTCP

weil mein Vater mit der deutschen Sprache groß geworden ist

(9) <.> wir <@> sollten die westdeutsche Zeitungen lesen <.> verschiedene (XX_HAM_GD)

<.> wir	<@> sollten	die	westdeutsche	Zeitungen
wir	soll.PST.3PL	die	westdeutsch.PL.INDEF	Zeitung.PL

lesen	<.> verschiedene
les.INF	verschieden.PL.INDEF

wir sollten die westdeutschen Zeitungen lesen

4.4 Typologie von Musterentlehnungen und soziolinguistische Settings

Der Ausgangspunkt des LangGener-Projekts war die Frage nach der Validität der Typologie von Musterentlehnungen (PAT typology), die sich an Heine & Kuteva (2005) und Matras & Sakel (2007) orientiert und in Wiemer, Wälchli & Hansen (2012) auf slavischsprachige Daten ausgerichtet wurde. Diese Typologie unterscheidet (a) sprachkontaktinduzierte Grammatikalisierung, (b) Polysemiekopierung, (c) Restrukturierung sowie (d) Verlust grammatischer Muster. Dabei wurde vorausgesetzt, dass die genannten Typen in – soziolinguistisch gesehen – unterschiedlichen Settings festzustellen sein würden, d.h. sowohl bei Bilingualen mit Deutsch als Erst- und Polnisch als Zweitsprache in Polen als auch bei Bilingualen mit Polnisch als Erst- und Deutsch als Zweitsprache in Deutschland. Die Typen an Musterentlehnungen weisen allerdings einzelsprachspezifische Besonder-

heiten auf, denn etwa der Aspekt kann durch Lexikalisierung lediglich im Polnischen oder Tschechischen als grammatisches Muster verlorengehen. Das deutsch-tschechische Teilprojekt sollte die im deutsch-polnischen Projekt angewandte Typologie mit Blick auf die angesprochenen sprachspezifischen Konkretisierungen in einer vergleichbaren Sprachenkombination validieren.

Allerdings ging es darin weniger um die Richtung der Musterreplikation von Deutsch zu Polnisch in der GP und von Polnisch zu Deutsch in der GD, sondern vielmehr um die Frage, ob und wie diese Muster in demselben soziolinguistischen Setting vorliegen, und damit auch um die Frage, ob und zu welchem Ausmaß dieselben Typen von Musterentlehnungen und deren Konkretisierungen sich bei Spracherosion der Erstsprache und dem unvollständigen Spracherwerb der Zweitsprache identifizieren lassen. Die Aufarbeitung des Materials im doppelt annotierten deutsch-polnischen Korpus, in dem sowohl die konkreten Abweichungsarten sprachformbezogen als auch soziolinguistisch annotiert werden, liefert dafür entsprechende Daten, die eine feinkörnige und doch auch quantitativ verankerte Analyse erlauben (im deutsch-tschechischen ist der soziolinguistische Rahmen nur durch Metadaten erfasst). Die Frage nach der Anwendbarkeit dieser oben erwähnten Typologie sowie nach der Vergleichbarkeit von deren sprachspezifischen Konkretisierungen ließe sich auch auf sprachliche Daten übertragen, die in anderen soziolinguistischen Settings als den herkunftssprachlichen erhoben wurden (Montrul 2008; Polinsky 2018). Hier zeichnet sich die Möglichkeit einer feineren Einbindung der soziolinguistischen Settings in die variationslinguistische Analyse ab.

5 Ausblick: Data Sharing in weiteren Forschungskontexten

Die im Rahmen des LangGener-Projekts erstellten Korpora liefern wichtige Daten auch jenseits der eng gefassten Frage nach der Wechselwirkung zwischen der korpuslinguistisch basierten Analyse von sprachformbezogen annotierten sprachlichen Daten und den soziolinguistisch annotierten metasprachlichen Daten im deutsch-polnischen Korpus (bzw. den soziolinguistischen Metadaten im deutsch-tschechischen und deutsch-polnischen Korpus). Im deutsch-polnischen Korpus sind durch die soziolinguistische Annotation die besten Voraussetzungen für die Herausarbeitung von prototypischen Sprachbiographien der GP und der GD gegeben, die einen wichtigen Beitrag leisten können bei der Rekonstruktion von mitteleuropäischen Sprachbiographien (mehr dazu siehe in Jańczak, Nekula & Zielińska in diesem Band) und damit auch der Rekonstruktion der Sprachsituation von Sprachgruppen in sozial und soziolinguistisch prekären Verhältnissen in Mitteleuropa nach dem Zweiten Weltkrieg. Für das deutsch-tschechische Projekt bietet sich hier eine Korrelation mit den Interviews an, die im Rahmen der soziolinguistischen und anthropologischen sprachbiographischen Forschung erhoben, dokumentiert und interpretiert wurden (vgl. etwa Nekvapil 1998, 2001, 2003, 2004 oder Kreisslová 2013, 2014).

Im deutsch-polnischen Korpus wird durch seine soziolinguistische Annotation der Weg aufgezeigt, wie die Rekonstruktion von (prototypischen) Sprachbiographien und die Erfassung von sprachbiographischen Narrativen als ein Teil der Textrealität in den sprachbiographischen Interviews operationalisiert werden kann (dazu vgl. u. a. auch Pavlenko 2007). Damit leisten die Korpora und das LangGener-Projekt einen Beitrag zur *oral history* nicht nur durch die digital verfügbaren Transkripte sprachbiographischer Interviews,

in denen sich die Wechselwirkung von Sprache und Gesellschaft auf der Mikroebene offenbart. Durch ihre Annotation, die Operationalisierung bei der Rekonstruktion von Sprachbiographien und die differenzierte Diskussion von Lebens-, Subjekt- und Textrealität in den metasprachlichen Daten leisten sie dies auch methodisch.

Nicht zu vergessen ist schließlich das dokumentarische Ziel des deutsch-polnischen Projekts, das die Sprache der Respondentinnen und Respondenten der GP, deren Sprachvarietät durch die ausgebliebene oder fragmentarische Weitergabe an die nächste Generation vom Aussterben bedroht ist, für die weitere Forschung konservieren soll. Davon kann vor allem die deutsche Dialektologie profitieren, die auf die Daten im deutsch-polnischen Korpus zurückgreifen und diese für ihre Zwecke – etwa mit Blick auf „neue Varietäten“ (Kerswill 2010) – weiterbearbeiten kann.

Literatur

- Baker, Paul (2010): *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Benmamoun, Elabbas/Montrul, Silvina/Polinsky, Maria (2013): Heritage languages and their speakers. Opportunities and challenges for linguistics. *Theoretical Linguistics* 39.3-4, 129–181.
- Blommaert, Jan/Rampton, Ben/Spotti, Massimiliano (2011): Language and superdiversity. *Diversities* 13, 1–22.
- Brehmer, Bernhard (2011): Hamburg Corpus of Polish in Germany (HamCoPoliG). Archived in Hamburger Zentrum für Sprachkorpora. Version 0.2. Publication date 2011-09-02. <http://hdl.handle.net/11022/0000-0000-63CE-9>.
- Busch, Brigitta (2017): *Mehrsprachigkeit*. 2. Aufl. Wien: facultas.
- Coseriu, Eugenio (1980): ‚Historische Sprache‘ und ‚Dialekt‘. In: Joachim Göschel, Pavle Ivić, Kurt Kehr (Hgg.), *Dialekt und Dialektologie. Ergebnisse des Internationalen Symposions „Zur Theorie des Dialekts“* (Zeitschrift für Dialektologie und Linguistik. Beihefte. Neue Folge Nr. 26 der Zeitschrift für Mundartforschung). Wiesbaden: Steiner, 106–122.
- Czachór, Agnieszka (2012): Corpus of Polish spoken in Germany. Collecting and analysing written & spoken data for investigating contact-induced change. In: Thomas Schmidt, Kai Wörner (Hgg.), *Multilingual Corpora and Multilingual Corpus Analysis*. Amsterdam, Philadelphia: John Benjamins, 153–161.
- Garrett, Peter (2010): *Attitudes to Language*. Cambridge: Cambridge University Press.
- Kendall, Tyrel (2010): Corpora from a sociolinguistic perspective. *Revista Brasileira de Linguística Aplicada* 11.2, 361–389.
- Kerswill, Paul (2010): Contact and new varieties. In: Raymond Hickey (Hg.), *The Handbook of Language Contact*. Oxford: Wiley-Blackwell, 230–251.
- Koester, Almut (2010): Building small specialised corpora. In: Anne O’Keefe, Michael McCarthy (Hgg.), *The Routledge Handbook of Corpus Linguistics*. New York: Routledge, 66–79.
- Kotková, Radomila (2017): *Čeština nerodilých mluvčích s mateřským jazykem neslovanským*. Praha: Karolinum.
- Kreisslová, Sandra (2013): *Konstrukce etnické identity a kolektivní paměti v biografických vyprávěních českých Němců na příkladu vzpomínek Němců na Chomutovsku*. Praha: FF UK.
- Kreisslová, Sandra (2014): Sprache und Identität der deutschsprachigen Bevölkerung in den böhmischen Ländern nach 1945. In: Klaas-Hinrich Ehlers, Marek Nekula, Martina Niedhammer, Hermann Scheuringer (Hgg.), *Sprache, Gesellschaft und Nation in Ostmitteleuropa. Institutionalisierung und Alltagspraxis*. Göttingen: Vandenhoeck & Ruprecht, 269–285.

- Küsters, Ivonne (2009): *Narrative Interviews. Grundlagen und Anwendungen*. 2. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lamnek, Siegfried/Krell, Claudia (2016): *Qualitative Sozialforschung*. Basel: Beltz.
- Lüpke, Friederike (2016): Uncovering small-scale multilingualism. *Critical Multilingualism Studies* 4.2, 35–74.
- Montrul, Silvina A. (2008): *Incomplete Acquisition in Bilingualism. Re-Examining the Age Factor*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Nekula, Marek/Šichová, Kateřina (Hgg.) (2017): *Variety češtiny a čeština jako cizí jazyk*. Praha: Akropolis.
- Nekvapil, Jiří (Hg.) (1998): *(Language) Biographies of Czech German. A Collection of German Transcripts*. Prague: Charles University.
- Nekvapil, Jiří (2001): From the biographical narratives of Czech Germans. Language biographies in the family of Mr and Mrs S. *Journal of Asian Pacific Communication* 11.1, 77–99.
- Nekvapil, Jiří (2003): Language biographies and the analysis of language situations. On the life of the German community in the Czech Republic. *International Journal of the Sociology of Language* 162, 63–83.
- Nekvapil, Jiří (2004): Sprachbiographien und Analyse der Sprachsituation. Zur Situation der Deutschen in der Tschechischen Republik. In: Rita Franceschini, Johana Miecznikowski (Hgg.), *Leben mit mehreren Sprachen. Sprachbiographien*. Bern: Peter Lang, 147–172.
- Pavlenko, Aneta (2007): Autobiographic narratives as data in applied linguistics. *Applied Linguistics* 28.2, 163–188.
- Polinsky, Maria (2018): *Heritage Languages and Their Speakers*. Cambridge: Cambridge University Press.
- Riehl, Claudia Maria (2014): *Mehrsprachigkeit*. 3. Aufl. Darmstadt: WBG.
- Rosen, Alexander et al. (2020): *Compiling and Annotating a Learner Corpus for a Morphologically Rich Language*. Praha: Karolinum.
- Sorace, Antonella (2011): Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism* 1.1, 1–33.
- Štindlová, Barbora (2013): *Žákovský korpus češtiny a evaluace jeho chybové anotace*. Praha: Nakladatelství FF UK.
- Weinreich, Uriel/Labov, William/Herzog, Marvin (1968): Empirical foundations for a Theory of language change. In: Winfried P. Lehmann, Malkiel Yakov (Hgg.), *Directions for Historical Linguistics. A Symposium*. Austin: University of Texas Press, 95–195.
- Wiese, Heike/Alexiadou, Artemis/Allen, Shanley/Bunk, Oliver/Gagarina, Natalia/Iefremenko, Kateryna/Jahns, Esther/Klotz, Martin/Krause, Thomas/Labrenz, Annika/Lüdeling, Anke/Martynova, Maria/Neuhaus, Katrin/Pashkova, Tatiana/Rizou, Vicky/Tracy, Rosemarie/Schroeder, Christoph/ Szucsich, Luka/Tsehaye, Wintai/Zerbian, Sabine/Zuban, Yulia (2019): *RUEG Corpus (Version 0.3.0) [Data set]*. Zenodo. DOI: 10.5281/zenodo.3236069

Immigrantinnen und Immigranten aus der Tschechoslowakei in Deutschland: Musterentlehnungen in ihren sprachbiographischen Interviews

Aneta Bučková – Marek Nekula

Abstract:

Migrants from Czechoslovakia in Germany: Pattern replications in interviews on their language biography

The chapter introduces the project on the language of immigrants from Czechoslovakia to Germany based on narrative interviews concerning language biographies. The analysis of this material focuses on pattern replications of German in Czech and of Czech in German in the interviews of immigrants belonging to two groups: those with a Sudeten German and those with a Czech linguistic background. The respondents are of a similar age, they were born around 1955, and share the fact that they emigrated after the critical (puberty) age of 12 and left Czechoslovakia for Germany before 1989. The German-Czech project presented in this chapter is a part of the research network “Language across Generations: Contact Induced Change in Morphosyntax in German-Polish Bilingual Speech” (Lang-Gener). This network’s main research question concerns the effects of language contact, specifically semantic and syntactic replications. However, the German-Czech project also has a specific research question. It is based on the assumption that both groups differ in their language ideologies concerning Czech and German and that different language ideologies in the “same” environment result in different language management. The project then addresses the question whether and to what extent language ideologies have an effect on language acquisition, use and maintenance and thus also on the quantity and quality of pattern replications. With regard to Czech, the language erosion (attrition) of the primary or first language is investigated more than 30 years after resettlement in Germany. With regard to German, the substrate interference of Czech with German, the incompletely acquired first or second language, is analyzed. Pattern replications in both languages are examined focusing on syntactic pattern replications, more specifically on sentence and phrase patterns. The chapter interprets the first results for attrition using typologies developed for language contact and emerging grammars in heritage languages and second language acquisition and proves and develops them in regard to attrition.

1 Fragestellung und ihre Einordnung

Das deutsch-tschechische Teilprojekt des Verbundprojektes „Language across Generations: Contact Induced Change in Morphosyntax in German-Polish Bilingual Speech“ (LangGener) befasst sich ähnlich wie das in diesem Band von Centner, Jańczak, Nekula & Zielińska und Hansen & Nekula beschriebene deutsch-polnische Projekt mit der Morphosyntax in der gesprochenen Sprache von Bilingualen. Genauer gesagt werden in auf Deutsch und Tschechisch geführten narrativen, sprachbiographischen Interviews mit deutsch-tschechischen Bilingualen Musterentlehnungen (*pattern replications*, PAT) aus der Modell- in die Replikasprache und umgekehrt elizitiert (Matras 2009: 234–265), die anschließend in Transkripten der geschnittenen Aufnahmen annotiert und analysiert werden. Mit Musterentlehnungen werden solche Sprachkontakterscheinungen bezeichnet, bei denen nicht konkretes phonologisches Material unterschiedlichen Umfangs (Phonem, Morphem, Wort), sondern syntaktische Konstruktionen und topologische Muster sowie die Organisation, Distribution oder das Mapping semantischer oder grammatischer Bedeutungen von einer Sprache in die andere übertragen werden (Sakel 2007: 15). In der aus dem Teilprojekt hervorgehenden Dissertation (Bučková 2022) wird auf syntaktische Musterentlehnungen aus dem Deutschen ins Tschechische bzw. aus dem Tschechischen ins Deutsche fokussiert.

Derartige sprachkontaktbedingte Abweichungen von der Norm der jeweiligen tschechischen bzw. deutschen ‚Baseline‘, deren Interpretation regiolektale Besonderheiten im Blick behält (Ehlers 2018b) und sich im Detail an Korpora gesprochener Sprache (Kopřivová et al. 2017) orientiert, werden im deutsch-tschechischen Teilprojekt anhand der Erzählungen deutsch-tschechischer Bilingualer untersucht, die im Rahmen sprachbiographischer Interviews erhoben wurden. Die interviewten Personen wurden um 1955 im tschechischen Teil der Tschechoslowakei geboren und wanderten nach der sog. kritischen (neuerdings sensitiven) Periode (vgl. Montrul 2008: 266; Riehl 2014: 79) zwischen 1964 und 1989 nach Deutschland aus, als ihr Spracherwerb der Erstsprache/n abgeschlossen war und stabilisiert wurde. Zum Zeitpunkt der Datenerhebung leben die Interviewten seit mehr als zwanzig Jahren in Deutschland. Sie bilden zwei Gruppen. Die eine Gruppe sind sog. Spätaussiedlerinnen und Spätaussiedler (im Weiteren SP, so auch im Kürzel der Beispiele), d. h. Nachfahren von Personen (sudeten)deutscher Herkunft, die nicht unmittelbar nach dem Krieg aus der Tschechoslowakei zwangsausgesiedelt wurden, sondern dort bis in die 1960er Jahre lebten, bevor sie nach Westdeutschland auswanderten. Die andere Gruppe sind sog. Migrantinnen und Migranten (im Weiteren MI, so auch im Kürzel der Beispiele), d. h. Personen ohne deutsche Wurzeln, die zwischen 1968 und 1989 aus unterschiedlichen Gründen die Tschechoslowakei verließen (1986 emigrierte der letzte einbezogene Respondent), um in Westdeutschland ein neues Leben anzufangen. Beide Gruppen werden im 2. Abschnitt näher charakterisiert.

Vor der Auswanderung lebten die interviewten Personen in dem tschechischsprachigen Teil der Tschechoslowakei, in dem das Deutsche aufgrund des tschechisch dominierten Sprachregimes im öffentlichen Raum so gut wie nicht präsent war (zum Konzept des Sprachregimes vgl. Busch 2017: 135). Daher war das Tschechische nicht nur für die MI, sondern auch die SP vor der Emigration die Erstsprache (das heißt die zuerst erworbene Sprache, vgl. Riehl 2014: 12) bzw. die primäre Sprache (die öfter und in mehreren Kontexten verwendete Sprache, vgl. Montrul 2016: 92). Dies veränderte sich nach der

Auswanderung, als Deutsch allmählich zur primären und Tschechisch zur sekundären Sprache geworden ist.

Das deutsch-tschechische Teilprojekt arbeitet mit derart unterschiedlichen Gruppen, weil es von der Annahme ausgeht, dass sich innerhalb der erwähnten Gruppen in Bezug auf Deutsch und Tschechisch unterschiedliche Vorstellungen über die Sprachen bzw. Sprachideologien herausbildeten. Diese wirken sich – so die weitere Annahme – bei den Interviewten und ihren Familien auf die Netzwerkbildung und das Sprachmanagement in Bezug auf den Spracherwerb, Sprachgebrauch und Spracherhalt von Deutsch und Tschechisch aus (zum Zusammenhang von Sprachideologien und Sprachmanagement siehe auch Nekula 2021b sowie Jańczak, Nekula & Zielińska in diesem Band). Es ist zu erwarten, dass sich dies quantitativ und qualitativ auch auf die Musterentlehnungen in der jeweils anderen Sprache auswirkt.

Die klassische Definition von Silverstein (1979: 193), der Sprachideologien als „sets of beliefs about language articulated by the users as a rationalization or justification of perceived language structure and use“ versteht, deckt die bewussten und diskursiv explizierbaren Überzeugungen über die Sprache/n ab. Das Sprachmanagement wird aber auch durch habitualisierte Sprachideologien einer Sprachgemeinschaft mitbestimmt (Errington 2001), die man sich mit dem Erwerb von kommunikativen Normen unbewusst aneignet (zum Beispiel öffentlich vs. privat verwendete Sprachen) und die einem etwa erst durch eine Kommunikationsstörung oder einen Bruch in der Sprachbiographie bewusst werden. Bewusst oder unbewusst manifestieren sich Sprachideologien im Sprachgebrauch, beispielsweise in dem Bevorzugen einer Sprache vor einer anderen oder in der Motivation, eine Sprache zu behalten und den Sprachwechsel (*language shift*) zu vermeiden. Damit geht auch das Bilden und Aufrechterhalten sozialer Netzwerke mit Sprechern und Sprecherinnen der jeweiligen Sprache einher. Daraus wird ersichtlich, dass die Sprachideologien auch die soziale Identität einer Person prägen können (Kroskirty 2004: 509; 2010: 203). Oft geschieht dies im Rahmen der Abgrenzung zu anderen Gruppen (Kroskirty 2010: 204), wobei die Sprache als ein Abgrenzungskriterium dient (Kreisslová 2014: 270; in Anlehnung an Barth 1998 [1969]: 14).

Das Sprachmanagement, das heißt das Verhalten von Sprecherinnen und Sprechern gegenüber der Sprache und damit auch das Behandeln sprachlicher Probleme (vgl. Neustupný 2002: 436), basiert auf der Identifizierung der Abweichung von sprachlichen, kommunikativen und soziokulturellen Normen. Der Identifizierung dieser Abweichung folgt ihre Evaluierung, die mit der Einstellung verbunden ist. Durch ihre durch Migration bedingte Mehrsprachigkeit stellen die MI und SP in dem einsprachigen, in unserem Falle deutsch dominierten Sprachregime ihres Ziellandes eine Abweichung von der Norm dar, die sie in dem *family language management* auflösen müssen, wie dies auch die *language management summaries* (Nekvapil 2004b) in den sprachbiographischen Interviews erahnen lassen, die in dem Projekt aufgenommen wurden.

Die Sprachideologien, die dabei explizit werden und in soziale Normen rückverbunden werden (Woolard 1998: 3), scheinen das Sprachmanagement in den Familien zu prägen (Neustupný 2002: 435) oder umgekehrt, sie werden bei der argumentativen Rationalisierung des Sprachmanagements eingesetzt, das die migrations- oder anderweitig bedingten Abweichungen von der Norm auflöst:

- (1) já jenom vím že jsem říkala jakmile zavřeme dveře za # za za _ za z bytu se mluví německy venku jsem nikdy nemluvila česky s dětma # _ # protože _ já jsem měla ten dojem že to je drzost kvůli těm _ co tam seděj vedle mě na hřišti když já se budu bavit cizí řečí kterou oni nerozuměj tak je to _ neslušnost _ bis auf drzost _ jo ? (KG_MUN_MI_CZ)

ich weiß nur dass ich gesagt habe sobald wir die Tür schließen hinter # hinter hinter _ hinter von der Wohnung spricht man Deutsch draußen habe ich nie Tschechisch gesprochen mit den Kindern # _ # weil _ ich hatte den Eindruck dass es eine Frechheit gegenüber denen die neben mir auf dem Spielplatz sitzen ist wenn ich mich in einer Fremdsprache unterhalte die sie nicht verstehen so ist es eine Unhöflichkeit _ oder sogar Frechheit _ ja ?

Im folgenden Beispiel werden Vorstellungen über die Sprache sichtbar, die sich auf alle Ebenen des Sprachmanagements und ihre Sprachnormen beziehen:

- (2) Int: A vy jste říkala že máte dceru ? a s tou jste mluvila česky ? nebo umí česky ?
 CI_BUB: To je něco co nás **seh*** # moc # **beschäftiguje** to je to je slovíčko že ? # moje sestry to ne* nedokázaly a já taky ne ale # mně to fakt # chybí nebo říkám nevím proč jsem to nezačala ona ona zná tak zmrzlinu nebo dobrý den nebo takový věci ale že bych s ní mluvila jako to jsem taky neuměla a někdy mně tak # jeden _ nevím co byl _ on říká **das ist nicht deine Muttersprache** _ (2) okay ale i když to já ráda mluvím _ proč jsem to nedala dále ? (CI_BUB_SP_CZ)

Int: Und Sie sagten Sie haben eine Tochter ? Und mit der haben Sie tschechisch gesprochen ? Oder kann sie Tschechisch ?

CI_BUB: Das ist etwas was uns **seh*** [CS] # sehr # **beschäftigt** [MAT mit tschechischer Flexion] das ist das ist ein Wort was ? # meine Schwestern haben es ni* nicht geschafft und ich auch nicht aber # mir # fehlt es wirklich oder ich sage ich weiß nicht warum ich damit ich nicht angefangen habe sie sie kennt so Eis oder guten Tag oder solche Dinge aber dass ich mit ihr geredet hätte wie das konnte ich auch nicht und einmal hat mir der _ ich weiß nicht wer er war _ er sagt **das ist nicht deine Muttersprache** _ (2) okay aber auch wenn ich es gerne spreche warum habe ich es nicht weitergegeben ?

Die Interviewte, die der Gruppe der SP angehört, erklärt hier die ausgebliebene Weitergabe des Tschechischen sowohl durch die soziokulturellen Normen, die Gebrauch und Weitergabe der Sprache durch Herkunft und Identität, d. h. „Muttersprache“ begründen, als auch auf der sprachlichen Ebene im engeren Sinne. Auf dieser Ebene schreibt sich die Interviewte mit der Formulierung „aber dass ich mit ihr geredet hätte wie [über...] das konnte ich auch nicht“ eine begrenzte Beherrschung des Tschechischen zu, dessen Erosion im vorliegenden Zitat das Codeswitching und die Integration deutschen Wortguts in den Idiolekt der Interviewten illustrieren. Diese Erosion wird zwar im Sinne des Sprachmanagements vor dem Hintergrund der Heimatland-Baseline des Tschechischen wahrgenommen und selbstironisch bewertet, nicht aber durch eine Selbstkorrektur korrigiert, was durch eine Orientierung an einer Sprachkontaktnorm erklärbar ist, die sich von der Baseline-Norm unterscheidet und gegenüber dieser präferiert wird. In dem Beispiel wird auch das Wissen über die kommunikativen Normen deutlich, die Tschechisch in Deutschland lediglich im privaten Bereich bei Alltagsphänomenen vorsehen.

Im deutsch-tschechischen Teilprojekt wird erwartet, dass die soziale und sprachliche Praxis der zwei untersuchten Gruppen durch unterschiedliche Sprachideologien geprägt ist. Die sprachliche Praxis in sozialen Netzwerken trägt zum Erwerb der Zweitsprache

bei und entscheidet über die Aufrechterhaltung der Erstsprache, wobei das Sprachbewusstsein sowie die Möglichkeit und Fähigkeit zur Netzwerkbildung eng mit dem Alter verbunden sind. Folglich korrelieren die Sprachideologien mit den Sprachkontakterscheinungen in den Narrationen der Interviewten. Aufgrund der Wirkung, die die Überzeugungen und Rationalisierungen auf das Sprachmanagement haben, wird angenommen, dass die Spracherosion (Riehl 2014: 89) des Tschechischen bei den SP stärker ausgeprägt sein wird als bei den tschechischen MI. Aus diesem Grund wird davon ausgegangen, dass es bei den SP mehr syntaktische und semantische Musterentlehnungen im Tschechischen gibt, die auch typologisch spezifisch ausgeprägt sein können. Die Spracherosion des Tschechischen ist dabei, in Abhängigkeit von der Migrationsdauer und anderen Variablen, wie endogame Familien- bzw. Partnerbildung oder Kontakt zum Herkunftsland, auch bei den tschechischen MI zu erwarten, zumal diese – selbst in München – nur begrenzt über soziale Netzwerke verfügen (Hämmerle 2009).

Bei allen Interviewten ist ihr Alter zum Zeitpunkt der Auswanderung entscheidend. Die Sprecherinnen und Sprecher in dieser Studie sind erst nach ihrer sensitiven Periode nach Deutschland ausgereist. Das bedeutet, dass die Erstsprache bereits hinreichend stabilisiert war, während der Erwerb der Zweitsprache im späten Schul- bzw. Erwachsenenalter nicht nur durch das ausgeprägte Bewusstwerden des Zweitspracherwerbsprozesses, sondern auch die Lateralisation von rechts nach links geprägt war, bei der die Spracherkennung geschwächt wird (u. a. Riehl 2014: 80). Mit dem Migrationsalter ist auch die Möglichkeit und Fähigkeit zur Netzwerkbildung mit Deutschsprachigen eng verbunden. Diese dürfte bei den MI im Schnitt weniger ausgeprägt gewesen sein als bei den SP, die mit nahen oder fernen Verwandten und sudetendeutschen Organisationen in Verbindung treten konnten. Daher ist bei den MI in Bezug auf Deutsch mit weniger Input zu rechnen, der auch Sprachkorrekturen einschließt, sodass bei ihnen eine Fossilisierung bzw. Stabilisierung (Selinker 1972: 215; Matras 2009: 75) des Deutschen anzunehmen ist. So kann man bei ihnen im Deutschen nicht nur die Absenz der sog. inneren Mehrsprachigkeit, sondern auch kleinere Ausdifferenzierung und mehr Normabweichungen von der Baseline des Deutschen erwarten.

2 Interviewte Personen

Die im Projekt schließlich genutzten Aufnahmen richten sich auf zwanzig Bilinguale, die in der Nachkriegszeit in der ehemaligen Tschechoslowakei aufgewachsen sind und zwischen Ende der 1960er Jahre (konkret ab 1964) und 1989 (konkret bis 1986) in die Bundesrepublik Deutschland ausgewandert sind. Die allermeisten von ihnen leben heute in Bayern. Man kann sie in zwei Gruppen unterteilen, die jeweils aus zehn Personen bestehen. Im Folgenden werden diese Gruppen, die wir bereits als SP und MI bezeichnet haben, im Hinblick auf ihre sprachbiographischen Charakteristika beschrieben. Ihre genaue Übersicht einschließlich ihrer Geburtsjahre befindet sich im Anhang dieser Monographie.

Die erste Gruppe bilden Personen mit einem sudetendeutschen Familienhintergrund, die SP. Es handelt sich um Angehörige der deutschen Minderheit vornehmlich aus den Grenzgebieten der Tschechoslowakei sowie größeren Städten im Landesinneren wie Prag/Praha, Iglau/Jihlava oder Brünn/Brno. Deutschsprachige Bevölkerung besiedelte seit dem 11. Jahrhundert Teile der böhmischen Länder in mehreren Wellen (vgl. Schwarz

1934; Nekula 2021c). Ihre Stellung im Habsburgerreich war die einer Majorität, auch wenn sie in den böhmischen Ländern quantitativ gesehen eine Minorität waren. Die sozial herausgehobene Stellung war auch dadurch gegeben, dass staatliche, wirtschaftliche, edukative und kulturelle Institutionen in den böhmischen Ländern im langen 19. Jahrhundert deutsch geprägt waren, bzw. dass sich in den böhmischen Ländern eine Doppelsprachigkeit in staatlichen Institutionen herausbildete. Nach der Zerschlagung des Habsburgerreiches im Jahre 1918 blieben diese parallelen Strukturen in der Wirtschaft, Bildung, Kultur und auf der kommunalen Ebene auch in der Verwaltung erhalten, was sich neben der Tradition auch durch die Größe, die ökonomische Stärke und das kulturelle Niveau der deutschen Minderheit in der Tschechoslowakei begründete. Diese hat dadurch sprachlich eine weitgehende funktionale Autonomie genossen. Nach der Abtretung der Sudetengebiete zum 1. Oktober 1938 sind diese ein Teil des Deutschen Reiches geworden. Für die Deutschen auf dem Gebiet des Protektorats Böhmen und Mähren wurde im Rahmen von Protektoratsbehörden das deutsche Sprachregime vorgesehen. Nach der Wiederherstellung der Tschechoslowakei nach dem Ende des Zweiten Weltkrieges kam es allerdings zu einem Bruch. 1945–1947 wurde in Reaktion auf die Nazi Herrschaft die Mehrheit der in der wiederhergestellten Tschechoslowakei befindlichen „Deutschen“ zwangsausgesiedelt. Die Anzahl der Deutschen ist von 3.149.820 nach dem Zensus im Jahr 1930 auf 159.938 nach dem Zensus im Jahr 1950 gesunken, wobei die Tendenz weiter fallend war (Nekula 2021c: 72f.).

Diejenigen, die bleiben durften oder mussten, waren die sog. Spezialisten, die unentbehrlich für die Wirtschaft waren, Personen aus interethnischen (exogamen) Ehen, eine kleine Anzahl an anerkannten Antifaschisten, Personen jüdischer Abstammung und Personen, die für den nicht mehr durchgeführten Transport im Jahr 1947 gelistet waren (Kreisslová 2014: 271), wie dies auch in den im Projekt aufgenommenen sprachbiographischen Interviews dokumentiert wird:

- (3) und _ unsere _ deutsche Familie _ von meiner Großmutter und vom Großvater seine Seite die wurden ja sechszwanzig vertrieben und mein Großvater musste als Spezialist weil er Meister war in einer Weberei musste er bleiben und _ er hat dann immer wieder versucht da gab es immer glaube ich so eine Fünfjahresfrist immer um die A* Aussiedlung angeru* ange* eingereicht das wurde ein paar mal abgelehnt und wie er dann glaube ich die Rente bekommen hatte der Großvater dann _ ist die Genehmigung gekommen (QC_WUN_SP_DE)

Die sprachliche Situation der nicht ausgesiedelten Deutschen war durch die Stigmatisierung des Deutschen und dessen Ausschließen aus der Öffentlichkeit geprägt (Kreisslová 2014: 276; Kreisslová & Novotný 2018: 126). Die Assimilationspolitik des Staates führte zu einer Einschränkung der Kommunikationsdomänen auf das Familienleben, was zu einer Abschwächung der Kompetenz im Deutschen der Nachkriegsgeneration führte (Kreisslová & Novotný 2018: 127). In den exogamen Familien war der Verzicht auf das Deutsche die Regel. Selbst in den endogamen deutschsprachigen Familien entstand ein Konflikt zwischen der Primärsozialisation der Kinder in der Familie und der Sekundärsozialisation in der tschechischsprachigen Gesellschaft, die bis zum Übergang zum Tschechischen als Erstsprache und zum Identitätsshift von der deutschen zur tschechischen ethnischen Identität führen konnte (Kreisslová 2014: 280). Dies trifft vor allem für jene zu, die in den 1950er Jahren geboren wurden, als sich die Stellung der erwachsenen Angehörigen der deutschen Minderheit nach dem Staatsvertrag mit der DDR normalisierte,

der Assimilationsdruck auf die nächste Generation aber aufrechterhalten wurde und die Auswanderungschancen noch gering waren (im Detail vgl. Nekula 2021c).

Die Wirkung des tschechisch dominierten Sprachregimes auf die Gruppe der SP aus der Tschechoslowakei ist mit der Wirkung des polnisch dominierten Sprachregimes auf die Angehörigen der „Generation Deutschland“ im deutsch-polnischen Teil des Lang-Gener-Projektes vergleichbar. Die Personen, die im deutsch-tschechischen Teilprojekt befragt wurden, sind in deutschsprachigen oder deutsch-tschechischen Familien aufgewachsen und haben sich außerhalb und teilweise auch innerhalb der Familie tschechisch sozialisiert. Sie besuchten tschechische Schulen. Deutschunterricht konnten nur einige von ihnen wahrnehmen, und zwar in Form von deutschen Sprachzirkeln (Kreisslová 2014: 277), die in der Regel einmal pro Woche nach dem Unterricht angeboten wurden. Tschechisch ist demnach zu ihrer primären Sprache geworden. In den 1960er Jahren stellten ihre Familien einen Antrag auf Aussiedlung nach Westdeutschland, nach dessen Bewilligung sie die Tschechoslowakei verlassen haben. Die Aussiedlung wurde durch den Anspruch auf die deutsche Staatsbürgerschaft möglich (mehr dazu siehe Dietz 2011), was sich auch auf die Stellung der SP sowie die Sprachideologien und das Sprachmanagement in Deutschland auswirkte.

Nach der Aussiedlung kam es bei dieser Gruppe der Interviewten zu einem abrupten Sprachwechsel bzw. zur Umstellung von Mehr- auf Einsprachigkeit. Sie sprachen dann außerhalb wie innerhalb der Familie nur noch Deutsch, was – weil sie dazu auch von ihren Eltern angehalten wurden – als Ergebnis des Sprachmanagements in den Familien verstanden werden kann. Für die meisten gab es nach der Ankunft in Deutschland keine Möglichkeit, Tschechisch zu sprechen, sodass man bei ihnen in Bezug auf Deutsch und Tschechisch in ihrem Alltag von einem monolingualen Sprachmodus ausgehen kann (zum mono- und bilingualen Sprachmodus vgl. Grosjean 2001). Nur einige haben gelegentlich mit ihren Geschwistern Tschechisch gesprochen. Dies änderte sich für manche später, als sie die Gelegenheit hatten, dienstlich oder privat in die Tschechoslowakei bzw. später in die Tschechische Republik zu reisen, wobei der Kontakt zum Tschechischen auch durch moderne Medien gefördert wird (vgl. Keijzer 2021: 221).

Die andere Gruppe bilden die MI. Dies sind Personen etwa gleichen Alters, die keinen deutschen Familienhintergrund haben und deren Emigration etwa zur selben Zeit erfolgt wie die Auswanderung der SP. Sie wurden auf dem Gebiet der heutigen Tschechischen Republik geboren und lebten – im Falle der Personen in unserem Sample – meistens in Prag. Sie sind monolingual in einer tschechischsprachigen Umgebung aufgewachsen und sind anderen Sprachen in der Regel nur im Schulunterricht begegnet. Ihre erste Fremdsprache war Russisch, später kamen auch andere Fremdsprachen dazu, mitunter auch Deutsch:

- (4) ich sprach kein Wort Deutsch also das war Deutsch war für mich der die das psi vocas das war das einzige (alle lachen) ich hatte zwar in der Schule ein bisschen Deutsch gehabt [...] also meine Sprachkenntnisse waren sehr _ null tendierten gegen null oder null (KG_MUN_MI_DE)

Sie emigrierten – abhängig von ihrem Alter entweder zusammen mit ihren Eltern oder allein bzw. mit ihren Partnerinnen und Partnern – vor allem aus politischen und wirtschaftlichen Gründen in Reaktion auf die Besetzung der Tschechoslowakei durch die Truppen des Warschauer Paktes im August 1968 und der politisch wie ökonomisch lähmenden Zeit der

sog. Normalisierung in den 1970er und 1980er Jahren (Nešpor 2005: 266; Hämmerle 2009: 320). Laut Hämmerle (2009: 321f.) wanderten in dieser Zeit die meisten Tschechinnen und Tschechen nach Bayern aus und siedelten sich in und um München an. Hämmerle (2009: 324) macht auf ihre Assimilationsbereitschaft aufmerksam, die dadurch gestärkt wurde, dass die Reemigration aufgrund der politischen Lage in der Tschechoslowakei und der Kriminalisierung der Emigration und Rückkehrer als nicht realistisch erschien. In München wurden zwar von diesen MI tschechische Vereine gegründet oder wieder ins Leben gerufen (zum Beispiel *Česká katolická misie* oder *Sokol*), insgesamt hat sich jedoch die tschechische Gemeinschaft nicht stark untereinander vernetzt (Hämmerle 2009: 325). Weitere Aspekte der niedrigen Vitalität des Tschechischen im Vergleich etwa mit dem Russischen oder den südslawischen Sprachen diskutiert Achterberg (2005), dessen Daten sich auf MI vor 1989 beziehen.

Obwohl es also keine stark vernetzte tschechische Sprachgemeinschaft in Bayern gab, kann man erhebliche Unterschiede zwischen dem Sprachgebrauch dieser Gruppe und den oben beschriebenen SP feststellen. Auch diesen Sprachgebrauch kann man anhand der *language management summaries* in den Interviews als Ergebnis der Sprachmanagements einschätzen. So verwendeten MI weiterhin Tschechisch im Kontakt mit Freunden und Bekannten, innerhalb ihrer Familie und mit ihren Partnerinnen und Partnern, falls sie in einer endogamen Beziehung lebten. Insgesamt kann man also in Bezug auf Deutsch und Tschechisch im Alltag der MI von einem bilingualen Sprachmodus ausgehen. Außerdem waren sie im Kontakt mit ihrem Heimatland, wo sie ihre Angehörigen zurückgelassen hatten, und nach Erlangen der deutschen Staatsbürgerschaft oder spätestens nach der Wende reisten sie beruflich oder privat in die Tschechoslowakei bzw. nach Tschechien. In den meisten Fällen sprachen sie mindestens zu Hause auch mit ihren Kindern Tschechisch und gaben die Sprache bewusst an die nächste Generation weiter.

3 Erhebung

Ähnlich wie im deutsch-polnischen Projekt wurden auch bei den deutsch-tschechischen Bilingualen die sprachlichen sowie metasprachlichen (soziolinguistischen) Daten mittels narrativer Interviews erhoben. Diese Methode wird in der qualitativen Sozialforschung generell für biographisch orientierte Fragestellungen verwendet (Kruse 2015: 151). Im Falle des LangGener-Projektes lag der Schwerpunkt der biographischen Erzählung auf der Sprache bzw. den Sprachen und ihren Varietäten, um die Sprachbiographie der interviewten Personen und ihren Umgang damit zu rekonstruieren. In einer Sprachbiographie wird die Verflechtung von Lebensgeschichte und Sprache(n) der Interviewten fokussiert, insbesondere der Spracherwerb, der Sprachgebrauch und der Erhalt erworbener Sprachen (Nekvapil 2003: 63; Meng 2004: 98). So gewinnt man einerseits die metasprachlichen Daten, die in sog. *language management summaries* verdichtet sind, der Rekonstruktion von Sprachbiographien der interviewten Personen dienen und zu einer prototypischen Sprachbiographie zusammengefasst werden können. Diese wurde jeweils für die Gruppe der SP und die Gruppe der MI oben skizziert. Andererseits dienen die Transkripte der sprachbiographischen Interviews auch der Gewinnung von sprachlichen Daten, an denen sich – zusammengefasst in einem Korpus – Isolations- und Sprachkontaktphänomene, sei es die der Spracherosion oder der Fossilisierung, untersuchen lassen.

Die Suche nach passenden Respondentinnen und Respondenten verlief auf unterschiedliche Art und Weise. Eine wesentliche Rolle spielte dabei, ob es sich um SP oder MI handelte.

Bei den SP war der Kontakt zum Verband *Sudetendeutsche Landsmannschaft* bedeutend. Über diese Organisation konnten allerdings vor allem ältere SP erreicht werden, die der Vorkriegs- oder Kriegsgeneration angehören (vgl. Kreisslová 2014: 270). Über ihre Bekannten konnten jedoch schließlich auch Personen passenden Alters gefunden werden, die wiederum ihre Freunde oder Verwandten vermittelten. Man kann daher sagen, dass bei dieser Gruppe der Interviewten das Schneeballprinzip (Schütze 1976: 163, 252; vgl. auch Küsters 2006: 49) genutzt wurde, dessen Anwendung zudem auch die persönliche Vernetzung der SP in die sudetendeutschen Netzwerke bestätigte.

Die MI waren durch das Kontaktieren verschiedener Vereine, die von Tschechinnen und Tschechen in München betrieben werden (siehe Abschnitt 2), schwer zu erreichen. Ein Grund dafür ist sicherlich die oben erwähnte eher schwache Vernetzung der tschechischen Sprachgemeinschaft in Bayern. Als besonders ergiebig erwies sich der Besuch einer Kulturveranstaltung anlässlich des 100. Gründungstages der Tschechoslowakei in München. Dort konnte die Interviewerin die Zuschauerinnen und Zuschauer direkt ansprechen und um Unterstützung bei dem Forschungsprojekt bitten. Erst durch diesen persönlichen Kontakt konnten auch Verbindungen zu den Bekannten der direkt angesprochenen Personen hergestellt werden.

Die narrativen Interviews wurden in den meisten Fällen bei den Interviewten zu Hause durchgeführt, das heißt an einem für die Erzählenden vertrauten Ort (Glinka 2009: 135), an dem sie sich wohlfühlen und an dem sie daher auch über persönliche Themen möglichst entspannt sprechen konnten. Die Interviews bestanden im Wesentlichen aus einem Vorgespräch, dem Erzählstimulus, der Ratifizierung des Stimulus, der Haupterzählung, dem Nachfrageteil und einem Nachgespräch (vgl. Küsters 2006: 54ff.; Helfferich 2009: 36).

Im Vorgespräch wurde zwischen den Interviewten und der Interviewerin zunächst Vertrauen aufgebaut. Dies geschah einerseits im anfänglichen Smalltalk, andererseits durch das Vorstellen des Forschungsvorhabens. Dieses wurde nicht allzu ausführlich erklärt, um die Erzählung nicht im Voraus zu beeinflussen (Küsters 2006: 54). Es wurde ungefähr folgendermaßen beschrieben: „In unserem Projekt sammeln wir Geschichten von Menschen, die in der Tschechoslowakei geboren sind und später nach Deutschland ausgewandert sind, und es interessieren uns vor allem die Sprachen, wie es damals mit Deutsch und Tschechisch war und wie es sich dann vielleicht änderte.“

Des Weiteren bat die Interviewerin um Erlaubnis, das Gespräch aufnehmen zu dürfen, und sicherte den vertraulichen Umgang mit den personenbezogenen Daten der Interviewten zu. Falls an dieser Stelle Unsicherheiten bezüglich des Aufzeichnens oder des Erzählens überhaupt aufgetreten sind, wurde die Einwilligungserklärung zum Durchlesen und Unterzeichnen vorgelegt, in der die Verarbeitung der personenbezogenen Daten ausführlicher geschildert wurde. Falls die Interviewten der Aufnahme des Interviews gleich zustimmten, wurde die Einwilligungserklärung erst am Ende im Rahmen des Nachgesprächs gezeigt und unterschrieben.

Am Ende des Vorgesprächs wurde das weitere Vorgehen ungefähr so erläutert: „Ich würde Sie zuerst erzählen lassen und dabei nicht unterbrechen, und wenn mir was einfällt, mache ich mir eine Notiz. Danach kommt ein Nachfrageteil, in dem ich Ihnen Fragen

stellen kann, falls mir vorher etwas unklar war oder so. Ich würde gerne auf Deutsch anfangen und im zweiten Teil auf Tschechisch fortsetzen.“

Das Aufteilen in eine Haupterzählung, in der die Erzählerin oder der Erzähler „das monologische Rederecht“ hat, und einen Nachfrageteil, der in ein Leitfadenterview übergehen kann, ist die charakteristische Besonderheit des narrativen Interviews (Helfferich 2009: 36). Den Erzählenden wird bewusst Raum für ein eigenständiges Gestalten der eigenen Lebensgeschichte gegeben, was den Erinnerungsprozess fördert (Rosenthal 1995: 193ff.). Ein nicht weniger wichtiger Aspekt des freien Erzählens ist das allmähliche Nachlassen der Selbstkontrolle (ebd.: 195f.). Die Interviewerin unterstützte den Erzählfluss durch ihr aktives Zuhören, das heißt durch Mimik, Blickkontakt, durch *hmm, aha* etc. gefüllte Pausen, die Anteilnahme signalisieren, und durch eine einfühlsame Reaktion auf die durch den Erzählenden geäußerten Emotionen (Rosenthal 1995: 196, 200; Küsters 2006: 58; Glinka 2009: 12f.).

Die erwähnte Aufteilung in eine deutschsprachige und eine tschechischsprachige Interviewphase ist durch die Sprachbiographie der befragten bilingualen Personen begründet. In der ersten Erzählung und in dem darauffolgenden ersten Nachfrageteil wurde auf Deutsch das Leben vor der Aussiedlung bzw. der Emigration, das heißt die Zeit in der Tschechoslowakei dargestellt. Es war anzunehmen, dass die befragten Personen diese Zeit vornehmlich mit der tschechischen Sprache verbinden, denn in diesem Lebensabschnitt war Tschechisch ihre primäre Sprache (siehe Abschnitt 2). Eine Erzählung auf Deutsch konnte demnach den kognitiven Druck bei der Selektion lexikalischer und grammatischer Strukturen erhöhen und die kognitive Kontrolle senken (vgl. Matras 2009: 151), was das Auftreten von Sprachkontakterscheinungen begünstigen konnte. Analoge Prozesse konnten in dem zweiten Interviewteil ausgelöst werden, in dem auf Tschechisch über die Zeit in Deutschland erzählt wurde.

Nach dem Erläutern des Vorgehens gab die Interviewerin den Erzählstimulus: „Erzählen Sie mir bitte Ihre Geschichte, bevor Sie nach Deutschland gekommen sind.“ Bei einigen wenigen befragten Personen folgte eine kurze Aushandlungsphase (Küsters 2006: 56), in der sie sich versicherten, dass die Zeit vor der Aussiedlung bzw. der Emigration auch relevant sei. Dies war tendenziell bei der Gruppe der MI der Fall, die ihr Leben vor der Auswanderung aufgrund ihrer Einsprachigkeit als für das Interviewthema nicht interessant empfanden. Wie Glinka (2009: 10f.) anmerkt, müssen die befragten Personen einen Sinn darin sehen, weshalb sie ihre Geschichte erzählen sollen. Dieser Aspekt war bei der Gruppe der SP weniger problematisch, was auch mit dem Kontext zusammenhängt, in dem sie von der Interviewerin erstmals kontaktiert wurden. Für sie war es selbstverständlicher, über die eigene Kindheit in einer deutschen Familie in der Tschechoslowakei und über die Geschichte ihrer Familie zu erzählen, denn sie verstanden es zum Teil als ein wichtiges Zeugnis über eine nicht mehr existierende (Sprach-)Gemeinschaft.

Nachdem das Thema von den interviewten Personen ratifiziert wurde, begannen sie mit der Haupterzählung. Im Verlauf des Interviews legte die Interviewerin Wert auf das Entstehen von Nähe zwischen ihr und den interviewten Personen, sowohl auf der emotionalen als auch auf der kognitiven Ebene (vgl. Helfferich 2009: 119f.). Ersteres wurde vor allem durch das oben erwähnte einfühlsame Zuhören erreicht. Letzteres konnte teilweise dadurch hergestellt werden, dass sowohl die interviewten Personen als auch die Interviewerin aus dem heutigen Tschechien kommen und dort einige Zeit lang lebten. Das heißt, dass die interviewten Personen über für beide Seiten bekannte Orte, Ereignisse, Persönlichkeiten,

Speisen etc. sprechen konnten, ohne befürchten zu müssen, dass sie dabei nicht verstanden werden. Auf der anderen Seite konnte ihre Registerwahl im Deutschen dadurch beeinflusst werden, dass die Interviewerin keine deutsche Muttersprachlerin war. Dies dürfte bei Einigen zur Wahl einer eher standardnahen Varietät des Deutschen geführt haben.

Nach den Haupterzählungen in beiden Sprachen folgte jeweils ein Nachfrageteil. Zuerst wurden immanente Fragen gestellt, die sich auf mögliche Lücken aus der vorangegangenen Erzählung bezogen und weitere Erzählungen generieren sollten (Küsters 2006: 61). Außerdem hielt die Interviewerin einen exmanenten Fragenkatalog bereit, der die einzelnen Aspekte der Sprachbiographie, vor allem den Spracherwerb und den Sprachgebrauch betreffend, abdeckte. Die Fragenbereiche für die Zeit vor der Auswanderung umfassten die soziale Einbettung, die Familie und andere Domänen des Sprachgebrauchs (Freundeskreis, Religion, Berufsleben der Eltern, Institutionen und Öffentlichkeit, Vorschule und Schule). Die Fragenbereiche für die Zeit nach der Emigration umfassten die Familiensituation nach der Emigration, den weiteren Spracherwerb und den Sprachgebrauch in den Domänen Schule, Ausbildung oder Studium, Berufsleben, eigene Familie, Freizeit und breiteres soziales Umfeld in Deutschland (Engagement in tschechischen oder sudetendeutschen Vereinen) und in Tschechien (Kontakte und Reisen), sowie die Mediennutzung passiv (Lesen, Radiohören, Fernsehen) und aktiv (Telefonieren, Internetnutzung, E-Mail- oder Briefeschreiben) und Einstellungen (Selbsteinschätzung der eigenen Sprachkompetenz etc.).

An das Interview schloss sich ein Nachgespräch an. Oft wurde in dieser Phase die Einwilligungserklärung unterschrieben (siehe oben) und die interviewten Personen fragten nach der weiteren Bearbeitung des aufgenommenen Materials. Oft erfolgten auch Nachfragen bezüglich der Biographie der Interviewerin, die sich bis dahin bewusst zurückgehalten hatte.

Insgesamt wurden mit den SP 23 Audioaufnahmen und mit den MI 12 Audioaufnahmen jeweils auf Deutsch und auf Tschechisch angefertigt, die anschließend als ein Ganzes evaluiert wurden. Einige Audioaufnahmen wurden daraufhin aussortiert, um das Sample einheitlich zu gestalten. Dies betraf Audioaufnahmen, bei denen sich im Laufe des Interviews zeigte, dass die interviewten Personen doch nicht in eine der Gruppen fallen. Die verbleibenden Audioaufnahmen wurden geschnitten und weiter technisch bearbeitet. Danach wurden sie transkribiert und durch Annotation der sprachkontaktbedingten lexikalischen und grammatischen Phänomene weiterbearbeitet (mehr dazu siehe Bučková, Centner, Książek & Prawdzic in diesem Band). Der Prozess der Umwandlung des Interviews in eine bearbeitete Audioaufnahme, die einerseits in ein Transkript übertragen und andererseits mit diesem verbunden wird, wird in diesem Band von Bučková & Centner erläutert. Die von den Audioaufnahmen angefertigten Transkripte wurden in ein kleines projektbezogenes Korpus (mehr dazu siehe Hansen & Nekula sowie Woźniak in diesem Band) eingeordnet, das im deutsch-tschechischen Teilprojekt dem Tschechischen Nationalkorpus zur Verfügung gestellt werden soll, das auch für die weitere informationstechnische Bearbeitung des Korpus zuständig ist.

Auf dieser Basis werden die sprachlichen sowie die metasprachlichen Daten ausgewertet. Individuelle und prototypische Sprachbiographien sowie Sprachideologien, durch die das für den Spracherwerb, -gebrauch und -erhalt bedeutende Sprachmanagement begründet wird, werden in Bezug auf *language management summaries* rekonstruiert, die das einstige Sprachmanagement in den Familien retrospektiv abrufen und zusammenführen.

Die Sprachkontakthänomene können mit Verwendung der Annotation in Bezug auf die Fragestellung dieses Teilprojekts sowie andere Fragestellungen sowohl quantitativ als auch qualitativ ausgewertet und interpretiert werden, was im 6. Abschnitt dieses Beitrags beispielhaft vorgeführt wird.

4 Unterschiede zum deutsch-polnischen Teilprojekt

Den Teilprojekten des LangGener-Projekts ist gemeinsam, dass man in dem Verbundprojekt Fragen nach *pattern replication* (PAT) bei Bilingualen unter den Bedingungen eines langanhaltenden intensiven Sprachkontakts nachgeht und so eine grammatikorientierte Forschung mit der Soziolinguistik verbindet. Zwischen den Teilprojekten gibt es allerdings auch einige inhaltlich und organisatorisch bedingte Unterschiede.

Das deutsch-polnische Teilprojekt, das auf ein relativ großes Team zurückgreifen konnte, setzte sich zwei grundlegende Ziele: (1) die Sprache der in Polen verbliebenen deutschen Minderheit zu dokumentieren, wobei eine Ausdifferenzierung nach der Geburtsregion innerhalb des heutigen Polens und dem Zeitpunkt des Zweitspracherwerbs geschah; (2) Auswirkungen des Sprachkontakts bei den Verbliebenen und den SP zu untersuchen und zu vergleichen, wobei bei Letzteren auch die Dauer des Aufenthalts in Deutschland berücksichtigt wurde. Daher sowie angesichts der grundsätzlich anderen Sprachregimes, in denen die Angehörigen der beiden Gruppen aufgewachsen sind, wurde im deutsch-polnischen Teilprojekt zwischen der „Generation Polen“ (Wechsel vom deutsch zum polnisch dominierten Sprachregime) und der „Generation Deutschland“ (Wechsel vom polnisch zum deutsch dominierten Sprachregime) unterschieden (zum Generationskonzept und der Unterscheidung beider Generationen siehe Centner in diesem Band).

Das erste Ziel entfiel bei dem deutsch-tschechischen Teilprojekt komplett. Dies hing sowohl mit den begrenzten personellen Ressourcen für das deutsch-tschechische Teilprojekt als auch mit der unterschiedlichen Ausprägung der deutschen Minderheit in der Tschechoslowakei und Polen zusammen. Durch die äußere und innere Zwangsaussiedlung war und ist die deutsche Minderheit in Tschechien stärker zersplittert (zur inneren Zwangsaussiedlung vgl. etwa Dvořák 2012) als die lokal kondensierte deutsche Minderheit in Polen. Die niedrigere Kondensation hatte dabei Auswirkungen auf den Erhalt und die transgenerationale Weitergabe des Deutschen im tschechischen Teil der Tschechoslowakei. Auch deshalb gehören die Interviewten altersmäßig nur einer Generation an, für die die oben skizzierte Fragestellung erarbeitet wurde. So wird im deutsch-tschechischen Teilprojekt statt der Generationen mit Gruppen – den SP und den MI – gearbeitet. Dabei leben die meisten Interviewten in Bayern. Der Vorteil dieser Beschränkung besteht nicht nur in der Einsparung von Ressourcen, sondern auch darin, dass in Bezug auf das Deutsche nur eine regional geprägte Baseline des gesprochenen Deutsch in Betracht gezogen werden musste (zum Regiolekt als Baseline beim Sprachkontakt vgl. u. a. Ehlers 2018b). Da im Sinne der Theorie der Akkommodation vorausgesetzt werden kann, dass in der informellen Interaktion mit den Interviewten von deren Gesprächspartnerinnen und -partnern statt Dialekt eher der höher anzusetzende bayerische Regiolekt genutzt wurde und dieser für die Interviewten zur Modellsprache geworden ist, gilt er als Baseline für die Interpretation der deutschen Interviewteile (zur Akkommodation allgemein vgl. Giles et al. 1991; spezifisch für

Regiolekt im Sprachkontakt vgl. Ehlers 2018a). Dagegen sind bei den untersuchten Gruppen unterschiedliche Ausprägungen von Sprachideologien, Sprachmanagement, Sprachgebrauch und Sprachkontaktphänomenen (stärkere Spracherosion des Tschechischen bei den SP, stärker und anders ausgeprägte Substratinterferenz aus dem Tschechischen ins Deutsche bei den MI) vorauszusetzen und festzustellen.

Im deutsch-polnischen Team geschah die Verbindung der grammatikorientierten Forschung mit der Soziolinguistik in Arbeitsgruppen, die zwar bei der Erhebung der „Generation Polen“ und der „Generation Deutschland“ sowie methodologisch miteinander kooperierten, im Grunde genommen jedoch unterschiedliche Ziele verfolgten. Das deutsche Team des deutsch-polnischen Teilprojektes fokussierte bei der Auswertung von Interviews auf die *pattern replications*, während das polnische Team neben der formorientierten Analyse auch die Sprachdokumentation der „Generation Polen“ und die Sprachbiographien in den Vordergrund stellte. Im deutsch-tschechischen Teilprojekt hingegen rückte die Verbindung der grammatikorientierten Forschung mit der Soziolinguistik stärker ins Zentrum der im 1. Abschnitt formulierten Fragestellung. Die Sprachideologien und das Sprachmanagement der Bilingualen werden als entscheidende Kriterien für den Spracherhalt oder Sprachverlust und damit auch für die Frequenz und die Typen der Musterentlehnungen bei den Bilingualen angesehen, worauf auch das Sample und die darin vorgenommene Gruppenbildung ausgerichtet ist.

Das Vorgehen bei den Interviews und der Fragenkatalog sind in beiden Teilprojekten wiederum grundsätzlich vergleichbar. Im deutsch-tschechischen Teilprojekt wurden alle Gespräche von derselben Interviewerin geführt, was eine gewisse Einheitlichkeit garantiert. Gleichzeitig war jedoch kein deutscher Muttersprachler anwesend, der den Interviewteil in deutscher Sprache hätte führen können. Dies könnte die Registerwahl der interviewten Personen beeinflussen haben (siehe Abschnitt 3). Aufgrund der Abwesenheit eines zweiten Interviewers war auch die unter 3. beschriebene thematische Aufteilung in eine deutsche und tschechische Erzählung notwendig, um Sprachdaten in beiden untersuchten Sprachen zu erhalten. Der Prozess der Datenverarbeitung einschließlich der Transkription und Annotation von Sprachkontaktphänomenen verlief ebenfalls im Einklang mit dem deutsch-polnischen Teilprojekt. Aus Kapazitätsgründen verzichtete jedoch das deutsch-tschechische Teilprojekt auf eine soziolinguistische Annotation, die das deutsch-polnische Teilprojekt auszeichnet (mehr dazu siehe Jańczak, Nekula & Zielińska in diesem Band). Das Korpus des tschechischen Teilprojektes soll im Tschechischen Nationalkorpus (Portál | Český národní korpus 2020) zugänglich sein.

5 Theoretische Einordnung

Durch die Erforschung morphosyntaktischer *pattern replications* deutsch-tschechischer Bilingualer mit unterschiedlichem soziolinguistischem Profil steht das deutsch-tschechische Teilprojekt wie das gesamte LangGener-Projekt im Rahmen der Mehrsprachigkeitsforschung an der Schnittstelle der grammatikorientierten Forschung, der Psycholinguistik und der Soziolinguistik. Die soziolinguistische Dimension des Projektes wurde bereits im Zusammenhang mit der Diskussion der Gruppen der Interviewten und den Hypothesen in Bezug auf ihre Sprachideologien und ihr Sprachmanagement erläutert. Die erwarteten unterschiedlichen Effekte auf der morphosyntaktischen Ebene haben aber sowohl bei der

Erosion des Tschechischen als auch beim Zweitspracherwerb des Deutschen neben der grammatischen auch eine psycholinguistische Dimension.

Die besondere soziolinguistische Charakteristik der untersuchten Gruppen tritt deutlich hervor, wenn man sie in die germanobohemistische Sprachkontakt- bzw. Mehrsprachigkeitsforschung einordnet, in der es nach 1945 um unterschiedliche Gruppen von Sprecherinnen und Sprechern geht:

(a) „Vertriebene“ mit Deutsch als L1, die 1945–1947 im Umfang von ca. 2,9 Mio. zwangsausgesiedelt wurden und für welche Ehlers (2015) eine „Vertriebenenlinguistik“ einfordert, in deren Rahmen er ihr Deutsch sowie den intralingualen Sprachkontakt beleuchtet;

(b) „unsere Deutschen“, die nach der Zwangsausiedlung der Deutschen aus der Tschechoslowakei dort verblieben sind (nach dem Zensus im Jahre 1950 lediglich 159.938, im Jahre 2001 nur noch 41.238), deren Dialekt bei den in den 1930er Jahren Geborenen als bedrohte Sprache idealtypisch dokumentiert wird (Bachmann et al. 2014–2019), während die nach 1945 Geborenen für die Spracherosion und den Sprachverlust des Deutschen als Ergebnis des indirekten Sprachkontakts bzw. des unvollständigen bzw. ausgebliebenen Spracherwerbs des Deutschen als L1 stehen (Tišerová 2008; Nekula 2021c);

(c) SP, die in der Tschechoslowakei geboren wurden und dort in Abhängigkeit von dem (endogamen vs. exogamen) Familientypus und dem Geburtsjahr Deutsch und/oder Tschechisch als L1 erworben und eine auf das Zielland und die dort zwangsausgesiedelten Sudetendeutschen (siehe a) bezogene ethnische Identität bewahrt oder entwickelt haben; nach der Aussiedlung haben sie Deutsch als L2 erworben bzw. als ihre primäre Sprache wiedergewonnen, während Tschechisch als L1 bzw. als einstige primäre Sprache sich durch Spracherosion veränderte, was bisher wenig erforscht, doch analog zur Erforschung der Sprache/n der „Russlanddeutschen“ in Deutschland (Meng & Protassova 2001 u. a. m.) zu sehen ist;

(d) MI der ersten Generation mit Tschechisch als L1, die zwischen 1948 bis 1989 hauptsächlich als politische Migranten nach Deutschland kamen und dort Deutsch als L2 (unvollständig) erwarben und Tschechisch als L1 infolge der Isolation durch Spracherosion (Feldmeier 2014) teilweise verloren. Der Sprachverlust wird mit Blick auf die soziolinguistischen Voraussetzungen der sprachlichen Vitalität (Achterberg 2005; Hämmerle 2009) diskutiert und kann im Kontext der Migrationslinguistik (Krefeld 2004) eingeordnet werden; dabei kann man zwischen dieser Generation, die als „Generation Exil“ charakterisiert wird (vgl. Nekula 2021b), und der „Generation Europa“ unterscheiden, für die durch den politischen Wandel von 1989 der Kontakt zum Heimatland selbstverständlich und der Spracherhalt naheliegender geworden ist;

(e) Herkunftssprecherinnen und -sprecher als MI der zweiten Generation bzw. der „Generation Anderthalb“, bei denen der unvollständige Spracherwerb (sowie Spracherhalt) des Tschechischen als L1 im Vordergrund steht (Seidlmayer 2010, 2013; Meier 2017; Panochová 2017; Plesníková 2018; Nekula 2021a-b) und dessen Erforschung im Kontext der sich dynamisch entwickelnden *heritage linguistics* zu sehen ist (Polinsky 1995, 2006, 2018; Montrul 2008; Kagan et al. 2017 oder Brehmer & Mehlhorn 2015); für einen Überblick für Russisch in Deutschland vgl. Anstatt (2009, 2011), für Polnisch vgl. Blaszczyk (2015), für BKS vgl. Hansen (2018).

Das deutsch-tschechische Teilprojekt untersucht die Angehörigen der Gruppen (c) und (d), die in einem Land mit tschechisch dominiertem Sprachregime sozialisiert worden sind, aus dem sie erst nach der sensitiven Periode in ein Land mit deutsch dominiertem Sprachregime emigriert sind, in dem sie seit über 20 Jahre leben. Damit leistet das Projekt einen Beitrag zur Erforschung des intensiven lang andauernden Sprachkontakts, der

unter einem besonderen kulturellen Druck geschieht und sprachkontaktinduziert auch das scheinbar resistente grammatische System verändern kann (mehr dazu u. a. auch Thomason 2001; Riehl 2014). In diesem Rahmen stehen die Erfassung und Typologisierung von deutsch-tschechischen *pattern replications* im Mittelpunkt, die analog zum deutsch-polnischen Teilprojekt untersucht und mit dessen Ergebnissen verglichen werden können. Das LangGener-Projekt schließt dadurch an die bisherige Forschung zum intensiven slawisch-germanischen Sprachkontakt von Bilingualen an, deren Untersuchung mit Blick auf die Typen von deutsch-tschechischen Sprachkontaktphänomenen eine typologische (zum deutsch-tschechischen bilingualen Sprachvergleich und Typologie vgl. Nekula et al. 2013) und mit Blick auf die Fragen der alters- bzw. zeitbedingten Stabilisierung von Sprachen, der umwelt- und zeitbedingten Spracherosion sowie der Fossilisierung der Zweitsprache Deutsch auch eine psycholinguistische Dimension hat. Durch die Erfassung und Deutung von Sprachwandelprozessen *in statu nascendi* leistet das LangGener-Projekt einen Beitrag auch zur historischen Linguistik.

Im 6. Abschnitt präsentieren wir Typen von Isolations- und Kontaktphänomenen im Tschechischen, wie Restrukturierung des Kasussystems, Aspektabbau oder Anzeichen für die Herausbildung eines Artikels und der gebundenen Wortfolge, die man auf die Erosion des Tschechischen und die deutsche Modellsprache zurückführen kann und bei denen wir uns die Frage stellen, ob und welche Unterschiede hier bei SP und MI festzustellen sind. Damit ist der vorliegende Text der Spracherosionsforschung (*attrition research*) zuzuordnen, in der es um den partiellen Verlust der L1 unter dem Einfluss der L2 geht (Besters-Dilger 2013: 192). Dies ist mit Blick auf die Gruppe (c) dahingehend zu spezifizieren, dass die einstige primäre Sprache zugunsten der Umgebungssprache zur sekundären bzw. schlafenden Sprache (Green 1998) geworden ist. Den partiellen Verlust führt Montrul (2008) auf den reduzierten Sprachinput zurück, der aufgrund des monolingualen Sprachmodus bei der Gruppe (c) kleiner ist als bei Gruppe (d), für die der bilinguale Sprachmodus charakteristisch ist. Durch den reduzierten Input wird die Sprache (Tschechisch in Deutschland) nicht hinreichend aktiviert und der „Schwellenwert“ (Paradis 2004) für ihre Verwendung steigt, weil einzelne Aspekte nicht (gleich) erinnert werden können. Dies führt zu Einschränkungen im Bereich der Performanz (Montrul 2008: 65) infolge der Lockerung bzw. Destabilisierung der einstigen kognitiven Verfestigung. Dies manifestiert sich in Isolationsphänomenen wie Wortfindungsprobleme, Codeswitching oder Nivellierung von Varietäten, Registern und Funktionsstilen, ohne dass sich dadurch das System der einst erworbenen Sprache grundsätzlich ändert. Andere nehmen an (zur Literatur vgl. Riehl 2014: 90, 93), dass die Isolations- sowie Kontaktphänomene (zur Unterscheidung vgl. Krefeld 2004: 39–45) Zeichen des umweltbedingten Vergessens (Riehl 2014: 89) sind, zu (partiellen) Restrukturierungsprozessen führen (können), wie dies beim Abbau der Deklination von Eigennamen der Fall ist, und generationsübergreifend den (kollektiven) Sprachverlust zur Folge haben.

Die Unterscheidung von Isolations- und Kontaktphänomenen ist dabei zum Teil schwierig. So kann die Reduktion des Kasussystems einerseits als Isolationsphänomen verstanden werden, indem nicht die Kategorie als solche in Frage gestellt wird, sondern nur ihre Anwendung auf wenig frequente Wörter durch fehlende Praxis Schwierigkeiten bereitet. Andererseits kann eine Kasusreduktion, wie sie bei den russischen Herkunftssprecherinnen und -sprechern im englischsprachigen Umfeld vorliegt (Polinsky 2006), als Resultat des Sprachkontaktes verstanden werden, bei dem ein einfacheres (englisches,

deutsches) System der Modellsprache auf ein komplexeres (russisches, tschechisches) System der Replikasprache der Sprachbenutzerinnen und -benutzer einwirkt.

Das Ausmaß der Spracherosion hängt vom Sprachinput ab, der je nach Familientypus (Romaine 1995), Netzwerkbildung, Kontakt zum Herkunftsland und Verwendung von Medien – sei es dauerhaft oder revitalisierend – variiert. Die Spracherosion (Attrition) ist jedoch auch vom (relativen) Migrationsalter (Besters-Dilger 2013: 192) und damit vom Sprachinput und der Stabilisierung der erworbenen Sprache vor der Migration oder Ausreise abhängig und setzt bei den erwachsenen MI in der Regel erst nach 10 Jahren (Besters-Dilger 2013: 194) ein und vertieft sich danach. Dass es sich somit um *forgetters* unterschiedlichen Ausmaßes handelt, ist in der Forschung nicht neu. Neu ist dagegen die Verbindung der Spracherosion mit Sprachideologien und Sprachmanagement in Sprachkontaktsituationen (Nekvapil & Sherman 2009), auf die auch die Aufteilung der Gruppen in SP und MI abzielt.

Von den *forgetters* der L1 sind die *incomplete learners* der L1 zu unterscheiden (u. a. Bester-Dilger 2013), die der Gruppe (e) zuzuordnen sind und die man mit der Gruppe der *forgetters* im Hinblick auf die Frequenz und Ausprägung von Isolations- sowie Kontaktphänomenen vergleichen kann (Rakhilina et al. 2016). Von den *forgetters* unterscheiden sich *incomplete learners* allerdings dadurch, dass sie die L1 aufgrund des geringen Impulses und Inputs so fragmentarisch erworben haben, dass dies bei der L1 unter dem indirekten Einfluss der L2 zu grundsätzlichen Restrukturierungsprozessen etwa beim Kasussystem (Restrukturierung des sechs- bzw. siebengliedrigen Kasussystems in ein zwei- bzw. dreigliedriges) führen kann. Diese Abgrenzung, die auf den Abschluss der sensitiven Periode bezogen wird bzw. sich des 12. Lebensjahres (Riehl 2014: 91) bedient, ist kaum scharf zu ziehen. Vielmehr ist sie als Kontinuum zu verstehen. Das Migrationsalter müsste daher auch bei Erwachsenen differenzierter gesehen werden, worauf auch die Auswahl der interviewten Personen innerhalb der Gruppe der SP und MI und die darauf basierende Auswertung der Ergebnisse im deutsch-tschechischen Teilprojekt abzielen.

Als *incomplete learners* sind in Bezug auf Deutsch neben den tschechischen MI (Gruppe d) auch einige SP (Gruppe c) einzuordnen, an die von ihnen um oder vor 1935 geborenen Eltern zwar keine Reliktvarietät des Deutschen weitergegeben wurde, die jedoch infolge des unvollständigen Spracherwerbs des Deutschen als Ergebnis des Drucks der Umgebungssprache im Begriff war zu entstehen. Nach der Aussiedlung konnten die SP mit Deutsch als L1 ihre Erstsprache – in Abhängigkeit vom bisherigen Erwerb des Deutschen sowie vom Migrationsalter – reaktivieren und weitgehend vollständig erwerben. Der Erwerb des Deutschen als L2 trifft prototypisch für die tschechischen MI (Gruppe d) zu, wobei sich die *incomplete acquisition* der L2 durch junge erwachsene MI grundsätzlich von der *incomplete acquisition* der L1 der Herkunftssprecher im frühkindlichen oder im Vorschulalter (Gruppe e) unterscheidet (vgl. Vergleich in Rakhilina et al. 2016). Die damit verbundenen Sprachkontaktphänomene werden zwar im 6. Abschnitt des vorliegenden Textes nicht diskutiert, das deutsch-tschechische Teilprojekt ist jedoch diesbezüglich aktiv und auch in der Erforschung des Zweitspracherwerbs zu verorten, bei dem die L2 die Replikasprache der L1 als Modellsprache ist.

Die Studien zur migrationsbezogenen Spracherwerbs- und Sprachverlustforschung, die sich zum Teil dokumentarische Ziele setzen, sind des Öfteren deskriptiv-typologisch geprägt. Sie gehen von authentisch erhobenen Sprachdaten aus und fokussieren im gewonnenen Material die typologisch bedingten Isolations- und Kontaktphänomene, die

sich in Studien zu *forgetters* und/oder *incomplete learners* unterschiedlicher slawischer Sprachen im Umfeld germanischer Sprachen wiederholt finden (zum Beispiel Polinsky 2006, 2009, 2018; Besters-Dilger 2013; Dubinina & Polinsky 2013; Hansen 2018 u. a.). Diese können als Ausgangspunkt und Orientierung bei der Analyse der erhobenen sprachlichen Daten dienen. Daraus ergeben sich sprachübergreifend vergleichbare Abweichungen von der Baseline slawischer Sprachen, wie bei der Valenz, den Präpositionalphrasen, der nicht-kanonischen Subjektposition, der Wortfolge, der Kongruenz, der Negation, den reflexiven Verben und den Possessivpronomina sowie bei der Kasusreduktion oder dem Aspektabbau bzw. den Anzeichen eines Artikels.

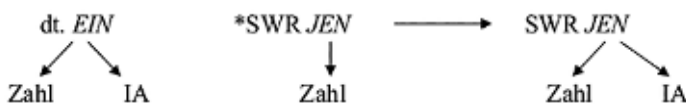
Diese Phänomene sollen anschließend in die universelle Typologie syntaktischer *pattern replications* eingeordnet werden, die von Heine & Kuteva (2005) bzw. Kuteva & Heine (2012) für einen intensiven und langanhaltenden Sprachkontakt entwickelt und von Hansen (2018) auf Herkunftssprachen angewandt wurde. Der vorliegende Beitrag überträgt diese Typologie von den Herkunftssprachen auf die Spracherosion der Erstsprache sowie den Zweitspracherwerb unter den Bedingungen eines intensiven und langanhaltenden Sprachkontakts, um neben der Feststellung von Differenzen auch die Universalität des theoretischen Modells sowie der Wandelprozesse zu testen.

Hansen (2018: 130) unterscheidet drei Typen von *pattern replications*:

1. sprachkontaktinduzierte Grammatikalisierung
2. Polysemiekopierung (grammatischer Funktionen)
3. Restrukturierung (d. h. Umgang mit dem Verlust)

Bei der Grammatikalisierung, die durch den Sprachkontakt ausgelöst bzw. unterstützt werden kann, verändern sich lexikalische in grammatische bzw. weniger grammatische in stärker grammatische Formen. Das zeigt Hansen (2018: 131f.) an der Grammatikalisierung der Präposition *za* zur Infinitivpartikel in der Sprache kroatischer Herkunftssprecher im deutschsprachigen Umfeld, die die Tendenzen im gesprochenen Kroatischen verstärkt. Bei der Polysemiekopierung (grammatischer Funktionen) erhält das bestehende Element eine neue (grammatische) Funktion, was Hansen (2018: 132f.) am Moliseslawischen ausführt, in dem unter dem Einfluss des Italienischen bei dem Imperfektiv neben der temporal-aspektuellen Funktion auch die Funktion des Konditionals aufkommt, wobei die Funktion des Konditionals in der Modell- und Replikasprache identisch ist, der Prozess abrupt geschieht und das Niveau der Grammatikalisierung ebenfalls dasselbe ist.

Durch diese Abruptheit unterscheidet sich die Polysemiekopierung grammatischer Funktionen von der Grammatikalisierung, für welche die Polysemiekopierung als Übergangsstufe zur Grammatikalisierung eine Rolle spielt. So kommt es bei der Entstehung des Artikels im Obersorbischen nach dem Vorbild des Deutschen zunächst zur Polysemiekopierung bei den Lexemen *jen* ‚ein‘ und *ten* ‚der‘, die schließlich grammatikalisiert wird, wobei sich – im Unterschied zum Deutschen – bei dem bestimmten Artikel die grammatische Bedeutung auf Kosten des Demonstrativpronomens durchsetzt:



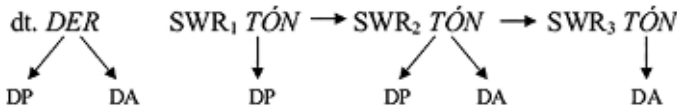


Abbildung 1: Polysemiekopierung im Obersorbischen (Scholze 2006: 150, 167)

Die Differenz zur Polysemiekopierung grammatischer Funktionen wird am Tschechischen deutlich, in dem das deiktische Lexem *ten* DP ‚der‘ im 19. Jahrhundert nach dem Vorbild des Deutschen polysem (DP ‚dieser‘ und DA ‚der‘ bzw. ‚der.DEM‘ und ‚der.DET‘) wird, ohne dass diese Innovation grammatikalisiert wird, bis sie letztendlich in den gesprochenen Substandard abgedrängt wird, in dem sie schließlich als ein die Nähe indizierendes Kontaktmittel genutzt wird (vgl. die Diskussion in Berger 1993 und in Nekula 1997).

Die Restrukturierung macht sich beim Umbau der Wortfolge, der Schwächung oder gar dem Verlust der mehrfachen Negation sowie der Kasusreduktion und dem Aspektabbau bemerkbar, die Polinsky (2006, 2008) am Beispiel des amerikanischen Russisch diskutiert, das russische Herkunftssprecher der „Generation Anderthalb“ nutzen. Die Kategorie des Kasus wird in Abhängigkeit vom Grad des unvollständigen Spracherwerbs von sechs auf zwei Subkategorien (Substantiva) bzw. zwei oder drei Subkategorien (frequenter Pronomina) restrukturiert, während häufig vorkommende Wortformen in Präpositionalphrasen als „chunks“ lexikalisiert werden (Polinsky 2006: 214–221). Die Restrukturierung ist auch für den Aspekt festzustellen, bei dem grammatische Aspektpaare durch Lexikalisierung aufgelöst werden (Polinsky 2006: 226–233).

6 Erste Ergebnisse in Bezug auf die Erosion des Tschechischen

Aus Platzgründen beschränken wir uns in diesem Beitrag auf die Konkretisierung der Spracherosion des Tschechischen bei deutsch-tschechischen Bilingualen, die – teilweise vom Tschechischen isoliert – seit mehr als zwanzig Jahren in Deutschland leben. Ausgehend von Abweichungen von der Baseline des gesprochenen Tschechischen im Heimatland (weiter nur Heimatland-Baseline) werden Daten präsentiert, die als Isolations- und Kontaktphänomene gelten können, die in den oben zitierten Studien zu Herkunftssprachen in Bezug auf den unvollständigen Spracherwerb präsentiert wurden. Es wird von Interesse sein, ob und welche Typen von Isolations- und Kontaktphänomenen bei beiden Gruppen der Interviewten auch bei der Spracherosion zu finden bzw. wie diese darin ausgeprägt sind. Diese werden anschließend mit Blick auf die Typologie von *pattern replications* nach Hansen (2018) interpretiert, wobei sich die Frage stellt, welche von den in dieser Studie zugrunde gelegten Typen – je nach der Gruppe der Interviewten und innerhalb der Gruppen der Interviewten – auch bei der Spracherosion zu finden sind und ob bzw. inwieweit die Typologie, die anhand der herkunftssprachlichen Daten entworfen wurde, auch bei der Spracherosion anwendbar ist. Diese werden auch im Hinblick auf die Frequenz bei den SP, die zum monolingualen Sprachmodus tendieren, und bei den MI, die zum bilingualen Sprachmodus tendieren, verglichen.

Neben den morphosyntaktischen PATs, denen in diesem Abschnitt vornehmlich die Aufmerksamkeit gewidmet werden soll, wurden auch andere Arten von Isolations- und

Kontaktphänomenen festgestellt, die an dieser Stelle jeweils kurz angesprochen werden. Die Beispielklausen aus dem Projektkorpus werden mithilfe der *Leipzig Glossing Rules* (Bickel et al. 2015) glossiert und um ein Äquivalent in der Baseline ergänzt. Da die Baseline-Äquivalente zu der authentischen „Bilingualen-Sprache“ primär konstruiert und erst sekundär in den Korpora und/oder durch Baseline-Sprecher validiert werden, werden diese Äquivalente in den Beispielen als „BaselinKa“ bezeichnet. Bei jedem Beispiel wird außerdem die Quelle aufgeführt, die das Kürzel der Sprechergruppe enthält, in die die betreffenden Personen gehören (SP bzw. MI).

Eines der typischen Anzeichen für Spracherosion sind Wortfindungsprobleme (Olsh-tain & Barzilay 1991). Eine der Strategien, die eigene Erzählung trotzdem fortzusetzen, ist Codeswitching, das heißt ein Wechsel von einer Sprache zu einer anderen. Dieser Sprachwechsel wird bei den von Spracherosion betroffenen Bilingualen oft durch Re-starts, Hästitationsgeräusche, Unterbrechungen sowie auch Reformulierungen eingeleitet, weshalb man in solchen Fällen auch von „flagged code-switching“ spricht (Bolonyai 2009: 260). Dies ist in (5) auch in unseren Daten der Fall:

- (5) a já jsem dělala ## ## # _ uklidit ne **Zimmermädchen** jo ?¹
 und 1SG AUX mach.3SG.PST _ aufräum.INF NEG Zimmermädchen ja
*und ich machte ## ## # _ aufräumen nicht **Zimmermädchen** ja ?*
 (LS_ZEI_SP_CZ)

Darin macht sich die Latenz der deaktivierten, sekundär gewordenen Erst- oder der einstigen primären Sprache bemerkbar. Codeswitching wird in unserem Korpus wiederholt auch durch die tschechische metasprachliche Formel *jak se řekne (česky) XXX* mit der Bedeutung ‚wie sagt man (auf Tschechisch) XXX‘ (vgl. (6)) eingeleitet bzw. durch die aus dem Deutschen lehnübersetzte tschechische Variante *XXX jak se to jmenuje (česky)* mit der Bedeutung ‚XXX wie heißt es (auf Tschechisch)‘ (vgl. (7)) eingeordnet. Diese Formel geht von einem bilingualen Gegenüber aus und das eingeführte deutsche Wort ist in diesem Rahmen nicht als eine Ad-hoc-Entlehnung zu verstehen:

- (6) # my jsme třikrát s* # **jak se to řekne (angesucht)?** žád* žádali
 1PL AUX dreimal wie REFL DEM sag.3SG angesucht ansuch.3PL.PST
 za to povolení
 für DEM Genehmigung
*wir haben dreimal **wie sagt man das angesucht?** um die Genehmigung angesucht*
 (IL_BLA_SP_CZ)

- (7) m* můj táta # von byl # # _ u **Eisenbahn** **jak se (to)**
 1SG.POSS Papa 3SG sein.3sg.pst bei Eisenbahn wie REFL DEM
 jmenuje jo?
 heiß.3SG ja
*mein Papa war bei der **Eisenbahn** wie heißt es ja?*
 (LS_ZEI_SP_DE)

¹ Zu Transkriptionsregeln und Sonderzeichen siehe Bučková & Prawdziej in diesem Band.

Auch bei (8a) kann von Codeswitching ausgegangen werden, das auf der Ebene der Morphosyntax (unmarkierte Form statt Vokativendung) erkennbar wird. Der Grund dafür ist die Einleitung der Klausel, die eine Wiedergabe einer direkten Rede signalisiert (*a říkali* ‚und sie sagten‘). Diese paraphrasiert auf Tschechisch einen Dialog, der in der Realität auf Deutsch stattfand und bei der Anrede zitiert wird. Im Tschechischen müsste der Vorname in der Anrede im Vokativ stehen (vgl. (8b)). Die Respondentin bleibt jedoch bei der unmarkierten Form bzw. Grundform im Nominativ:

(8) a. **Bilingualen-Sprache**

a	říkali	Krista	tvoje	děti	jsou
und	sag.3PL.PST	Krista.NOM	2SG.POSS.NOM.PL	Kind.NOM.PL	COP.3PL
už	velký	teďka			
schon	groß.NOM.PL	jetzt.DIM			
(LS_ZEI_SP_CZ)					

b. **BaselinKa**

a	říkali	Kristo	tvoje	děti	jsou
und	sag.3PL.PST	Krista.VOC	2SG.POSS.NOM.PL	Kind.NOM.PL	COP.3PL
už	velký	teďka			
schon	groß.NOM.PL	jetzt.DIM			
<i>und sie sagten Krista deine Kinder sind schon groß jetzt</i>					

Mit den Wortfindungsschwierigkeiten, die Codeswitching auslösen können, hängen ebenfalls PATs auf der lexikalischen Ebene zusammen, wie in (9). Die Sprecherin in (9a) bildet im Tschechischen das Wortgefüge *díl města*, was dem deutschen Kompositum *Stadtteil* entspricht. In der Heimatland-Baseline (9b) verwendet man jedoch in diesem Kontext nicht das Lexem *díl*, sondern den synonymen Ausdruck *část*, wobei für das deutsche *Stadtteil* auch das tschechische Derivat *čtvrť* ‚Viertel‘ steht:

(9) a. **Bilingualen-Sprache**

to	byl	potom	novej	díl	města	jako
das	sein.3SG.PST	dann	neu.M	Teil	Stadt.GEN	wie
(IL_REG_SP_CZ)						

b. **BaselinKa**

to	byla	potom	nová	část	města	jako
das	sein.3SG.PST	dann	neu.F	Teil	Stadt.GEN	wie
<i>das war dann praktisch ein neuer Stadtteil</i>						

Sprachkontakterscheinungen im Bereich der lexikalischen Semantik betreffen ebenfalls die Verwendung tschechischer Lexeme, deren Bedeutung anhand eines deutschen Musters erweitert wird. Im Unterschied zum Deutschen hat beispielsweise das Verb *moci* ‚können‘ in der Heimatland-Baseline des Tschechischen nur die Bedeutung ‚die Möglichkeit haben, etwas zu tun‘. Die Bedeutung ‚die Fähigkeit haben, etwas zu beherrschen‘ trägt im Tschechischen das Verb *umět*. Dieser Unterschied wird von der Respondentin in (10a) zugunsten der Kopierung des deutschen Polysemiemusters aufgegeben:²

(10) a. **Bilingualen-Sprache**

vona	ještě # če* # u* _	noviny	může	číst
3SG	noch	Zeitung	könn.3SG	les.INF
(LS_ZEI_SP_DE)				

b. **BaselinKa**

vona	ještě	umí	číst	noviny
3SG	noch	könn.3SG	les.INF	Zeitung
sie kann noch Zeitung lesen				

Ähnlich wird die Bedeutung des Verbs *přijít*, das in der Heimatland-Baseline nur die Bedeutung ‚gehend kommen‘ hat, nach dem Vorbild des deutschen Verbs *kommen* in der Sprache der Bilingualen um die Bedeutung ‚fahrend kommen‘ erweitert (11a). Im Unterschied zum Tschechischen ist nämlich das deutsche Verb bezüglich der Art der Bewegung unsensibel und schließt die Bedeutungen ‚gehend kommen‘ und ‚fahrend kommen‘ mit ein, sodass es im Vergleich zum Tschechischen als polysem erscheint und diese Polysemie in der Sprache der Bilingualen reproduziert wird, während die Bedeutung ‚fahrend kommen‘ in der Heimatland-Baseline durch ein eigenständiges Verb *přijet* ausgedrückt wird.³

(11) a. **Bilingualen-Sprache**

ale	vona	přijde # _	s vlakem	sem
aber	3SG	geh.3SG	mit Zug.INS	her
(LS_ZEI_SP)				

b. **BaselinKa**

ale	sem	vona	přijede	vlakem
aber	her	3SG	komm.3SG	Zug.INS
sie kommt mit dem Zug her				

Diese Polysemiekopierung ist auch bei den Bilingualen verbreitet, die das Tschechische im (früh)kindlichen Alter erworben haben und für die Deutsch im weiteren Spracherwerb zur primären Sprache geworden ist (am Beispiel Franz Kafkas vgl. Nekula 2003: 291f.; 2016: 101f.).

Weitere Abweichungen auf der lexikalischen Ebene, die mit Wortfindungsproblemen zusammenhängen und vom partiellen Sprachverlust zeugen, betreffen die Wortbildung und Kollokationen. Diese sind nicht immer auf ein Muster der deutschen Sprache zurückzuführen, wie bei der Wortbildung in (12). Darin wird anstatt *inzertní* (12b) das Wort *inzerátní* (12a) gebildet und verwendet, das im Tschechischen in dieser Form nicht existiert, auch wenn es nach den Wortbildungsregeln des Tschechischen gebildet ist. Darin al-

² Die Klausel weicht außerdem in der Wortstellung innerhalb der Verbalphrase von der Heimatland-Baseline ab (vgl. (10b)). Dieser Aspekt soll an dieser Stelle nicht weiter ausgeführt werden.

³ Die Klausel weicht außerdem in der Wortstellung (*sem* wird in der Heimatland-Baseline nicht als trennbares Präfix, sondern als Enklitikum platziert) und durch die analytische Bildung der instrumentalischen Bedeutung (*s vlakem* statt *vlakem*) von der Heimatland-Baseline ab (vgl. (11b)). Zu diesen Aspekten siehe unten.

terniert die Wortbildungsbasis einerseits für das Verb **{inzer}** {ova} {t} und das Substantiv **{inzer}** {a:t} {Ø}, andererseits für das Adjektiv **{inzer}** {n} {i:}, wobei diese Alternation offensichtlich nicht ausgeprägt ist und unfiziert bzw. infolge der Wortfindungsprobleme nicht genutzt wird. So hängt die Respondentin – analog zu *prodej* ‚Verkauf‘ – *prodejní* ‚Verkaufs-‘ oder *trest* ‚Strafe‘ – *trestní* ‚Straf-‘ – das Adjektivsuffix *-n(i)* an das Substantiv *inzerát* ‚Anzeige‘ – *inzerátní* ‚Anzeigen-‘ an:

(12) a. **Bilingualen-Sprache**

tady	jsme	měli	takovej	bezplatnej	inzerátní
hier	AUX	hab.1PL.PST	solch.ACC.SG	kostenlos.ACC.SG	Anzeige.ADJ.ACC
časopis					
Zeitschrift					
(NT_GAU_MI_CZ)					

b. **BaselinKa**

tady	jsme	měli	takovej	bezplatnej	inzertní
hier	AUX	hab.1PL.PST	solch.ACC.SG	kostenlos.ACC.SG	Anzeige.ADJ.ACC
časopis					
Zeitschrift					
<i>hier hatten wir so eine kostenlose Anzeigenzeitschrift</i>					

Die Einschränkung in der Wortbildung ist markant sichtbar auch bei der Austauschbarkeit der Nennformen des Adjektivs und des Adverbs in der Sprache der Bilingualen (13a), die sich im Tschechischen durch die Konversion des Adjektivs in das Adverb (bei *dobrý* ‚gut. ADJ‘ in *dobře* ‚gut.ADV‘) unterscheiden und in der Heimatland-Baseline daher die Form *dobře* (13b) zu erwarten wäre, während es im Deutschen nur eine Nennform (hier *gut*) gibt, die in der prädikativen und adverbialen Verwendung unverändert bleibt und in der attributiven Verwendung flektiert wird:⁴

⁴ Diese Interpretation unterstützt auch die fehlende Kongruenz des nominalen Teils des Prädikats mit dem Subjekt, in dem (in voller Lautung der Stelle) ebenfalls Abweichungen von der Norm für die Kongruenz in der Heimatland-Baseline festzustellen sind, wobei die futurale Perspektive der konditionalen *aby*-Konstruktion hinter den Verben des Wollens durch das Futur (*budeme* + INF ‚werden.1PL. + INF‘) supplementiert wird. Auffällig ist auch die Aufgabe der Konjunktion *aby* zugunsten der Konjunktion *že* nach den Verben des Wollens im Hauptsatz. Auch hier scheint die deutsche Norm auf die Sprache der Bilingualen einzuwirken, in der die Konjunktion *dass* auf die Verben des Wollens nicht sensibel und im Vergleich mit dem Tschechischen in diesem Sinne polysem ist. Diese Polysemiekopierung kann allerdings auch als Vertiefung der Grammatikalisierung von *že* bzw. als Verwendung von *že* als ‚Primitiv der Subordination‘ (vgl. Rakhilina et al. 2016: 7) verstanden werden (für das Tschechische vgl. Nekula 1997). Dafür würden Beispiele sprechen, in denen das finale *aby*, dem im Deutschen die Konjunktion *damit* entspräche, ebenfalls durch *že* ersetzt wird: *a # ale voni to dělali k* _ kvůli nám děti že to děti mají lepší* ‚aber sie haben es wegen uns Kindern gemacht, **damit** die Kinder es besser haben‘ (LS_ZEI_SP_CZ), wobei in diesem Beispiel in *kvůli nám děti* (wegen 1PL.DAT Kinder.NOM) zudem auch die Kongruenz innerhalb der Nominalphrase aufgegeben wird, die in *kvůli nám dětem* (wegen 1PL.DAT Kinder.DAT) in der Heimatland-Baseline auch im Gesprochenen zu erwarten wäre.

(13) a. **Bilingualen-Sprache**

chtěli	že_	budeme #	dobrý _ #	mluvit	ně* # německy
woll.3PL	dass	werd.FUT.1PL	gut.ADJ	sprech.INF	Deutsch

(LS_ZEI_SP_CZ)

b. **BaselinKa**

chtěli	abysme	dobře	mluvili	německy
woll.3PL	dass.1PL	gut.ADV	sprech.PTCP.PL	Deutsch

*sie wollten, dass wir **gut** deutsch sprechen*

Wie bei (11) ist auch diese Polysemiekopierung bei jenen Bilingualen verbreitet, die das Tschechische neben dem Deutschen im (früh)kindlichen Alter erworben haben, wobei Deutsch für sie im weiteren Verlauf zur primären Sprache geworden ist (am Beispiel Franz Kafkas vgl. Nekula 2003: 284; 2016: 98f.).

Die Spracherosion betrifft bei einigen Sprecherinnen und Sprechern nicht nur die lexikalische Semantik und Wortbildung, sondern auch die Aussprache. Im Beispiel (14a) sind gleich zwei Abweichungen von der im Tschechischen üblichen Aussprache festzustellen. Erstens wird der Vokal /ɛ/ in *ale* ‚aber‘ von dem Interviewten zwar nicht in voller Analogie zu deutschen Ausspracheregeln in unbetonter Position am Wortende als [ə] reduziert, aber dennoch zu Mitte als [ě] zentralisiert ausgesprochen. Zweitens wird das Wort *kostel* ‚Kirche‘ in Anlehnung an die Existenz des deutschen Konsonantenclusters /kʃt/ im Unterschied zur Heimatland-Baseline als [kʃtɛl] ausgesprochen:⁵

(14) a. **Aussprache der Bilingualen**

a	bylo	to	ale	vod #	koštela _	katolickej [internát]
[a	bɪlɔ	tɔ	alě	vɔt ə	kʃtɛla	katɔlʃtskej]
und	sein.3SG.PST	es	aber	von	Kirche.GEN	katholisch.M.NOM

(II_ROD_SP_CZ)

b. **BaselinKa**

a	bylo	to	ale	od #	kostela _	katolickej [internát]
[a	bɪlɔ	tɔ	alɛ	ɔt ə	kɔstɛla	katɔlʃtskej]
und	sein.3SG.PST	es	aber	von	Kirche.GEN	katholisch.M.NOM

*und es war aber von einer **Kirche** ein katholisches [Internat]*

Im Folgenden werden nun die in unserem Korpus auffindbaren Beispiele für PATs aus dem Bereich der Morphosyntax unter die sprachübergreifenden Typen der Isolations- und Kontaktphänomene eingeordnet, die in Polinsky (2006, 2008, 2018), Besters-Dilger (2013), Dubinina & Polinsky (2013), Blaszczyk (2015), Hansen (2018) u. a. m. anhand der herkunftssprachlichen Daten erkennbar sind. In Bezug auf Hansen (2018) werden diese im oben angeführten Sinne mit Blick auf die Typen der Systemveränderung weiter analysiert.

⁵ Die Klausel (14a) enthält dabei auch weitere lexikalische und syntaktische Abweichungen, auf die aber an dieser Stelle nicht eingegangen wird. In Bezug auf die Heimatland-Baseline in (14b) wurden lediglich die die Aussprache betreffenden Aspekte dargestellt.

6.1 Restrukturierung des Kasussystems

Die vorgefundenen Abweichungen gegenüber der Nominaldeklinations-Baseline weisen in ihrer Gesamtheit auf eine Vereinfachung des komplexen synthetischen Kasussystems hin. Polinsky (2006: 214) spricht im Zusammenhang mit einer vergleichbaren Entwicklung bei russischen Herkunftssprecherinnen und -sprechern in den USA von einer Re-Analyse des Kasussystems und folglich auch der Deklinationsparadigmen. Gleichzeitig macht jedoch Polinsky (2006: 214) darauf aufmerksam, dass derartige ungrammatische Formen nicht zwangsläufig durch ein fehlendes Grammatikwissen (Kompetenz bzw. Repräsentation, vgl. Schmid & de Leeuw (2019: 183)) zu erklären sind, denn sie können nur Einschränkungen der Sprachproduktion (Performanz bzw. aktuelle Verarbeitungsprozesse, vgl. ebd.) widerspiegeln. Welcher Fall zutrifft, ermittelt sie durch Testen der rezeptiven Sprachkompetenz der interviewten Personen.

In jedem Fall stellt Polinsky (2006: 183, 198) bei Herkunftssprecherinnen und -sprechern die Tendenz fest, anstelle von einfachen, synthetischen Kasusmarkierungen Adpositionen zu verwenden. In unserem Material findet sich eine systematische Verwendung der Präpositionalphrase *od* bzw. *vod*⁶ ‚von‘ mit Genitiv (15a) statt eines einfachen Genitivs (15b) in der possessiven Funktion, was von einer Erosion zeugt, die in eine ähnliche Richtung geht. Die Präposition ersetzt zwar nicht die Kasusmarkierung, ihre Verwendung kann jedoch als eine Tendenz interpretiert werden, mehr analytische Konstruktionen statt einfacher synthetischer Kasusformen zu verwenden. Die präpositionale Konstruktion, die auf den Kontakt mit dem Deutschen zurückgeht, könnte man dann in Anlehnung an Kuteva & Heine (2012) als den ersten Schritt bei der sprachkontaktinduzierten Grammatikalisierung der Präposition *od* ‚von‘ als Possessivmarker interpretieren:

(15) a. **Bilingualen-Sprache**

jeden	z	nich	měl	dokonce		
ein.M	aus	3PL.GEN	hab.3SG.PST	sogar		
nějaký		akcie		vod	tý	firmy
irgendein.ACC.PL		Aktie.ACC.PL		von	DEM.GEN	Firma.GEN
(QC_MUN_MI_CZ)						

b. **BaselinKa**

jeden	z	nich	měl	dokonce		
ein.M	aus	3PL.GEN	hab.3SG.PST	sogar		
nějaký		akcie		tý	firmy	
irgendein.ACC.PL		Aktie.ACC.PL		DEM.GEN	Firma.GEN	
<i>einer von ihnen hatte sogar irgendwelche Aktien von der Firma</i>						

Diese Konstruktion ist nämlich in unserem Korpus auch in einer Reihe weiterer Beispiele vorhanden, wie *z rodiny vod jejího otce* (QC_WUN_SP_CZ), *děti od mojí sestřenice* (LS_ZEI_SP_CZ), *děti od ní* (LS_ZEI_SP_CZ), *měli _grab_ ## od od # o* ## ot' její _#_ babička* (LS_ZEI_SP_CZ), *sestřenice o* od mojí mámě* (LS_ZEI_SP_CZ), *sestra od babičce* (LS_ZEI_SP_CZ), *od mojí babičce* (LS_ZEI_SP_CZ) u. a. In diesen wird

⁶ *Vod* ist eine informelle Variante von *od* ‚von‘, ausgesprochen mit einem prothetischen *v*-.

zudem der Genitiv zugunsten der Nichtmarkierung bzw. des formalen Nominativs (*grab od její babička*) oder durch die Restrukturierung zugunsten des Dativs (*sestřenice od mojí mámě*) aufgegeben, und zwar eindeutig stärker in der Gruppe der SP, denen auch die Respondentin zuzuordnen ist, von der diese Beispiele stammen.

Die Tendenz zu mehr analytischen Konstruktionen zeigt sich auch an der Verwendung der Präposition *s* ‚mit‘ (16a) anstatt eines einfachen Instrumentals (16b):

(16) a. **Bilingualen-Sprache**

když jedu s autem
wenn fahr.1SG mit Auto.INS
(IL_REG_SP_CZ)

b. **BaselinKa**

když jedu autem
wenn fahr.1SG Auto.INS
wenn ich mit dem Auto fahre

Auch diese Konstruktion ist mehrfach belegt, wie in *a když _ u nás jsou nějaký # s autobusem když můžeš ject* (LS_ZEI_SP_CZ), *tak jsme jeli s autem* (LS_ZEI_SP_CZ), *jednou pojedou s busem nebo co nebo vlakem* (LS_ZEI_SP_CZ), *a my sme jeli s s tou metro* (LS_ZEI_SP_CZ), wobei auch bei dieser Konstruktion die Restrukturierung des Kasus-systems verbunden mit der Nichtmarkierung beim Substantiv *metro* in *my sme jeli s s tou metro* auffällt.

Es geht aber nicht nur um die Durchsetzung von analytischen Konstruktionen nach dem Vorbild des Deutschen, wo der Instrumental nicht vorhanden ist und die Bedeutung entsprechend durch die Adposition *mit* zum Ausdruck kommt, bzw. der possessive Genitiv im Gesprochenen durch die Konstruktion mit der Adposition *von* ersetzt bzw. der Genitiv im Allgemeinen geschwächt und durch den Dativ ersetzt wird (von Polenz 1999: 345; zu den Schwankungen in der Kasusmarkierung und -verwendung in der deutschen Gegenwartssprache vgl. auch Dürscheid (2007) und Vater (2009)). Eine ähnliche Tendenz zur Schwächung des Genitivs im Sprachsystem des Tschechischen macht sich auch bei den untersuchten Bilingualen bemerkbar. Als einer der Erklärungsansätze ist daher auch eine Akzeleration der ebenfalls in der Heimatland-Baseline zu verzeichnenden Prozesse in Betracht zu ziehen (vgl. Cope & Dittmann 2020). Die Genitiv-Schwächung tritt allerdings am stärksten beim Abbau des Genitiv Numerativ auf, wobei der Abbau in der Heimatland-Baseline kein Vorbild hat, während man ihn auf die Absenz des Genitiv Numerativ im Deutschen im Sinne eines negativen Transfers zurückführen kann:

(17) a. **Bilingualen-Sprache**

když chceš můžeš u mě pracovat půl den
wenn woll.2SG könn.2SG bei 1SG.GEN arbeit.INF halb Tag.NOM
(LS_ZEI_SP_CZ)

b. **BaselinKa**

jestli chceš můžeš u mě pracovat půl dne
wenn woll.2SG könn.2SG bei 1SG.GEN arbeit.INF halb Tag.GEN
wenn du willst kannst du bei mir halbtags arbeiten

Die Deaktivierung des Genitivs Numerativ durch die Lockerung der kognitiven Verfestigung manifestiert sich in den authentischen Daten in der Verlangsamung der Artikulation, den Restarts und den Häsitationsgeräuschen, wie in *jsem to _ dělala deset roky* (LS_ZEI_SP_CZ), *nedostal tak # moc peníze* (LS_ZEI_SP_CZ), *a a to s* to bylo moc peníze* (LS_ZEI_SP_CZ), *bydlí ne ? už padesát jedna _ roky* (LS_ZEI_SP_CZ), die dennoch nur ausnahmsweise zu seiner Aktivierung, wie in *tak jsme měli _ osm sto _ acht hundred já (1) o* o* _ osm set _ # # jídla každý den _ jo ?* (LS_ZEI_SP_CZ), führen.

Die bei den befragten Personen im Einzelnen unterschiedlich intensiv ausgeprägte Restrukturierung des Kasussystems zeigt sich dann in der Aufgabe von Instrumental (18), Genitiv (19) und Dativ (20) zugunsten des Nominativs bzw. Akkusativs, die Anzeichen eines dualen Kasussystems anstatt eines Kasussystems mit sechs Subkategorien erkennen lässt:

(18) a. **Bilingualen-Sprache**

ta	byl	dialekt	babička	to	asi	mluvila
dort	sein.3SG.PST	Dialekt.NOM	Oma.NOM	DEM.NOM	vielleicht	sprech.3SG.PST

(CI_BUB_SP_CZ)

b. **BaselinKa**

tam	byl	dialekt	babička	jím	asi	mluvila
dort	sein.3SG.PST	Dialekt.NOM	Oma.NOM	3SG.M.INS	vielleicht	sprech.3SG.PST

dort gab es einen Dialekt die Oma hat es vielleicht gesprochen

(19) a. **Bilingualen-Sprache**

_ vnoučata	_ se _ se	neptají	rodiče	_	ale	_
Enkel.PL	REFL	frag.3PL.NEG	Elter.PL.ACC	aber		
dědečka	a	babičku	jak	to		
Opa.ACC	und	Oma.ACC	wie	DEM		
tehďa	bylo					
damals	sein.3SG.PST					

(CI_BUB_SP_CZ)

b. **BaselinKa**

_ vnoučata	_ se _ se	neptají	rodičů	_	ale	_
Enkel.PL	REFL	frag.3PL.NEG	Elter.PL.GEN	aber		
dědečka	a	babičky	jak	to		
Opa.GEN	und	Oma.GEN	wie	DEM		
tehďa	bylo					
damals	sein.3SG.PST					

die Enkel fragen nicht die Eltern, sondern den Opa und die Oma wie es damals war

(20) a. **Bilingualen-Sprache**

a	já	já	jsem	to	nerozuměla
und	1SG	1SG	AUX	DEM.ACC	versteh.3SG.PST

(KN_MUN_MI_CZ)

b. **BaselinKa**

a já jsem **tomu** nerozuměla
 und 1SG AUX DEM.ACC versteh.3SG.PST
und ich habe es nicht verstanden

Die Beispiele (18), (19) und (20) lassen sich dabei auch als Übertragungen deutscher Valenzmuster deuten. Auf diese Interpretation wird im Abschnitt 6.7 eingegangen.

In (21a) verwendet die Sprecherin nach der Präposition *na* ‚auf‘ den Akkusativ. Dies ist insofern eine Abweichung von der Heimatland-Baseline, als es sich um eine Verbergänzung mit statischer Bedeutung handelt, für die der Lokativ verlangt wird (vgl. (21b)). Bei (21a) geht es im Unterschied zu (18a), (19a) und (20a) um kein PAT, sondern um ein Beispiel eines indirekten Sprachkontakts, der in unserem Korpus als AA (andere Abweichung) annotiert wurde. Es ist anzunehmen, dass die Form *na návštěvu* die Funktion eines Nicht-Nominativ-Markers hat und daher vom Umbau des Kasussystems zeugt.

(21) a. **Bilingualen-Sprache**

jednou jenom to jsme byli _ na **návštěvu** v Mostě
 einmal nur DEM AUX sein.1PL.PST auf Besuch.ACC in Most.LOC
 (IL_BLA_SP_CZ)

b. **BaselinKa**

jednou jenom to jsme byli _ na **návštěvě** v Mostě
 einmal nur DEM AUX sein.1PL.PST auf Besuch.LOC in Most.LOC
einmal nur da waren wir zu Besuch in Most

Die Restarts und Hästitionsgeräusche scheinen dabei gerade für die Restrukturierung des Kasussystems infolge der Erosion charakteristisch zu sein, indem sie die Spuren des deaktivierten Wissens über alternative Ausdrucksformen erkennen lassen, sei es bei Akkusativ statt Genitiv, wie in *a pak jsem # zase šla do tu do tu kantynu* (LS_ZEI_SP_CZ), *a von byl fakt od pondělí do # pátek _ celej týden pryč* (LS_ZEI_SP_CZ), *a pak jsem já šla # do tý jiný _ # třídu* (LS_ZEI_SP_CZ) oder bei Nichtmarkierung statt Instrumental, wie in (25) – siehe Abschnitt 6.3.

6.2 Herausbildung des Artikels

Die folgenden Beispiele für die Verwendung von Indefinit- und Demonstrativpronomina deuten darauf hin, dass die Bilingualen im Tschechischen das deutsche Artikelsystem replizieren. In (22a) verwendet die Respondentin scheinbar das Indefinitpronomen *nějaký* ‚irgendein‘ in der Funktion eines unbestimmten Artikels. In der Heimatland-Baseline wird dagegen die Unbestimmtheit des bezeichneten Referenten durch die bloße Platzierung der Bezeichnung in der rhematischen Position am Ende der Äußerung hinreichend markiert, sodass die funktionale Verdoppelung der Unbestimmtheit durch lexikalische Mittel als Replik des morphosyntaktischen Musters des Deutschen, in dem die Wortstellung keine so bedeutende Rolle spielt, verstanden werden kann.

(22) a. **Bilingualen-Sprache**

a	nejtěžší	je	překládat	nějaký	vtip
und	schwierig.SUP ⁷	sein.3SG	übersetz.INF	irgendein.M	Witz

(CI_BUB_SP_CZ)

b. **BaselinKa**

a	nejtěžší	je	překládat	vtipy ⁸
und	schwierig.SUP	sein.3SG	übersetz.INF	Witz.PL

*und am schwierigsten ist es, **einen** Witz zu übersetzen*

Dieses Beispiel ist bei Weitem nicht isoliert, sondern bildet mit anderen Beispielen, wie (28), *tak s* v Aše v Chebu nedostal **nějakej** voblek pro pro kluka patnáctiletýho* (QC_WUN_SP_CZ), *ale moje ses* # moje holka vona bydlí # **_jednu** hodinu od tady weg _Passau tam _ne ?* (LS_ZEI_SP_CZ), *že jako_ # mi nemůžou nabídnout **nějaký** stálý místo* (BJ_MUN_SP_CZ) u. a. m., ein erkennbares Sprachkontakthänomen.

(23) a. **Bilingualen-Sprache**

ani	to	jedno	ani	to	
weder	DEM	ein.N	weder	DEM	
druh*	jo	tak			
zweit	ja	so			
dopisuju	mailem	jo_	vobčas	nějaký	věci
korrespondier.1SG	E-Mail.INS	ja	ab und zu	irgendein.PL	Ding.PL

(QC_MUN_MI_CZ)

b. **BaselinKa**

ani	jedno	ani	druhé	jo	tak
weder	ein.N	weder	zweit.N	ja	so
dopisuju	mailem	jo_	vobčas	nějaký	věci
korrespondier.1SG	E-Mail.INS	ja	ab und zu	irgendein.PL	Ding.PL

*weder **das** eine noch **das** andere ja also ich korrespondiere per E-Mail ab und zu irgendwelche Dinge*

In (23a) gebraucht der Respondent das Demonstrativpronomen *to* ‚das‘ scheinbar als bestimmten Artikel. In diesem Fall ist die Bestimmtheit des bezeichneten Referenten durch die Platzierung der Bezeichnung in der thematischen Position am Anfang der Äußerung hinreichend markiert, sodass die funktionale Verdoppelung der Bestimmtheit durch lexikalische Mittel ebenfalls als Replik des deutschen Musters verstanden werden kann.

Dies trifft auch auf die Replik des bestimmten Artikels zu, die sich auch in anderen Beispielen, wie *ale my sm* my jsme mohli s **tím** českým penízem něco koupit ne ?* (LS_ZEI_SP_CZ), *a ted'kon vždyt' vyklízím vod matky byt tak najdu_ **ty** s* **ty** papíry z dřívějška* (QC_WUN_SP_CZ), *a já jsem **ta** jediná která tam SM nechodí* (KG_MUN_MI_CZ)

⁷ Die Glosse SUP steht für den Superlativ und wurde im Inventar der *Leipzig Glossing Rules* (Bickel et al. 2015) von den Autoren dieses Aufsatzes ergänzt.

⁸ Aufgrund der imperfektiven Verbform wurde das Objekt *vtip* ‚Witz‘ in der BaselinKa (20b) in den Plural gesetzt.

u. a., finden lässt. In jedem Fall scheint bei den untersuchten Bilingualen die Tendenz, das kurze Demonstrativum *to* im gesprochenen Tschechisch zur Sicherung der Referenz und Verständigung zu verwenden, noch weiter verstärkt zu sein (zur Akzeleration im Allgemeinen vgl. Cope & Dittmann 2020). Diese Tendenz im Tschechischen wird von Berger (1993) diskutiert und als Erbe des intensiven deutsch-tschechischen Sprachkontaktes im 19. Jahrhundert interpretiert, das auf die Bilingualen zurückgeht, durch die die Herausbildung des bestimmten Artikels angestoßen wurde.

In beiden Fällen kann die artikelartige Verwendung der Pronomina als eine Polysemiekopierung interpretiert werden (Heine & Kuteva 2005: 100; Wiemer & Wälchli 2012: 27). Angesichts des Grammatikalisierungsprozesses im Obersorbischen, in dem sich durch den Kontakt mit dem Deutschen sowohl der bestimmte als auch der unbestimmte Artikel herausgebildet haben (siehe Abschnitt 5), kann auch im Falle des Tschechischen angenommen werden, dass die Polysemiekopierung bei den untersuchten Bilingualen der Anfang einer sprachkontaktinduzierten Grammatikalisierung ist.

6.3 Restrukturierung des Aspekts und der Determiniertheit bei Bewegungsverben

Während sich beim Nomen die Herausbildung der grammatischen Kategorie der Determiniertheit abzeichnet, scheint der verbale Aspekt durch einen Abbau gekennzeichnet zu sein. Für die Herkunftssprache Russisch in den USA stellt Polinsky (2006: 226) die Lexikalisierung des Aspekts fest, d. h. dass entsprechend dem Inhalt und der Frequenz des jeweiligen Verbs nur eine Form des Aspektpaars erworben und genutzt wird. Komplementär dazu scheint es, als ob sich in unserem Material auch bei der Erosion des Tschechischen eine Lexikalisierung des Aspekts – verbunden mit der Auflösung von Aspektpaaren und damit dem Abbau der grammatischen Kategorie des Aspekts – zeigen würde. Das bedeutet, dass die interviewten Personen aufgrund des Inhalts und der Frequenz nur eine Form des Aspektpaares beibehalten und sie anstelle der anderen Form nutzen, wie dies bei der Verwendung des imperfektiven Verbs anstelle des perfektiven Verbs in (24a) sichtbar wird, wobei dies auch in anderen Beispielen wie in *jo to to # nějaký jo _jo _já musím říkat* (LS_ZEI_SP_CZ) zu finden ist. Für das Durchsetzen und Beibehalten dieser Aspektform spricht ihre höhere Frequenz im gesprochenen Tschechisch, die durch die Anzahl entsprechender Tokens in den Korpora der gesprochenen Sprache bestätigt wird (Kopřivová et al. 2017):

(24) a. **Bilingualen-Sprache**

možná	že	bych	se z*	to	zase	učila
vielleicht	COMP	1.SG.COND	REFL	DEM	wieder	lern.1SG.COND.IPFV

(LS_ZEI_SP_CZ)

b. **BaselinKa**

možná	že	bych	se z*	to	zase	naučila
vielleicht	COMP	1.SG.COND	REFL	DEM	wieder	lern.1SG.COND.PFV

*vielleicht würde ich es wieder **erlernen***

Markant ist in diesem Zusammenhang auch der Gebrauch von analytischen Formen in Konstruktionen wie *dostat* ‚bekommen‘ + Inf., *udělat* ‚machen‘ + deverbales Substantiv oder *dát* ‚geben‘ + deverbales Substantiv, die Vollverben ersetzen und so auch die Einschränkungen im Bereich der Wortbildung des Verbs, die für den Aspekt von Bedeutung ist, offenlegen: *mám ? uděláme světlo že jo* (KL_HEI_SP_CZ), *a dáme ji německý # _ odpověď* (IL_BLA_SP_CZ), *tak v* vod tý všechno _ dostane vědět* (KL_HEI_SP_CZ) usw.

Aufgelöst scheinen auch Verbpaare der Bewegungsverben, bei denen in der Regel das indeterminierte zugunsten des determinierten Verbs verloren geht, wie in (25a). So wird auch für die Kontexte, die eine wiederholte Aktivität oder – wie in (25a) – eine Fähigkeit zu einer Art von Bewegung ausdrücken, die determinierte Form verwendet, die in der Heimatland-Baseline bei absichtlichen, zielgerichteten Prozessen gebraucht wird (Nübler 2017):

(25) a. **Bilingualen-Sprache**

# _ von	jel	traktor ⁹	velký	LKW
3SG	fahr.3SG.PST.DET	Traktor.ACC	groß.ACC	LKW
a	to	všechno j*		
und	DEM	alles		
	(LS_ZEI_SP_CZ)			

b. **BaselinKa**

# _ von	jezdil	traktorem	velkým	nákladřákem
3SG	fahr.3SG.PST.IDET	Traktor.ACC	groß.INS	LKW.INS
a	to	všechno j*		
und	DEM	alles		
	<i>er fuhr</i>	<i>Traktor einen großen LKW</i>	<i>und das alles</i>	

Während das beibehaltene determinierte Verb offensichtlich seine Bedeutung ausweitet und auch die Bedeutungen des indeterminierten Verbs abdeckt, bleibt die präfixale Futurbildung mit *po-*, die nur für determinierte Bewegungsverben charakteristisch ist, auch in der indeterminierten Bedeutung erhalten, wie in *ale moje máma tam _ vždycky pot* # po-jedou _ na návštěvu _ ne ?* (LS_ZEI_SP_CZ), *tak pojedou vždycky s sebou ne ?* (LS_ZEI_SP_CZ), *# jinak pojedeme jenom na jeden den a zase zpátky* (IL_BLA_SP_CZ) usw.

6.4 Restrukturierung des Tempussystems

Auch beim Tempus, dessen System im Tschechischen mit dem Aspekt eng verflochten ist (die Präsensform der perfektiven Verben hat eine futurale Bedeutung, sodass sie das analytische Futur ersetzt, das nur bei den imperfektiven Verben vorliegt), kommt es zum Abbau bzw. zur Restrukturierung der grammatischen Kategorie. So verwenden einige Interviewte das Präsens in der Bedeutung des Futurs, so wie in (26a). Bei einem imperfektiven Verb, wie *bydlet* ‚wohnen‘, würde man aber in der Heimatland-Baseline ein analytisches Futur mit *budu* + Infinitiv erwarten:

⁹ Zu der sprachkontaktbedingten Abweichung im Bereich Kasusmorphologie in diesem Beispiel siehe Abschnitt 6.1.

(26) a. **Bilingualen-Sprache**

vy	dva	bydlíte	u	mě	v	baráku
2PL	zwei	wohn.2PL.PRS	bei	1SG.GEN	in	Haus.LOC
tady	ty	dva	týdny			
hier	DEM	zwei	Woche.PL			

(KL_HEI_SP_CZ)

b. **BaselinKa**

vy	dva	budete	bydlet	u	mě	v	baráku
2PL	zwei	sein.2PL.FUT	wohn.INF	bei	1SG.GEN	in	Haus.LOC
tady	ty	dva	týdny				
hier	DEM	zwei	Woche.PL				

*ihr zwei **wohnt** jetzt in meinem Haus diese zwei Wochen*

Weitere Beispiele für die Verwendung der Präsensformen anstatt des Futurs sind *a _ možná že # **dostane mají** děti a my dáme pozor na to ne ?* (LS_ZEI_SP_CZ), *# tam se setkám s jistě s _ s lidma kterýma **musím musím** mluvit _ německy* (KC_ISM_MI_CZ), *dobře tak to _ zahodím tady do # do toho _ do tý plechovky a **je** to rozbitý* (KL_HEI_SP_CZ) u. a.

Markant ist dies auch bei der Futurbildung determinierter Verben mittels der Präfigierung der Präsensform durch das Präfix *po-*. Diese wird zwar, wie wir sehen konnten, noch genutzt, sie steht aber unter Anpassungsdruck und wird durch das einfache Präsens (16a) vedrängt: *když (**jdu**) tak už nemusím přijít a vůbec* (IL_REG_SP_CZ), *no jo kein p* žádněj problém _ **jdeme** s te* pr* zase s nima **jdem** domů* (KL_HEI_SP_CZ), *a von říká já bych chtěl ty komunistický_ svátky _ jednou zažit jsem říkal tak **jedem** v v květnu* (KL_HEI_SP_CZ), *tak _ jednou tam **jedeme*** (LS_ZEI_SP_CZ), *tak **jdeme** ještě něco pít* (KL_HEI_SP_CZ), *[a že jsem _ že jsme taky _ mohli _ učít # nějakou práci] ale _ maminka a tatínek řekl ty ted'ka máš práci [...] ted'ka **jdeš** do práce a hotovo* (TU_HAG_SP_CZ). In der Herkunftssprache Tschechisch in den USA wird dagegen die synthetische Form der Futurbildung mit *po-* bei determinierten Bewegungsverben gar nicht erworben und genutzt und im Einklang mit dem Englischen durch die produktive analytische Form, wie in *budú letět* ‚sie werden fliegen‘ statt *poletí* (vgl. Cope & Dittmann 2020), ersetzt, was mit der Restrukturierung des Tempussystems einhergeht.

6.5 Variation bei reflexiven Verben

In den untersuchten tschechischen Interviews finden sich Beispiele, in denen es bei *se* ‚sich‘ zu Abweichungen von der Heimatland-Baseline kommt. Dies betrifft Verben, die in der Heimatland-Baseline etwa inhärent (*dívát se* ‚schauen‘, *všimnout si* ‚bemerken‘) oder eigentlich (*mýt se* ‚sich waschen‘, *přát si* ‚sich wünschen‘) reflexiv sind (vgl. Paněvová & Karlík 2020), von den Bilingualen aber ohne Reflexiva *se/si* ‚sich‘, wie in (27a), verwendet werden. Auch werden *se* ‚sich.ACC‘ und *si* ‚sich.DAT‘, wie in (28a), vertauscht, bzw. nicht reflexive Verben, wie in (29a), reflexiv gebraucht:

(27) a. **Bilingualen-Sprache**

u těch závodů jsem taky podíval
 bei DEM Betrieb.GEN.PL AUX auch schau.3SG.PST
 (KL_HEI_SP_CZ)

b. **BaselinKa**

do těch závodů jsem se taky podíval
 in DEM Betrieb.GEN.PL AUX REFL auch schau.3SG.PST
also ich habe bei den Betrieben auch vorbeigeschaut

(28) a. **Bilingualen-Sprache**

a chcel jsem se domluvit na
 und woll.3SG.PST AUX REFL.ACC verabred.INF Auf
 nějakou schůzku
 irgendein.ACC Treffen.ACC
 (QC_MUN_MI_CZ)

b. **BaselinKa**

a chtěl jsem si domluvit
 und woll.3SG.PST AUX REFL.DAT verabred.INF
 nějakou schůzku
 irgendein.ACC Treffen.ACC
*ich wollte **mich** für ein Treffen verabreden*

(29) a. **Bilingualen-Sprache**

von si to doufal
 3SG REFL DEM hoff.3SG.PST
 (NT_GAU_MI_CZ)

b. **BaselinKa**

von v to doufal
 3SG in DEM hoff.3SG.PST
*er hat es **sich** erhofft*

Bei den Klausen (27) bis (29) wird jeweils anhand konkreter Lexeme (Verben) auf ein deutsches Muster zurückgegriffen. Insgesamt kommen in unseren Daten jedoch häufiger die Fälle vor, in denen ein in der Baseline inhärent oder eigentlich reflexives Verb ohne das Reflexivum gebraucht wird, zum Beispiel *a _ co měli o mně myslet ?* (CI_BUB_SP_CZ), *s tou pak můj manžel něco začal* (KG_MUN_MI_CZ), *vyučil jsem* (II_ROD_SP_CZ), *to jsem chtěla potom s dětma učit* (TU_HAG_SP_CZ), *a _ teď přejedeš hranice a myslíš hmm _ kde to bylo* (CI_BUB_SP_CZ). Belege für dieses Phänomen finden sich auch bei russischen Herkunftssprecherinnen und -sprechern in den USA (Polinsky 2006: 235), während bei der ersten Generation russischer Sprecherinnen und Sprecher in der Diaspora in Deutschland, die von ihrem Profil her den hier analysierten Personen näher stehen, derartiges nicht aufgefallen ist (Besters-Dilger 2013: 200).

6.6 Abbau des reflexiven Possessivpronomens

Im Unterschied zum Deutschen verfügt das Standard-Tschechische über ein reflexives Possessivpronomen *svůj*, das immer dann verwendet wird, wenn die Person, der durch das Pronomen etwas zugeeignet wird, auch das Subjekt des jeweiligen Satzes ist (Šticha et al. 2013: 394). In anderen Fällen werden bei der Zueignung – entsprechend der Person – die Possessivpronomina *můj* ‚mein‘, *tvůj* ‚dein‘ und *jeho* ‚sein‘ / *její* ‚ihr‘ bzw. *náš* ‚unser‘, *váš* ‚euer‘, *jejich* ‚ihr.PL‘ verwendet. Im Beispiel (30a) scheint sich daher eine Restrukturierung des tschechischen Systems zu manifestieren, denn man referiert darin auf die Eltern des Vaters, der auch in der Subjektposition steht, wobei anstatt *svůj* ‚sein.REFL‘ das Possessivpronomen *jeho* ‚sein‘ verwendet wird:

(30) a. **Bilingualen-Sprache**

# otec	ještě	měl # p* _	bratry _	a	jeho	rodiče
Vater	noch	hab.3SG.PST	Bruder.ACC.PL	und	3SG.POSS	Elter.ACC.PL
v	Čechách					
in	Böhmen.LOC					
	(II_ROD_SP_CZ)					

b. **BaselinKa**

# otec	ještě	měl # p* _	bratry _	a	svoje	rodiče
Vater	noch	hab.3SG.PST	Bruder.ACC.PL	und	REFL	Elter.ACC.PL
v	Čechách					
in	Böhmen.LOC					
	<i>der Vater hatte noch Brüder und seine Eltern in Böhmen</i>					

Das Ersetzen von *svůj* durch nicht-reflexive Possessivpronomina kommt auch an anderen Stellen in unserem Korpus vor: *já jsem nesměla _ zůstat _ stát _ s mým autem* (CI_BUB_SP_CZ), *že jsem to mojim dětem nena* nena* _ nea* _ ne* _ *naučila* (TU_HAG_SP_CZ), *jo trošku ale _ já mám teďka tady moji rodinu* (TU_HAG_SP_CZ), *a _ já jsem potom děti dala k mojí mamince* (TU_HAG_SP_CZ) u. a.

Über die Tendenz, das reflexive Possessivpronomen durch das Possessivpronomen der jeweiligen Person zu ersetzen, berichtet auch Polinsky (2006: 240). Allerdings steht das reflexive Possessivpronomen – besonders in der 1. und 2. Person – auch in der Heimatland-Baseline inzwischen in Konkurrenz mit nicht-reflexiven Possessivpronomina (Čmejková 2011), sodass seine klare Präferenz in unserem Korpus als Akzeleration einer Entwicklungstendenz verstanden werden kann, die es in der Heimatland-Baseline gibt (Cope & Dittmann 2020).

Die Frequenz der Possessivpronomina in der Heimatland-Baseline des Tschechischen scheint im Übrigen insgesamt niedriger zu sein als im Deutschen, in dem die Possessivpronomina die Funktion eines Artikels übernehmen. Indem man die Possessivpronomina im Tschechischen der Bilingualen im oberen Absatz mit Blick auf die Heimatland-Baseline tilgen würde, scheinen hier die Bilingualen dem deutschen Verwendungsmuster zu folgen.

6.7 Abweichungen in der Verbvalenz

Einige der oben angeführten Beispiele, wie (18a), (19a) und (20a), die als Beispiel für die Restrukturierung des Kasussystems genutzt wurden, könnte man allerdings auch als sprachkontaktbedingte Abweichungen von der Valenz konkreter Verben sehen, das heißt von ihrer „Fähigkeit [...], Ergänzungen zu verlangen“ (Welke 2011: 45). Diese verstärken lediglich den Eindruck der Restrukturierung des Kasus, während sie in Wirklichkeit aber anderweitig motiviert sind. So verlangt das Verb *ptát se* ‚fragen‘ in (19), mit der Bedeutung ‚jemandem eine Frage stellen‘, drei Ergänzungen, die mit spezifischen semantischen Rollen verbunden sind und denen im aktiven Satz auch spezifische Kasusformen zugewiesen werden – Agens im Nominativ, Adressat im Genitiv und Patiens in einer Präpositionalphrase oder einem Nebensatz (valex 3.0 2020). Das deutsche Verb *fragen* mit der Bedeutung ‚jemandem eine Frage stellen‘ verlangt dagegen eine Nominativergänzung und eine (fakultative) Akkusativergänzung in Form einer Nominal- (*er fragte Vera*) oder Pronominalphrase (*er fragte mich*) sowie die eines Nebensatzes (*er fragte, wie es dir geht*) (Institut für Deutsche Sprache 2018).¹⁰

Zwischen den Verben *ptát se* und *fragen* gibt es also in der gewählten Lesart keine Unterschiede in der Anzahl der Slots und in den semantischen Restriktionen ihrer Besetzung. Der Kontrast liegt lediglich in der qualitativen Valenz, das heißt in dem zugewiesenen Kasus bei der zweiten Ergänzung. Die Respondentin in (19a) wählt außer der Nominativergänzung (*vnoučata*) eine Akkusativergänzung (*rodiče, dědečka a babičku*), was dem deutschen Satzbauplan entspricht. Das Verb *ptát se* erhält durch diese Kasuszuweisung keine zusätzliche Bedeutung, die auf eine Polysemiekopierung hindeuten würde. Es handelt sich demnach um eine Restrukturierung des Satzbauplans der tschechischen Replikasprache nach dem Muster der deutschen Modellsprache (Kuteva & Heine 2012: 179, siehe Abschnitt 5). Analoge Übernahmen, die von einer Polysemiekopierung oder Restrukturierung der Valenzmerkmale anhand deutscher Muster zeugen, beschreibt auch Hansen (2018: 137ff.) in Bezug auf kroatische und serbische Herkunftssprecherinnen und -sprecher in Deutschland.

Auch im Falle von (18a), (19a) und (20a) kann jedoch nicht von einer Polysemiekopierung ausgegangen werden, denn die Bedeutung der Verben *mluvit* ‚sprechen‘, *ptát se* ‚fragen‘ und *rozumět* ‚verstehen‘ wird durch die Veränderung ihrer qualitativen Valenz semantisch nicht erweitert. Gleichzeitig tritt in diesen Beispielen nicht die Restrukturierung im Sinne eines Categorieschwunds auf, denn die Kategorie Kasus bleibt bestehen. Die Typologie von Heine & Kuteva (2005) erscheint daher bezüglich der Verbalvalenz als nicht völlig ausreichend.

Hansen (2018: 137ff.) verwendet für seine Interpretation von PAT im Bereich der Valenz den konstruktionsgrammatischen Ansatz (Fried & Östman 2004). Die Konstruktionsgrammatik in ihrer gebrauchsbasierten, kognitiv-linguistisch orientierten Form (Goldberg 1995a-b, 2006; Boas 2016) ermöglicht es, spontansprachliche Daten und die in ihnen enthaltenen Sprachkontaktphänomene aus der Sicht eines multilingualen Sprachwissens der Interviewten Personen zu untersuchen. Psycholinguistische Variablen (zum Beispiel

¹⁰ Vgl. *Grammis – Grammatisches Informationssystem*: <https://grammis.ids-mannheim.de/verbs/view/400607/1>, letzter Zugriff am 18.11.2021.

das Migrationsalter) sowie soziolinguistische Variablen (zum Beispiel die Integration in die neue Sprachgemeinschaft) erlauben es, auf das *Entrenchment* (kognitive Nicht-/Verfestigung bzw. Stabilität/Lockerung, vgl. Ziem & Lasch 2013: 103) einzelner Konstruktionen in den untersuchten Idiolekten zu schließen und deren konkrete Ausprägung demnach stärker oder schwächer als Ergebnis der Interaktion mit der anderen Sprache zu analysieren.

In diesem Sinne kann man (31a) als eine Fusion der schematischen Konstruktion der Prädikation [[NP_{NOM}] [,sein⁴] [NP_{NOM}]] mit einer lexikalisch gefüllten Konstruktion der Altersangabe [[NP_{NOM} „EXPERIENCER“] [,sein⁴] [NP_{NOM} „ALTER“]] deuten, die es in (31b) nicht gibt. Dadurch entsteht eine Abweichung von der Heimatland-Baseline in (31b), in der die Konstruktion der Altersangabe [[NP_{DAT} „EXPERIENCER“] [,sein⁴] [NP_{NOM} „ALTER“]] keine Instanziierung der Konstruktion [[NP_{NOM}] [,sein⁴] [NP_{NOM}]] ist:

(31) a. **Bilingualen-Sprache**

a	ted'	je	Hanička	dvacet	osm
und	jetzt	sein.3SG	Hanička.NOM	zwanzig	acht
(CI_BUB_SP_CZ)					

b. **BaselinKa**

a	ted'	je	Haničce	dvacet	osm
und	jetzt	sein.3SG	Hanička.DAT	zwanzig	acht
<i>und jetzt ist Hanička achtundzwanzig</i>					

Ähnlich lassen sich auch folgende Beispiele interpretieren: *lékařka si to nevšimla a já taky ne* (NT_GAU_MI_CZ), # *vona všechno rozumí* (LS_ZEI_SP_CZ), *ale voni všechno rozumí ne?* (LS_ZEI_SP_CZ), *jsem si koukala # Ptzen* (LS_ZEI_SP_CZ), *voní tam si to všechny koukali* (LS_ZEI_SP_CZ), *jo jo _jo _ a když _ u nás jsou nějaký # s autobusem když můžeš ject _ na jeden den* (LS_ZEI_SP_CZ), *tak jsme jeli s autem* (LS_ZEI_SP_CZ), *a my jsme jeli s s tou metro* (LS_ZEI_SP_CZ), *a # možná _ že jednou pojedou s busem nebo co nebo vlakem* (LS_ZEI_SP_CZ), *jo jo když pojeděš # v autem* (LS_ZEI_SP_CZ), *ale my sm* my jsme mohli s tím českým penízem něco koupit ne?* (LS_ZEI_SP_CZ), *a _ chcel jsem se domluvit na nějakou schůzku* (QC_MUN_MI_CZ) oder *jsem _ přemejšlela odjet jo?* (NT_GAU_MI_CZ).

6.8 Nicht-kanonische Subjektkonstruktion

Eine der auf der syntaktischen Ebene beobachteten Differenzen zu der Heimatland-Baseline ist die Verwendung von resumptiven Pronomina. In Anlehnung an die Studien zu Kreol- und Pidginsprachen (Bickerton 1993) wird unter diesem Begriff ein pronominales Element verstanden, das mit dem Subjekt derselben Klausel ko-indexiert ist (vgl. Polinsky 2006: 238). Wie aus (32a) ersichtlich wird, folgt nach dem durch eine Nominalphrase ausgedrückten Subjekt ein Personalpronomen, das die Referenz noch einmal herstellt und dadurch verstärkt. Es ist anzunehmen, dass der Grund für diese Wiederholung die Eindeutigkeit der Koreferenz ist. Im Falle des Russischen als Herkunftssprache wurden resumptive Pronomina in einigen Kontexten, zum Teil als Kompensation für erodierte Konjugationsparadigmen, grammatikalisiert (ebd.: 238ff.). Unsere Daten zeigen keine

obligatorische Verwendung, sodass es sich womöglich erst um den Anfang eines Grammatikalisierungsprozesses handelt:

(32) a. **Bilingualen-Sprache**

moje holka **vona** bydlí # _ jednu hodinu od tady weg
 1SG.POSS Mädchen 3SG wohn.3SG ein.ACC Stunde.ACC von hier weg
 (LS_ZEI_SP_CZ)

b. **BaselinKa**

moje dcera bydlí hodinu cesty odtud
 1SG.POSS Tochter wohn.3 SG Stunde. ACC Weg.GEN von hier
meine Tochter wohnt eine Stunde weg von hier

Dennoch sind solche Beispiele, die auch durch (7), *a můj _ dědeček von byl v Božičanech* (TU_HAG_SP_CZ), *moje děti voni říkaly maminko udělej to* (TU_HAG_SP_CZ) oder *ty druhý ty jsou v Kanada* (TU_HAG_SP_CZ) belegt sind, auffällig. In der Heimatland-Baseline wird nämlich die Koreferenz über die Klausenebene hinweg mittels der Demonstrativa *ten, ta, to* ‚der, die, das‘ und *takový* ‚solcher‘ hergestellt (Kolářová 2011).

Die Verständnissicherung steht auch hinter anderen nicht-kanonischen Subjektkonstruktionen. So erscheint das Personalpronomen auch in der Subjektposition von Klausen, in denen in der Heimatland-Baseline *pro*-drop die nicht markierte Variante wäre, und zwar sowohl im Neben- (33a) als auch im Hauptsatz (34a). In den Beispielen (33a) und (34a) wird das pronominale Subjekt nicht weggelassen, obwohl es sich um keine Hervorhebung, Ausklammerung, Konfrontierung oder Rhematisierung handelt (Štícha et al. 2013: 672). Über den Verzicht auf *pro*-drop bei der ersten Generation in der russischsprachigen Diaspora in den USA, wobei dieser im Russischen nicht so stark ausgeprägt ist wie im Tschechischen, berichten auch Dubinina & Polinsky (2013: 167ff.). Diese Abweichung kann dadurch erklärt werden, dass in der deutsch- bzw. englischsprachigen Umgebung ein Satzmuster repliziert wird, das keine Subjektweglassung zulässt:

(33) a. **Bilingualen-Sprache**

a potom jsem řekla že **já** chci do školy
 und dann AUX sag.3SG.PST dass 1SG woll.1SG in Schule.GEN
 (TU_HAG_SP_CZ)

b. **BaselinKa**

a potom jsem řekla že chci do školy
 und dann AUX sag.3SG.PST dass woll.1SG in Schule.GEN
und dann habe ich gesagt dass ich in die Schule will

(34) a. **Bilingualen-Sprache**

a **já** jsem přišla _ přijela
 und 1SG AUX komm.3SG.PST [zu Fuß] komm.3SG.PST [mit einem Fahrzeug]
 s německým autem
 mit deutsch.INS Auto.INS
 (CI_BUB_SP_CZ)

b. **BaselinKa**

a	přijela	jsem	německým	autem
und komm.3SG.PST	AUX	deutsch.INS	Auto.INS	

und **ich** kam mit einem deutschen Auto

Dubinina & Polinsky (2013: 168) machen jedoch darauf aufmerksam, dass ein Verzicht auf *pro*-drop auch bei Bilingualen vorkommt, die von einer *pro*-drop-Sprache (zum Beispiel Hebräisch) umgeben sind. Sie schlagen daher als Erklärungsansatz das Verhältnis zwischen Syntax und Informationsstruktur vor. Die in der Diaspora lebenden Sprecherinnen und Sprecher sind während der Sprachproduktion nicht in der Lage, eine genügend robuste Verbindung zwischen der Informationsstruktur und der syntaktischen Struktur herzustellen. Die Variante, in der das Subjekt durch das Personalpronomen ausgedrückt wird, wählen sie deshalb, weil sie aus kognitiver Sicht einfacher (ebd.: 170) und für die Referenzherstellung und damit auch die Verständnissicherung eindeutiger ist.

6.9 Abweichungen bei Adpositionen und Nebensatz-Konjunktionen

Sprachkontaktbedingte Abweichungen in der Verwendung von Adpositionen, die wir in unserem Material ebenfalls identifizieren konnten, fallen zum einen in den Bereich der Verbalvalenz (vgl. etwa (35a)), zum anderen findet man sie auch in Präpositionalphrasen, die nicht vom Verb regiert werden, wie in (36a). Der Respondent erweitert an dieser Stelle die Verwendung der Präposition *na* ‚auf‘ auf den Kontext der Adverbialphrase, die zur Angabe der Sprache dient, in der etwas gesagt, gesprochen, geschrieben usw. wird. Es handelt sich um eine Polysemiekopierung, die auch Hansen (2018: 149) bei den Präpositionalphrasen bei serbischen und kroatischen Herkunftssprecherinnen und -sprechern in deutschsprachiger Umgebung feststellt:

(35) a. **Bilingualen-Sprache**

ale	von	by	sám	přes	to	nikdy	nemluvil	jo ?
aber	3SG	COND	selbst	über	DEM.ACC	nie	sprech.3SG.COND.NEG	ja

(IL_REG_SP_CZ)

b. **BaselinKa**

ale	on	by	sám	o	tom	nikdy	nemluvil	jo ?
aber	3SG	COND	selbst	über	DEM.LOC	nie	sprech.3SG.COND.NEG	ja

*aber er selber würde **darüber** nie sprechen ja*

(36) a. **Bilingualen-Sprache**

já	jsem	nechtěl	mluvit	nic	na	česky
1SG	AUX	woll.1SG.PST.NEG	sprech.INF	nichts	auf	Tschechisch.ADV

(II_ROD_SP_CZ)

b. **BaselinKa**

já	jsem	nechtěl	mluvit	vůbec	česky
1SG	AUX	woll.1SG.PST.NEG	sprech.INF	gar nicht	Tschechisch.ADV

*ich wollte nichts **auf** Tschechisch sprechen*

Die im Zusammenhang mit der Verwendung einer Sprache auftretende Präpositionalphrase wie in (35) kommt in unserem Material recht häufig vor: *a _ a ale to bylo # # # víc # # na německy _ ne ? hmm* (LS_ZEI_SP_CZ), *# jak se to jmenuje Teppich na česky ? #* (LS_ZEI_SP_CZ), *jak si jakej tak # hrozně jak teďkon mluvím na česky* (II_ROD_SP_CZ), *se mnou _ povídají jenom na česky (víš) (1)* (II_ROD_SP_CZ) usw. Ebenfalls andere Präpositionen werden abweichend von der Heimatland-Baseline verwendet: *a my se _ uvidíme jenom jednou v měsíce _ ne ?* (LS_ZEI_SP_CZ), *a vo* a von přijde t* a vona má pflagedienst jo ? _ a ty přijdou třikrát # _ na den* (LS_ZEI_SP_CZ), *a já jsem _ my jsme přišli na třináctýho listopadu* (TU_HAG_SP_CZ), *no jo moj můj m* manžel jednou # v v roku _ # tak moc už nejede* (LS_ZEI_SP_CZ) usw.

6.10 Nicht-kanonische Wortstellung

Das Tschechische der Bilingualen weist Erosionserscheinungen auch bei der Wortstellung auf, die in der Heimatland-Baseline relativ frei und in jedem Fall frei von der Verbklammer ist. Dieses deutsche topologische Muster erscheint im Tschechischen der Bilingualen zum einen in Form der Verbletzstellung in den Nebensätzen:

(37) a. Bilingualen-Sprache

kde	náš	v	Neu-Ulmu	ten ten	Heim	byl
wo	1PL.POSS	in	Neu-Ulm	DEM	Heim	sein.3SG.PST

(KL_HEI_SP_CZ)

b. BaselinKa

v	Neu-Ulmu	kde	byl	náš	domov
in	Neu-Ulm	wo	sein.3SG.PST	1PL.POSS	Heim

wo in Neu-Ulm unser Heim war

Zum anderen wird die deutsche Satzklammer im Tschechischen der Bilingualen auch in Hauptsätzen repliziert:

(38) a. Bilingualen-Sprache

tak #	by	musel	jeho	důchod	jináč	počítat
so	COND	müss.3SG.COND	3SG.POSS	Rente	anders	rechn.INF

(KL_HEI_SP_CZ)

b. BaselinKa

tak	by	si	musel	jinak	počítat	důchod
	COND	REFL	müss.3SG.COND	anders	rechn.INF	Rente

dann müsste er seine Rente anders berechnen

Aus Sicht der Informationsstruktur liegt in (38) der Fokus auf der *Rente*, die in eine rhetorische Position rückt. Diese wäre in der Heimatland-Baseline in satzfinaler Position. In (38a) bleibt sie jedoch im Mittelfeld und die Klausel wird mit dem Infinitiv abgeschlossen.

Die Beispiele (37) und (38) sind also jeweils auf ein spezifisches deutsches Wortstellungsmuster zurückzuführen. Eine Korrespondenz mit der deutschen Satzstruktur stellt bei den russischsprachigen MI der ersten Generation ebenfalls Besters-Dilger (2013: 197) fest.

Im Gegensatz dazu lassen sich Abweichungen im Bereich der Klitika nicht eindeutig mit einem deutschen Muster verbinden. Als kanonische Position der unbetonten Klitika im Tschechischen wird die Stelle nach der ersten syntaktischen und phonetisch realisierten Phrase im Satz bezeichnet (Štícha et al. 2013: 556). Im Beispiel (39a) erscheint zwar das unbetonte Auxiliar (*jsem* ‚[ich] bin‘ nach der ersten phonetisch realisierten Einheit, aber die Konjunktion *und* (die gleichzeitig eine diskursstrukturierende materielle Entlehnung darstellt) bzw. ihr tschechisches Äquivalent *a* erlaubt im Standard unmittelbar nach sich kein Klitikon (Štícha 2003: 126; Štícha et al. 2013: 566). Deshalb entspricht die Position von *jsem* in (39a) nicht der Wortstellung in der Heimatland-Baseline (vgl. (39b)). Darin kommen zwar einsilbige, unbetonte Einheiten (wie das hier behandelte *jsem*) nicht nur an der besagten zweiten Stelle, sondern oft auch am Anfang einer Klausel vor (Hoffmannová & Richterová 2019), allerdings schließt dies dann die Verwendung von Konjunktionen aus, die in (39a) vorliegt:

(39) a. **Bilingualen-Sprache**

und **jsem** **tam** pracovala jako jenom takový _ práce
 und AUX dort arbeit.1SG.PST wie nur solch.F.PL Arbeit.F.PL
 (TU_HAG_SP_CZ)

b. **BaselinKa**

a tam **jsem** dělala jako jenom takový _ práce
 und dort AUX mach.1SG.PST wie nur solch.F.PL Arbeit.F.PL
und dort habe ich nur solche Arbeiten gemacht

Abweichungen in der Position der Klitika stellt auch Hansen (2018: 150ff.) bei den Herkunftssprecherinnen und -sprechern des Kroatischen in einer deutschsprachigen Umgebung fest. Wie Hansen (2018: 152) anmerkt, ist eine Identifizierung mit der Modellsprache Deutsch in diesem Fall nicht möglich, denn das Deutsche verfügt über keine Klitika, die auf die sog. zweite Position im Satz eingeschränkt sind. Daher sind weder das von Hansen (ebd.) untersuchte Kroatisch und Serbisch noch das in diesem Text behandelte Tschechisch mit dem Deutschen vergleichbar. Es kann daher von einer Restrukturierung der Wortstellungsregeln ausgegangen werden, infolgedessen verliert das Klitikon seine syntaktischen und prosodischen Eigenschaften.

6.11 Abweichungen in der Kongruenz

Sowohl im Deutschen als auch im Tschechischen wird die Kongruenz innerhalb der nominalen Phrase durch die Kategorien Genus, Numerus und Kasus gekennzeichnet. Bei der Kongruenz zwischen Subjekt und Prädikat kommt im Tschechischen neben der Kongruenz in den Kategorien Person und Numerus bei infiniten Verbformen auch die Kongruenz in der Kategorie Genus hinzu. Die bilinguale Sprecherin verletzt aber in (40a) die Kongruenz, indem sie das Prädikat *přišlo* ‚[es] kam‘ in Neutrum Singular statt – wie das dem Subjekt *nějaký peníze* ‚irgendein Geld‘ angemessen wäre – in Maskulin Plural verwendet. Dies kann dadurch erklärt werden, dass sie auf die grammatischen Eigenschaften des deutschen Äquivalents *Geld* zugreift, das im Sinne von ‚Münzen, Geldnoten‘ als Neutrum nur im Singular verwendet wird.

(40) a. **Bilingualen-Sprache**

všechno # jsem dělala aby _ přišlo nějaký peníze
 alles AUX mach.1SG.PST damit komm.3SG.N.COND irgendein.PL Geld.PL
 (TU_HAG_SP_CZ)

b. **BaselinKa**

všechno jsem dělala aby přišly nějaký peníze
 alles AUX mach.1SG.PST damit komm.3PL.COND irgendein.PL Geld.PL
ich habe alles getan damit irgendwie Geld kommt

Für Kongruenzabweichungen zwischen Subjekt und Prädikat gibt es in beiden Gruppen, häufiger jedoch in der Gruppe der SP, auch weitere Beispiele, wie in *ale voni měla # měli _ grab*, wo sich die Respondentin innerhalb der Äußerung selbst korrigiert (LS_ZEI_SP_CZ), *už to byly padesát let* (IL_BLA_SP_CZ), *to je vodsud # tři kilometry pryč* (II_ROD_SP_CZ), *že tam můžeš _ # může se číst český noviny i i česká televize* (KC_ISM_MI_CZ). Die eher seltenen Beispiele von Abweichungen in der nominalen Kongruenz im Numerus, wie sie in (41) vorliegen, lassen sich allerdings durch eine veränderte Äußerungsperspektive in der gesprochenen Sprache erklären:

(41) a. **Bilingualen-Sprache**

a koukám na tu Karlovy Vary # a na na Prahu
 und schau.1SG auf DEM.F.SG Karlsbad.M.PL und auf Prag.F.SG
 (LS_ZEI_SP_CZ)

b. **BaselinKa**

a koukám na ty Karlovy Vary # a na na Prahu
 und schau.1SG auf DEM.M.PL Karlsbad.M.PL und auf Prag.F.SG
und ich schaue mir Karlsbad und Prag an

So bezieht sich das Pronomen *tu* in (41a) auf das folgende Substantiv *Prahu* und nimmt es kognitiv vorweg. Ähnlich lässt sich auch die Kongruenz zwischen Subjekt und Prädikat in folgenden Beispielen erklären, in denen die Sprecherin durch die Kongruenz die Pluralreferenz indiziert: *protože už nejsou tak moc lidí* (IL_BLA_SP_CZ), *a hodně lidí věděli* (TU_HAG_SP_CZ), *ale _ voni tam už nejsou hodně lidí* (TU_HAG_SP_CZ), *a # _ ty _ nejsou tam hodně lidí* (TU_HAG_SP_CZ), *už to byly padesát let* (IL_BLA_SP_CZ) u. a.

In jedem Fall zeugen solche Beispiele von einer geminderten kognitiven Kontrolle über die Sprachproduktion, die auf die seltenere Verwendung der sekundär gewordenen Erstsprache und damit auch die Erosion der Erstsprache zurückzuführen ist. Bei den Herkunftssprecherinnen und -sprechern sind die Abweichungen im Bereich der nominalen Kongruenz, die Polinsky (2006: 238ff.) für Russisch in englischsprachiger oder Hansen (2018: 152ff.) für Kroatisch und Serbisch in deutschsprachiger Umgebung feststellen, weitaus stärker ausgeprägt und sie kommen bezeichnenderweise auch in den Tests vor, in denen die Belege für solche Abweichungen nicht durch eine sich verändernde Äußerungsperspektive der Sprecherinnen und Sprecher zu erklären sind.

6.12 Negation

Die germanischen Sprachen Englisch und Deutsch, von denen Deutsch sogar eine sog. Negationsklammer (Blühdorn 2012: 67ff.) aufweist, unterscheiden sich von den slawischen Sprachen wie Tschechisch typologisch dadurch, dass die Letzteren eine sog. mehrfache Negation haben. Das bedeutet, dass die unbestimmten Proformen mit *ně-* (wie in *někdo* ‚jemand‘, *něco* ‚etwas‘, *někdy* ‚manchmal‘, *někde* ‚irgendwo‘, *nějak* ‚irgendwie‘ oder *nějaký* ‚irgendein‘) bei einer Satznegation, deren Skopus sich auf den gesamten Satz erstreckt, im Tschechischen mit *ni-* (wie in *nikdo* ‚niemand‘, *nic* ‚nichts‘, *nikdy* ‚niemals‘, *nikde* ‚nirgends‘ oder *nijak* ‚keineswegs‘) oder mit dem negativen Totalisator *žádný* ‚kein‘ ebenfalls negiert werden (Štícha et al. 2013: 749). Davon weichen aber die bilingualen Sprecherinnen und Sprecher ab:

(42) a. **Bilingualen-Sprache**

ale	já	jsem	taky	nikam	jela
aber	1SG	AUX	auch	nirgendwohin	fahr.3SG.PST

(LS_ZEI_SP_CZ)

b. **BaselinKa**

ale	já	jsem	taky	nikam	nejela
aber	1SG	AUX	auch	nirgendwohin	fahr.3SG.PST.NEG

*aber ich bin auch **nirgendwohin** gefahren*

(43) a. **Bilingualen-Sprache**

tak_#	bylo	to	žad* žádným	problém
so	sein.3SG.N.PST	DEM	kein.M.SG.INS	Problem.NOM

(mně)
1sg.DAT
(LS_ZEI_SP_CZ)

b. **BaselinKa**

tak_#	nebyl	to	žádný	problém
so	sein.3SG.M.PST.NEG	DEM	kein.M.SG.NOM	Problem.NOM

(pro mě)
für 1sg.ACC
*also **war** es **kein** Problem (für mich)*

In den Phrasemen, wie (*ne*)*mit co dělat* ‚mit etwas (nichts) zu tun haben‘, wird von dieser Regel abgewichen, während der bilinguale Sprecher sie – hyperkorrekt – mit mehrfacher Negation verwendet:

(44) a. **Bilingualen-Sprache**

a_#	ta	neměla	s	češtinou	nic_ nic	co	dělat
und	DEM	hab.3SG.PST.NEG	mit	Tschechisch.INS	nichts	was	tun.INF

(QC_MUN_MI_CZ)

b. **BaselinKa**

a	ta	neměla	s	češtinou	co	dělat
und	DEM	hab.3SG.PST.NEG	mit	Tschechisch.INS	was	tun.INF

*und die hatte mit Tschechisch **nichts** zu tun*

Diese Abweichung ist in unserem Material, wie in *ale to nemá _ nic co dělat s tím ovládním toho jazyka* (NT_GAU_MI_CZ), bei mehreren Sprecherinnen und Sprechern verfestigt, bei denen die grammatischen Regeln noch bekannt sind, die Norm ihrer Anwendung aber durch die schwächere oder fehlende kognitive Verfestigung deaktiviert oder erloschen ist. So scheint sie für die Erosion des Tschechischen charakteristisch zu sein.

6.13 Abweichungen in der Genuszuweisung

Das Tschechische verfügt wie das Deutsche beim Nomen über ein dreigliedriges Genussystem, das beim tschechischen Maskulinum um die natürlich motivierte Unterscheidung belebt vs. unbelebt reicher ist. Das dem lexikalischen Wortstamm des Substantivs inhärente Genus wird an der Kongruenz der Pronomina und Adjektiva im Rahmen der Nominalgruppe erkennbar, wie bei den Indefinitpronomina *takovou* ‚solch.F.SG.ACC‘ und *nějakou* ‚irgendein.F.SG.ACC‘ in (45b), die mit dem Substantiv *závist* ‚Neid‘ kongruieren. In (45a) liegt die Genuskongruenz in dieser Nominalphrase hingegen nicht vor. Wir ordnen diese und andere Beispiele nicht unter 6.11 ein, weil dies auf den spezifischen Transfer des deutschsprachigen Äquivalents *der Neid* zurückgeführt werden kann, das bereits beim Substantiv, nicht erst bei der Kongruenz für die maskuline Genuszuweisung verantwortlich sein dürfte. Zum anderen wäre in diesem Fall auf eine Analogie zu schließen, denn die Nomina mit Null-Endung sind im Tschechischen typischerweise, wenn auch nicht ausschließlich, maskulin, wobei dies als Indiz einer Restrukturierung des Deklinationssystems bzw. der Lockerung des formalen Bandes zwischen dem Regenten und dem Dependents verstanden werden kann:

(45) a. **Bilingualen-Sprache**

a	takže	nikdo n* #	neměl	takovej
und	also	niemand	hab.3SG.PST.NEG	solch.M.SG.ACC
	nějakej		jako	závist
	irgendein. M.SG.ACC		wie	Neid

(KG_MUN_MI_CZ)

b. **BaselinKa**

a	takže	nikdo n* #	neměl	takovou
und	also	niemand	hab.3SG.PST.NEG	solch.F.SG.ACC
	nějakou		jako	závist
	irgendein. F.SG.ACC		wie	Neid

und also niemand war irgendwie neidisch

Eine falsche Genuszuweisung im Rahmen der Nominalgruppe und damit auch Abweichungen von der in der Heimatland-Baseline üblichen Kongruenz zwischen dem Subjekt und den nominalen oder infiniten Teilen des Prädikats finden sich auch an anderen Stellen

unseres Korpus, wie in *jeden euro co dostanete za jednu euru nic* (QC_WUN_SP_CZ), *tam je taky _ krásná _ penzion tam máme* (KL_HEI_SP_CZ), *jestli to je dobrý pro tu topení pro to pod le* po* _ (pod letní) nebo ne* (KL_HEI_SP_CZ), *z vesnice do druhýho tam jed* je* ty jedou jenom dvě auta za den* (KL_HEI_SP_CZ), *a tak se _ta _jo ? _ wie sagt man das _ das Leben _ život _ se tam taky # změni* _ změnilo* (IL_BLA_SP_CZ). Über Abweichungen bei der Genuszuweisung berichtet im Falle der russischen Herkunftssprecherinnen und -sprecher in den USA auch Polinsky (2006: 221ff.).

6.14 Zusammenfassung der Ergebnisse

Wie wir zeigen konnten, finden sich in den Interviews mit deutsch-tschechischen Bilingualen, die in der ersten Generation in Deutschland leben, Isolations- und Kontakthänomene, die Ergebnis der Erosion (Attrition) sind und sich den Typen von Isolations- und Kontakthänomenen zuordnen lassen, die auch für die Herkunftssprecherinnen und -sprecher slawischer Sprachen der zweiten Generation charakteristisch sind, bei denen sie Ergebnis des unvollständigen Spracherwerbs sind. Lediglich für den Abbau des Konditionals und seinen Ersatz durch den Indikativ, auf den Polinsky (2006: 233) aufmerksam macht, lassen sich in unserem Korpus so gut wie keine Belege finden. Dies macht deutlich, dass diese Typologie nicht nur von einem Typus der Bilingualen auf den anderen übertragbar ist, sondern auch, wie systematisch die jeweiligen Erscheinungen sind. Zugleich wird aber auch klar, wie schwierig es ist, den unvollständigen Spracherwerb und die Attrition auseinanderzuhalten. Dafür braucht man die Evidenz aus den metasprachlichen Daten, wie sie in den sprachbiographischen Narrationen vorliegen.

Geht man der Frage nach, wie in den Isolations- und Kontakthänomenen, die Ergebnis der Erosion (Attrition) sind, die Typen von *pattern replications* vertreten sind, die Hansen (2018: 130) für die Herkunftssprecherinnen und -sprecher formuliert, dann stellt man fest, dass alle drei, d. h. sprachkontaktinduzierte Grammatikalisierung, Polysemiekopierung (grammatischer Funktionen) und Restrukturierung potenziell vorhanden sind.

Als eine potenzielle sprachkontaktinduzierte Grammatikalisierung kann man die Verwendung der Präposition *vod* (standardsprachlich *od*) ‚von‘ bezeichnen. Diese wird von den interviewten Personen in possessiven Konstruktionen anstelle eines einfachen Genitivs gebraucht. Weiter könnte man die Verwendung des Indefinitpronomens (*nějaký* ‚irgendein‘) in der Funktion des unbestimmten Artikels als beginnende Grammatikalisierung sehen. Außerdem kommt in den Daten auch die semantische Entleerung gewisser Verben vor, die in einer Art *light verb construction* aus semantisch entleertem Verb plus Infinitiv oder Substantiv in Objektfunktion statt eines Verbs mit derselben Bedeutung benutzt werden (zum Beispiel *udělat světlo* ‚Licht machen‘ statt *rozsvítit* ‚(die Lampe) anschalten‘). Aus der Sicht der Sprecherinnen und Sprecher kann man diese Konstruktion als eine dekompositionelle Struktur interpretieren. Bei dekompositionellen Strukturen wird ein Konzept in einfachere semantische Einheiten aufgebrochen, wobei jede Einheit mit einem Wort lexikalisiert ist (vgl. Rakhilina et al. 2016: 10f.). Auch die nicht-kanonischen Subjektstrukturen können als eine erste Stufe der Grammatikalisierung von Personalpronomina als Referenzverstärker oder formales Subjekt angesehen werden.

Als Polysemiekopierungen können sowohl der Gebrauch des Präsens in der Funktion des Futurs als auch die Verwendung von Adpositionen interpretiert werden. In diesen Fällen erweitern die Elemente der untersuchten Sprache ihre Bedeutung im Vergleich zu der Heimatland-Baseline.

Bei dem markantesten Beispiel der möglichen Restrukturierung – dem Kasus – bleibt die Frage offen, ob es sich um isolierte Beispiele oder – je nach der/dem Interviewten – tatsächlich um die Restrukturierung des Kasusystems (vgl. Kuteva & Heine 2012: 179) handelt. Um festzustellen, ob die Interviewten zwischen verschiedenen Kasusformen unterscheiden, müssten wir wie Polinsky (2018: 223) Rezeptionsstudien anschließen und die in den Narrationen abweichenden Kasusformen systematisch testen.

Weitere Bereiche, die von einer Restrukturierung betroffen sind, stellen Aspekt und Determiniertheit, reflexive Verben sowie das reflexive Possessivpronomen und die Valenz konkreter Verben dar, wobei die Unterscheidung der Restrukturierung der Valenz konkreter Verben (6.7) und des Kasus (6.1) teilweise sehr schwierig ist. Die Restrukturierung der Valenz verstärkt in jedem Fall den Eindruck der Restrukturierung der Kasus.

Vergleicht man die Isolations- und Kontaktphänomene bei der Gruppe der tschechischen MI und der SP aus der Tschechoslowakei, stellt man fest, dass die Gruppe der SP fast fünfmal so häufig wie die Gruppe der MI im Tschechischen auf Muster aus dem Deutschen zurückgreift. Diese ersten Ergebnisse sind zu der Zeit, in der der Aufsatz entsteht, noch vorläufig, denn die Kontrolle der annotierten Daten wurde noch nicht abgeschlossen. Sie zeigen jedoch eine Tendenz, die die im 1. Abschnitt erwähnte Erwartung bestätigt: Die SP sind aufgrund ihrer Sprachbiographien von der Spracherosion des Tschechischen stärker betroffen als die MI. Eine erhebliche Rolle dürfte ihr Sprachmanagement und darin vor allem die Sprachwahl und die daraus resultierende Frequenz der Verwendung des Tschechischen nach der Emigration spielen. Es lässt sich behaupten, dass die Sprachideologien in Bezug auf das Tschechische und die damit verbundene fehlende Motivation zum Erhalt des Tschechischen Unterschiede zu den MI aufweisen, was sich auch in der Herausbildung unterschiedlich ausgeprägter sozialer Netzwerke in der Emigration spiegelt. Während die SP in lokale deutsche oder sudetendeutsche Netzwerke vor Ort fanden, trugen bei den MI beispielsweise ihre Verbindungen zu tschechischen Bekannten in der Emigration sowie intensivere Kontakte zu Verwandten und zum Heimatland dazu bei, dass das Tschechische bei ihnen weniger syntaktische Musterentlehnungen aus dem Deutschen enthält, als es bei den SP der Fall ist.

7 Zusammenfassung

Dieser Beitrag stellt das deutsch-tschechische Teilprojekt innerhalb des LangGener-Projektes vor. Das Verbundprojekt ist Musterentlehnungen gewidmet, die auf der Mehrsprachigkeit der interviewten Personen beruhen, und zeichnet sich durch die Verbindung von grammatikorientierter und soziolinguistischer Forschung aus. Diese Verbindung wird im deutsch-tschechischen Teilprojekt, das auf Sprachen im Migrationskontext abzielt, noch enger: Es stellt die Frage nach der Auswirkung von Sprachideologien auf den Sprachgebrauch und damit auch auf die Musterentlehnungen im Deutschen und Tschechischen. Untersucht wird sie in dem vorgestellten Teilprojekt anhand zweier Gruppen deutsch-tschechischer Bilingualer. Zum einen geht es um die SP, das heißt Angehörige der (sude-

ten)deutschen Minderheit auf dem Gebiet des heutigen Tschechiens, die nach dem Zweiten Weltkrieg nicht zwangsausgesiedelt wurden und erst in den 1960er Jahren oder später nach Westdeutschland ausgewandert sind. Zum anderen werden die MI untersucht, d. h. Personen ohne einen deutschen Familienhintergrund, die im Zuge des Prager Frühlings und seiner Folgen die Tschechoslowakei verließen und ebenfalls nach Westdeutschland emigrierten. Es zeigte sich, dass sich diese Gruppen in ihren Sprachideologien und in ihrem Sprachmanagement unterscheiden, was Auswirkungen etwa auf die Netzwerkbildung und damit auch die sprachliche und soziale Praxis hatte. Dies hat einen erheblichen Einfluss nicht nur auf den Erhalt des Tschechischen, sondern voraussichtlich auch den Erwerb des Deutschen nach der Auswanderung. In Bezug auf die Erosion des Tschechischen, auf die sich dieser Beitrag konzentriert, zeigte sich bei der Gruppe der SP, dass diese bei ihnen ausgeprägter ist als bei den MI.

Die interviewten Personen beider untersuchter Gruppen sind in den 1950er Jahren in der Tschechoslowakei geboren und wurden in einem von Tschechisch dominierten Sprachregime sozialisiert. Einige SP haben Deutsch als eine (der) Erstsprache(n) von ihren Eltern erworben, während die MI mit Tschechisch als Erstsprache monolingual aufgewachsen sind. Nach der Auswanderung ist bei beiden Gruppen der Interviewten das Deutsche zur primären Sprache geworden. Bei den SP erfolgte ein abrupter Sprachwechsel durch die Aufgabe des Tschechischen in allen Domänen der Sprachverwendung. Die MI behielten hingegen Tschechisch in der Kommunikation mit ihrer Familie und ihren Freunden sowie durch die Kontakte zum Heimatland. Was das Deutsche betrifft, wurde dieses von den SP reaktiviert bzw. als Zweitsprache komplett erworben. Bei den MI kann man im Gegensatz dazu von einem limitierten Spracherwerb und einem größeren Ausmaß an Fossilisierung ausgehen, denn ihre sozialen Netzwerke innerhalb der deutschen Sprachgemeinschaft waren weniger ausgeprägt als bei den SP. Dies wird aber in dieser Studie nicht untersucht.

Die sprachlichen sowie metasprachlichen Daten wurden mittels narrativer Interviews erhoben, in denen es um eine sprachbiographische Erzählung der interviewten Personen ging. Die Interviews wurden sowohl auf Deutsch als auch auf Tschechisch durchgeführt. Im deutschsprachigen Teil wurde die Zeit vor der Auswanderung, die die interviewten Personen in der Tschechoslowakei verbrachten, erzählt. Im tschechischsprachigen Teil ging es um das Leben nach der Auswanderung. Wir haben beabsichtigt, durch die Verkehrung der Sprachen angesichts der Gebrauchskontexte den kognitiven Druck zu erhöhen, der das Auftreten von Sprachkontaktphänomenen begünstigen kann. Nach der freien Erzählung in jeder Sprache folgte jeweils ein Nachfrageteil, wodurch es möglich war, die sprachbiographisch relevanten Informationen zum Spracherwerb und Sprachgebrauch zu erhalten.

Auch wenn die Forschungsfrage auf soziolinguistischen Einflussfaktoren basiert, spielt auch die psycholinguistische Dimension eine wichtige Rolle bei der theoretischen Einordnung der Studie, da sich die soziolinguistischen Einflussfaktoren psycholinguistisch auswirken. So können die vorgestellten Sprachkontakterscheinungen aus dem Blickwinkel der Sprachattritionsforschung durch Lockerung der verfestigten Strukturen infolge eines unzureichenden Inputs erklärt werden. Sie lassen sich in einer universellen Typologie als Polysemiekopierung, Restrukturierung oder kontaktinduzierte Grammatikalisierung deuten.

Diese Typologie liegt auch der Analyse der ersten Ergebnisse zugrunde, die sich in dieser Studie auf die Spracherosion des Tschechischen beschränkt. Sie umfassen

dieselben Typen von Isolations- und Kontaktphänomenen, denen man auch bei der Analyse von herkunftssprachlichen Daten begegnen kann: Restrukturierung des Kasussystems, Herausbildung des Artikels, Restrukturierung des Aspekts und der Determiniertheit bei Bewegungsverben, Restrukturierung des Tempussystems, Variation bei reflexiven Verben, Abbau des reflexiven Possessivpronomens, Abweichungen in der Verbvalenz, nicht-kanonische Subjektkonstruktionen, Abweichungen bei Adpositionen, Wortstellung, Kongruenz, Negation und Abweichungen in der Genuszuweisung. Die weitgehende Übereinstimmung der Muster- bzw. der strukturellen Entlehnungen in den herkunftssprachlichen Daten und in den sprachlichen Daten von bilingualen Sprecherinnen und Sprechern, die die nun erodierende Sprache vollständig erworben haben, bestätigt dann den Ansatz, dass man – an den Muster- bzw. den strukturellen Entlehnungen selbst – kaum eindeutig entscheiden kann, welche auf den unvollständigen Spracherwerb und welche auf die Attrition zurückzuführen sind. Hier müssen die Sprachbiographien in die Interpretation einbezogen werden.

Literatur

- Achterberg, Jörn (2005): *Zur Vitalität slavischer Idiome in Deutschland. Eine empirische Studie zum Sprachverhalten slavophoner Immigranten*. München: Sagner.
- Anstatt, Tanja (2009): Der Erwerb der Familiensprache. Zur Entwicklung des Russischen bei bilingualen Kindern in Deutschland. In: Irina Gogolin, Ursula Neumann (Hgg.), *Streitfall Zweisprachigkeit – The Bilingualism Controversy*. Wiesbaden: VS Verlag für Sozialwissenschaften, 111–131.
- Anstatt, Tanja (2011): Sprachattrition. Abbau der Erstsprache bei Russisch-Deutschen Jugendlichen. *Wiener Slawistischer Almanach* 67, 7–31.
- Bachmann, Armin R./Greule, Albrecht/Muzikant, Mojmir/Scheuringer, Hermann (Hgg.) (2014–2019): *Atlas der deutschen Mundarten in Tschechien*. 7 Bd. Tübingen: Narr, Francke.
- Barth, Fredrik (1998 [1969]): Introduction. In: Fredrik Barth (Hg.), *Ethnic Groups and Boundaries. The Social Organization of Cultural Difference*. Long Grove, Ill.: Waveland Press, 9–38.
- Berger, Tilman (1993): *Das System der tschechischen Demonstrativpronomina*. LMU München: Habilitationsschrift.
- Besters-Dilger, Juliane (2013): Russian in Germany. Intermediate results on L1 attrition. In: Michael Moser, Maria Polinsky (Hgg.), *Slavic Languages in Migration*. Wien: LIT Verlag, 189–204.
- Bickel, Balthasar/Comrie, Bernard/Haspelmath, Martin (2015): Leipzig glossing rules. Conventions for interlinear morpheme-by-morpheme glosses. Online verfügbar unter: <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>, zuletzt aktualisiert am 31.05.2015, letzter Zugriff am 26.06.2019.
- Bickerton, Derek (1993): Subject focus and pronouns. In: Francis Byrne, Donald Winford (Hgg.), *Focus and Grammatical Relations in Creole Languages. Papers from the University of Chicago Conference on Focus and Grammatical Relations in Creole Languages*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 189–212.
- Blaszczyk, Izabela Maria (2015): *Ausdruck von (In)definitheit bei polnisch-deutschen Bilingualen. Eine Analyse am Beispiel des Demonstrativums ten und des Zahlwortes jeden*. Saarbrücken: AV Akademikerverlag.
- Blühdorn, Hardarik (2012): *Negation im Deutschen. Syntax, Informationsstruktur, Semantik*. Tübingen: Narr.
- Boas, Hans (2016): Cognitive Construction Grammar. In: Thomas Hoffmann, Graeme Trousdale (Hgg.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press, 233–253.

- Bolonyai, Agnes (2009): Code-switching, imperfect acquisition, and attrition. In: Almeida Jacqueline Toribio, Barbara E. Bullock (Hgg.), *Cambridge Handbook of Linguistic Code-Switching*. Cambridge: Cambridge University Press, 253–269.
- Brehmer, Bernhard/Mehlhorn, Grit (2015): Russisch als Herkunftssprache in Deutschland. Ein holistischer Ansatz zur Erforschung des Potenzials von Herkunftssprachen. *Zeitschrift für Fremdsprachenforschung* 26.1, 85–123.
- Bučková, Aneta (2022): *Syntaktische Musterentlehnungen bei deutsch-tschechischen Bilingualen*. Regensburg: Dissertation.
- Busch, Brigitta (2017): *Mehrsprachigkeit*. 2. Aufl. Stuttgart, Wien: UTB; Facultas. Online verfügbar unter: <http://www.utb-studi-e-book.de/9783838537740>.
- Čmejrková, Světa (2011): Posesivní reflexivizace: zájmeno *svůj*, jeho užití a významy. In: František Štícha (Hg.), *Kapitoly z české gramatiky*. Praha: Academia, 655–688.
- Cope, Lida/Dittmann, Robert (2020): Language loss: Czech in the diaspora. In: *Encyclopedia of Slavic Languages and Linguistics Online: Brill*. Online verfügbar unter https://referenceworks.brillonline.com/entries/encyclopedia-of-slavic-languages-and-linguistics-online/language-loss-czech-in-the-diaspora-COM_035822.
- Dietz, Barbara (2011): Aussiedler/Spätaussiedler in Germany since 1950. In: Klaus J. Bade, Pieter C. Emmer, Leo Lucassen, Jochen Oltmer (Hgg.), *The Encyclopedia of Migration and Minorities in Europe: From the Seventeenth Century to the Present*. Cambridge: Cambridge University Press, 245–250.
- Dvořák, Tomáš (2012): *Vnitřní odsun 1947–1953: Závěrečná fáze „očisty pohraničí“ v politických a společenských souvislostech poválečného Československa*. Brno: Maticе moravská.
- Dubinina, Irina/Polinsky, Maria (2013): Russian in the U.S. In: Michael Moser (Hg.), *Slavic Languages in Migration*. Wien: Lit Verlag, 161–187.
- Dürscheid, Christa (2007): Quo vadis, Casus? Zur Entwicklung der Kasusmarkierung im Deutschen. In: Hartmut E. H. Lenk, Maik Walter (Hgg.), *Wahlverwandtschaften: Valenzen – Verben – Varietäten*. Hildesheim: Olms, 89–112.
- Ehlers, Klaas-Hinrich (2015): Vertriebenen-Linguistik. Geschichte und Profil der germanistischen Forschung zu den sprachlichen Folgen der Vertreibung nach dem Zweiten Weltkrieg. In: Gerda Hassler (Hg.), *Metasprachliche Reflexion und Diskontinuität. Wendepunkte – Krisenzeiten – Umbrüche*. Münster: Nodus, 208–221.
- Ehlers, Klaas-Hinrich (2018a): Niederdeutsch bei Alteingesessenen und zugewanderten Vertriebenen in Mecklenburg. In: *Bevensen-Jahrbuch*. 71. Bevensen-Tagung 2018: ‚Allens kloor – Bämßen 70 Jahr‘, 197–210.
- Ehlers, Klaas-Hinrich (2018b): *Geschichte der mecklenburgischen Regionalsprache seit dem Zweiten Weltkrieg. Varietätenkontakte zwischen Alteingesessenen und immigrierten Vertriebenen*. Berlin: Peter Lang.
- Errington, Joseph (2001): *Linguistics in a Colonial World. A Story of Language, Meaning, and Power*. Malden (MA): Blackwell.
- Feldmeier, Beate (2014): *Anrede im Sprachkontakt. Reflexion und Gebrauch von Anredestrategien durch tschechische Migranten im deutschsprachigen Umfeld*. München: Otto Sagner.
- Fried, Mirjam/Östman, Jan-Ola (Hgg.) (2004): *Construction Grammar in a Cross-Language Perspective*. Amsterdam: John Benjamins Publishing Company.
- Giles, Howard/Coupland, Justine/Coupland, Nikolas (Hgg.) (1991): *The Contexts of Accommodation*. Cambridge: Cambridge University Press.
- Glinka, Hans-Jürgen (2009): *Das narrative Interview. Eine Einführung für Sozialpädagogen*. 3. Aufl. Weinheim: Juventa-Verlag.
- Goldberg, Adele E. (1995a): *Constructions. A Construction Grammar Approach to Argument Structure*. Berkeley, University of California (Diss. von 1992 u.d.T. Goldberg, Adele E.: *Argument Structure Constructions*).

- Goldberg, Adele E. (1995b): *Argument Structure Constructions*. Chicago: University of Chicago Press. Online verfügbar unter: <http://www.loc.gov/catdir/description/uchi052/94020705.html>.
- Goldberg, Adele E. (2006): *Constructions at Work. The Nature of Generalization in Language*. Oxford, New York [etc.]: Oxford University Press.
- Green, David W. (1998): Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1.2, 67–81.
- Grosjean, François (2001): The bilingual's language modes. – In: Janet Nicol (Hg.), *One Mind, Two Languages: Bilingual Language Processing* (= Explaining Linguistics, 2). Malden (MA): Blackwell Publishers, 1–22.
- Hämmerle, Verena (2009): Soziolinguistischer Kontext des Zweitspracherwerbs bei Migranten. Das Tschechische in Bayern. *brücken - Germanistisches Jahrbuch Tschechien - Slowakei*. N.F. 17, 305–344.
- Hansen, Björn (2018): On the permeability of grammars: Syntactic pattern replications in heritage Croatian and heritage Serbian spoken in Germany. In: Jasmina Grković-Mejdžor, Björn Hansen, Barbara Sonnenhauser (Hgg.), *Diachronic Slavonic Syntax. The Interplay between Internal Development, Language Contact and Metalinguistic Factors*. Berlin: De Gruyter Mouton, 125–160. Online verfügbar unter: <https://www.degruyter.com/view/book/9783110531435/10.1515/9783110531435-006.xml>, letzter Zugriff am 09.10.2020.
- Heine, Bernd/Kuteva, Tania (2005): *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.
- Helfferich, Cornelia (2009): *Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews*. 3., überarb. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hoffmannová, Jana/Richterová, Olga (2019): Redukované (jednoslabičné) nepřízvukné začátky výpovědí a dialogických replik v mluvené češtině. In: Jana Hoffmannová, Jiří Homoláč, Kamila Mrázková (Hgg.), *Syntax mluvené češtiny*. Praha: Academia, 117–141.
- Institut für Deutsche Sprache (2018): *Wörterbuch zur Verbvalenz. Grammatisches Informationssystem grammis*. Online verfügbar unter: <https://grammis.ids-mannheim.de/verbs/view/400607/1>, letzter Zugriff am 21.09.2020.
- Kagan, Olga E./Carreira, Maria M./Hitchens Chik, Claire (Hgg.) (2017): *The Routledge Handbook of Heritage Language Education. From Innovation to Program Building*. London: Routledge.
- Karlík, Petr & Jarmila Panevová (2017). Reflexivní sloveso. In: Petr Karlík, Marek Nekula, Jana Pleskalová (Hgg.), *CzechEncy – Nový encyklopedický slovník češtiny online*. Online v <https://www.czechency.org/slovník/REFLEXIVN%C3%8D%20SLOVESO#reflexivum%20tantum>, letzter Zugriff am 27.10.2021.
- Keijzer, Merel (2021): First language attrition in the twenty-first century. How continued L1 contact in the digital age fuels language attrition theorizing. In: Evangelia Adamou, Yaron Matras (Hgg.), *The Routledge Handbook of Language Contact*. London, New York: Routledge, 221–233.
- Kolářová, Ivana (2011): Základní jevy koreference. In: František Štícha (Hg.), *Kapitoly z české gramatiky*. Praha: Academia, 980–1000.
- Kopřivová, Marie/Lukeš, David/Komrsková, Zuzana/Poukarová, Petra/Waclawicová, Martina/Benešová, Lucie/Křen, Michal (2017): *ORAL. korpus neformální mluvené češtiny*. Verze 1 z 2. 6. 2017. Praha: Ústav Českého národního korpusu FF UK. Online verfügbar unter: <http://www.korpus.cz>, letzter Zugriff am 20.09.2020.
- Krefeld, Thomas (2004): *Einführung in die Migrationslinguistik. Von der Germania italiana in die Romania multipla*. Tübingen: Narr.
- Kreisslová, Sandra (2014): Sprache und Identität der deutschsprachigen Bevölkerung in den böhmischen Ländern nach 1945. In: Klaas-Hinrich Ehlers, Marek Nekula, Martina Niedhammer, Hermann Scheuringer (Hgg.), *Sprache, Gesellschaft und Nation in Ostmitteleuropa. Institutionalisierung und Alltagspraxis*. Göttingen: Vandenhoeck & Ruprecht, 269–285.
- Kreisslová, Sandra/Novotný, Lukáš (2018): Between language revitalization and assimilation: On

- the language situation of the German minority in the Czech Republic. *Journal of Nationalism, Memory & Language Politics* 12.1, 121–139. DOI: 10.2478/jnmlp-2018-0007.
- Kroskrity, Paul V. (2004): Language ideologies. In: Alessandro Duranti (Hg.), *A Companion to Linguistic Anthropology*. Malden (MA), Oxford: Blackwell, 496–517.
- Kroskrity, Paul V. (2010): Language ideologies: Evolving perspectives. In: Jürgen Jaspers, Jef Verschueren, Jan-Ola Östman (Hgg.), *Society and Language Use*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 192–212.
- Kruse, Jan (2015): *Qualitative Interviewforschung. Ein integrativer Ansatz*. 2., überarb. u. erg. Aufl. Weinheim, Basel: Beltz Juventa.
- Küstners, Ivonne (2006): *Narrative Interviews. Grundlagen und Anwendungen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kuteva, Tania/Heine, Bernd (2012): An integrative model of grammaticalization. In: Björn Wiemer, Bernhard Wälchli, Björn Hansen (Hgg.), *Grammatical Replication and Borrowability in Language Contact. Trends in Linguistics*. Berlin: De Gruyter Mouton, 159–190.
- Matras, Yaron (2009): *Language Contact*. Cambridge: Cambridge University Press.
- Meier, Marius Felix (2017): *Sprachideologien und Sprachmanagement beim Spracherhalt im Ausland: Tschechische Schule Regensburg*. Universität Regensburg: Bachelorarbeit.
- Meng, Katharina (2004): Russlanddeutsche Sprachbiographien – Rückblick auf ein Projekt. In: Rita Franceschini, Johanna Miecznikowski (Hgg.), *Leben mit mehreren Sprachen. Sprachbiographien = Vivre avec plusieurs langues: biographies langagieres*. Bern, Oxford: Peter Lang, 97–117.
- Meng, Katharina/Protassova, Ekaterina (2001): *Russlanddeutsche Sprachbiografien: Untersuchungen zur sprachlichen Integration von Aussiedlerfamilien*. Tübingen: Narr.
- Montrul, Silvina (2008): *Incomplete Acquisition in Bilingualism. Re-examining the Age Factor*. Amsterdam: John Benjamins Publishing Company.
- Montrul, Silvina (2016): *The Acquisition of Heritage Languages*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139030502.
- Nekula, Marek (1997): *Text und Kontext am Beispiel der Konjunktionen*. Brno: Masarykova univerzita.
- Nekula, Marek (2003): *Franz Kafkas Sprachen. „...in einem Stockwerk des innern babylonischen Turmes...“*. Tübingen: Niemeyer.
- Nekula, Marek (2016): *Franz Kafka and His Prague Contexts. Studies in Language and Literature*. Praha: Karolinum.
- Nekula, Marek (2021a): Sprachideologien und Sprachmanagement in sprachbiographischen Interviews. Sprecher mit tschechischem Migrationshintergrund in Bayern. *Brücken – Zeitschrift für Sprach-, Literatur- und Kulturwissenschaft* 28.2, 63–82.
- Nekula, Marek (2021b): Čestina v Německu – jazyk v kontaktu a izolaci v „generaci exil“. *Naše řeč* 104.5, 357–373.
- Nekula, Marek (2021c): Varieties of German in the Czech Republic. *Wiener Slawistisches Jahrbuch*, N.F. 9, 69–103.
- Nekula, Marek/Šichová, Kateřina/Valdrová, Jana (Hgg.) (2013): *Bilingualer Sprachvergleich und Typologie: Deutsch – Tschechisch*. Tübingen: Julius Groos.
- Nekvapil, Jiří (2003): Language biographies and the analysis of language situations. On the life of the German community in the Czech Republic. *International Journal of the Sociology of Language* 162, 63–83. DOI: 10.1515/ijsl.2003.038.
- Nekvapil, Jiří (2004a): Sprachbiographien und Analyse der Sprachsituation. Zur Situation der Deutschen in der Tschechischen Republik. In: Rita Franceschini, Johanna Miecznikowski (Hgg.), *Leben mit mehreren Sprachen. Sprachbiographien*. Bern: Peter Lang, 147–172.
- Nekvapil, Jiří (2004b): Language biographies and management summaries. *Language Management in Contact Situations* 3, 9–33.

- Nekvapil, Jiří/Sherman, Tamah (Hgg.) (2009): *Language Management in Contact Situations. Perspectives from Three Continents*. Frankfurt/Main: Peter Lang.
- Nešpor, Zdeněk R. (2005): České migrace 19. a 20. století a jejich dosavadní studium. *Soudobé dějiny* 12.2, 245–284. Online verfügbar unter <https://www.ceeol.com/search/journal-detail?id=68>, letzter Zugriff am 26.02.2020.
- Neustupný, Jiří Václav (2002): Sociolinguistika a jazykový management. *Sociologický časopis / Czech Sociological Review* 38.4, 429–442. Online verfügbar unter: <http://sreview.soc.cas.cz/cs/issue/55-sociologicky-casopis-czech-sociological-review-4-2002/1002>, letzter Zugriff am 08.01.2020.
- Nübler, Norbert (2017): Determinovanost. In: Petr Karlík, Marek Nekula, Jana Pleskalová (Hgg.), *CzechEncy – Nový encyklopedický slovník češtiny online*. Online v <https://www.czechency.org/slovník/DETERMINOVANOST>, letzter Zugriff am 20.09.2020.
- Olshtain, Elite/Barzilay, Margaret (1991): Lexical retrieval difficulties in adult language attrition. In: Herbert W. Seliger, Robert Michael Vago (Hgg.), *First Language Attrition*. Cambridge: Cambridge University Press, 139–150.
- Paradis, Michel (2004): *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.
- Panochová, Anna (2017): *Zděděný jazyk a základní slovní zásoba češtiny*. FF UK Praha: Diplomarbeit.
- Plesníková, Hana (2018): *Syntaktische Musterentlehnungen (pattern replications) in der Herkunftssprache tschechischsprachiger Migranten in Deutschland*. Regensburg: Masterarbeit.
- Polenz, Peter von (1999): *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart*. Bd. 3: 19. und 20. Jahrhundert. Berlin: De Gruyter.
- Polinsky, Maria (1995): American Russian. Language loss meets language acquisition. *Formal Approaches to Slavic Linguistics*. Cornell Meeting. Michigan Slavic Publications, 370–406.
- Polinsky, Maria (2006): Incomplete acquisition. American Russian. *Journal of Slavic Linguistics* 14.2, 191–262.
- Polinsky, Maria (2008): Without aspect. In: Greville G. Corbett, Michael Noonan (Hgg.), *Case and Grammatical Relations. Studies in Honor of Bernard Comrie*. Amsterdam: John Benjamins Publishing Company, 263–282.
- Polinsky, Maria (2018): *Heritage Languages and their Speakers*. Cambridge: Cambridge University Press.
- Portál | Český národní korpus (2020). Online verfügbar unter: <https://www.korpus.cz/>, letzter Zugriff am 10.08.2020.
- Rakhilina, Ekaterina/Vyrenkova, Anastasia/Polinsky, Maria (2016): Linguistic creativity in heritage speakers. *Glossa: A Journal of General Linguistics* 1.1, 1–29. Online verfügbar unter: <https://doaj.org/article/64ce196235c84656971a9f5bf37fad64>, letzter Zugriff am 27.11.2019.
- Riehl, Claudia Maria (2014): *Sprachkontaktforschung. Eine Einführung*. 3., überarbeitete Aufl. Tübingen: Narr Verlag.
- Romaine, Suzanne (1995): *Bilingualism*. Oxford: Blackwell.
- Rosenthal, Gabriele (1995): *Erlebte und erzählte Lebensgeschichte. Gestalt und Struktur biographischer Selbstbeschreibungen*. Frankfurt/Main: Campus.
- Sakel, Jeanette (2007): Types of loan. Matter and pattern. In: Yaron Matras, Jeanette Sakel (Hgg.), *Grammatical Borrowing in Cross-Linguistic Perspective*. New York: Mouton de Gruyter, 15–31.
- Scholze, Lenka (2006): *Das grammatische System der obersorbischen Umgangssprache unter besonderer Berücksichtigung des Sprachkontakts*. Konstanz: Dissertation.
- Schütze, Fritz (1976): Zur Hervorlockung und Analyse von Erzählungen thematisch relevanter Geschichten im Rahmen soziologischer Feldforschung. Dargestellt an einem Projekt zur Erforschung von kommunalen Machtstrukturen. In: *Kommunikative Sozialforschung. Alltagswissen und Alltagshandeln, Gemeindeforschung, Polizei, politische Erwachsenenbildung*. München: Fink, 159–260.

- Schmid, Monika S./de Leeuw, Esther (2019): Introduction to linguistic factors in language attrition. In: Monika S. Schmid, Barbara Köpke, Mirela Cherciov (Hgg.), *The Oxford Handbook of Language Attrition*. Oxford: Oxford University Press, 181–190.
- Schwarz, Ernst (1934): Jazyk německý na území ČSR. In: *Československá vlastivěda*. Bd. 3: Jazyk. Praha: Sfinx, 524–597.
- Seidlmayer, Katarina (2010): *Spracherwerb und Sprachverlust bei tschechischen Migranten in Bayern. Analyse von zwei Sprachbiographien*. Regensburg: Bachelorarbeit.
- Seidlmayer, Katarina (2013): *Verbvalenz im deutsch-tschechischen Sprachkontakt bei Bilingualen*. Universität: Masterarbeit.
- Selinker, Larry (1972): Interlanguage. *International Review of Applied Linguistics in Language Teaching* 10.1, 209–231. DOI: 10.1515/iral.1972.10.1-4.209.
- Silverstein, Michael (1979): Language structure and linguistic ideology. In: Paul R. Cline et al. (Hgg.), *The Elements. Parasession on Linguistic Units and Levels*. Chicago: Chicago Linguistic Society, 193–247.
- Štícha, František (2003): *Česko-německá srovnávací gramatika*. Praha: Argo.
- Štícha, František/Vondráček, Miroslav/Kolářová, Ivana/Hoffmannová, Jana/Bílková, Jana/Svobodová, Ivana (2013): *Akademická gramatika spisovné češtiny*. Praha: Academia.
- Thomason, Sarah G. (2001): *Language Contact. An Introduction*. Edinburgh: Edinburgh University Press.
- Tišerová, Pavla (2008): Tschechien. In: Ludwig M. Eichinger, Albrecht Plewnia, Claudia M. Riehl (Hgg.), *Handbuch der deutschen Sprachminderheiten in Mittel- und Osteuropa*. Tübingen: Narr, 171–242.
- vallex 3.0 (2020). Online verfügbar unter: <https://ufal.mff.cuni.cz/vallex/3.0/#/lexeme/ptat-1/0>, letzter Zugriff am 21.09.2020.
- Vater, Heinz (2009): Kasus-Schwankungen im gegenwärtigen Deutsch. In: Rita Brdar-Szabó, Elisabeth Knipf-Komlósi, Attila Péteri (Hgg.), *An der Grenze zwischen Grammatik und Pragmatik*. Frankfurt/Main: Peter Lang, 53–66.
- Welke, Klaus (2011): *Valenzgrammatik des Deutschen. Eine Einführung*. Berlin: De Gruyter. Online verfügbar unter: <http://www.reference-global.com/doi/book/10.1515/9783110254198>.
- Wiemer, Björn/Wälchli, Bernhard (2012): Contact-induced grammatical change. Diverse phenomena, diverse perspectives. In: Björn Wiemer, Bernhard Wälchli, Björn Hansen (Hgg.), *Grammatical Replication and Borrowability in Language Contact*. Berlin: De Gruyter Mouton, 3–64.
- Woolard, Kathryn A. (1998): Introduction. Language ideology as a field of inquiry. In: Bambi B. Schieffelin, Kathryn Woolard et al. (Hgg.), *Language Ideologies: Practice and Theory*. Oxford, New York: Oxford University Press, 3–47.
- Ziem, Alexander/Lasch, Alexander (2013): *Konstruktionsgrammatik. Konzepte und Grundlagen gebrauchsbasierter Ansätze*. Berlin, Boston: De Gruyter.

Anhänge

Anhang 1

Interviewleitfaden für die sprachbiographischen Gespräche

Allgemeines

1. Könnten Sie sich vorstellen? Wie heißen Sie?
2. Wann und wo wurden Sie geboren?
3. Woher kommen Sie? Haben Sie hier immer gewohnt? Wo haben Sie zuletzt gewohnt?
4. Was sind Sie von Beruf? Womit haben Sie sich beruflich beschäftigt?

Familiengeschichte

1. Woher stammt Ihre Familie (Eltern)?
2. Womit haben sich Ihre Eltern beruflich beschäftigt?
3. Wie hat man bei Ihrer Mutter/Ihrem Vater zu Hause gesprochen?
4. Sprachen Ihre Eltern noch andere Sprachen? Mit wem?
5. Woher stammen Ihre Großeltern?
6. Wie sprachen Ihre Großeltern? Mit Ihnen/mit den Nachbarn?
7. Sprachen sie noch andere Sprachen?
8. Welcher Religion gehören Sie an?
9. Haben Ihre Familienangehörigen ihre Religion irgendwann gewechselt?
10. Gab es Religionsunterricht zu Ihren Zeiten, zu Zeiten Ihrer Eltern/Großeltern? Wer hat damals unterrichtet? In welcher Sprache war der Religionsunterricht?
11. In welcher Sprache hat man in Ihrer Familie in Ihrer Kindheit gebetet und wie macht man es heute? Erinnern Sie sich an ein Gebet aus der Kindheit? Sprechen dieses Gebet auch Ihre Enkel?

Kleinkindalter bzw. Vorschulzeit

1. Welche Sprache haben Sie in Ihrer Familie zuerst gelernt?
2. Von wem haben Sie diese Sprache gelernt?
3. Wie sprachen Sie mit Ihrer Mutter/Ihrem Vater?
4. Besuchten Sie einen Kindergarten? In welcher Sprache haben Sie mit den Kindergärtnerinnen gesprochen?
5. In welcher Sprache haben Sie damals mit den anderen Kindern gesprochen?
6. In welcher Sprache unterhielten sich Ihre Eltern untereinander? Mit ihren Kindern? Mit Verwandten aus dem gleichen Ort/aus anderen Orten? Mit Freunden/Bekannten aus dem gleichen Ort/aus anderen Orten? Mit Fremden/Nachbarn? Mit dem Arzt/Pfarrer/mit Beamten?
7. In welcher Sprache unterhielten sich Ihre Großeltern untereinander? Mit ihren Kindern? Mit Verwandten aus dem gleichen Ort/aus anderen Orten? Mit Freunden/Bekannten aus dem gleichen Ort/aus anderen Orten? Mit Fremden/Nachbarn? Mit dem Arzt/Pfarrer/mit Beamten/im Geschäft?

Schulzeit

1. Welche Sprachen beherrschen Sie sonst noch aus Ihrer Kindheit/Jugendzeit? Wie haben Sie diese Sprachen gelernt?
2. Wo und wann besuchten Sie die Grundschule?
3. In welcher Sprache haben Sie außerhalb des Unterrichts mit Ihren Lehrern gesprochen?
4. In welcher Sprache haben Sie sich mit Ihren Mitschülern unterhalten? Haben Sie Ihre Mitschüler gerne gehabt? Haben sich die Schulfreunde gegenseitig Spitznamen gegeben? Welche?
5. Gab es an dieser Schule Deutschunterricht/Polnischunterricht? Wie viele Stunden in der Woche? Bitte erzählen Sie davon.

Ggf. zusätzliche Fragen nach:

- Anzahl der Jahre, in denen Deutschunterricht/Polnischunterricht besucht wurde
 - Einstellung zum Deutschunterricht/Polnischunterricht. Haben Sie diesen Unterricht gemocht? Wie war dieser Unterricht?
6. Wie viele Schulklassen haben Sie in einer deutschen, wie viele in einer polnischen Schule besucht?

Jugendalter

1. Welche Schule haben Sie nach der Grundschule besucht? Wie haben Sie dort mit den Lehrern und Mitschülern gesprochen?
2. Gab es an dieser Schule damals Deutschkurse/Polnischkurse?
3. Haben Sie diesen Unterricht besucht?
4. Was haben Sie dann gemacht? Sind Sie zur Arbeit gegangen oder haben Sie eine weitere Ausbildung gemacht?

Falls die Person studiert hat, Fragen 5–6:

5. Wie haben Sie mit den Dozenten/Mitstudierenden damals gesprochen?
6. Gab es an dieser Hochschule Deutschkurse/Polnischkurse? Haben Sie daran teilgenommen?
7. Haben die Bewohner Ihrer Ortschaft gewechselt? Sind neue Einwohner nach dem Krieg gekommen? Woher kamen die neuen Nachbarn? Wie sprachen sie? Wie haben Sie sich mit ihnen verständigt? Wie verhielten sich die neuen Einwohner Ihnen gegenüber? Wie sprach man auf Dorffeiern/Tanzabenden? Gab es Konflikte oder eher ein friedliches Zusammenleben?

Erwachsenenalter

1. Wie ist Ihr Familienstand? Sind Sie verheiratet?
2. Wann und wo wurde Ihr Ehegatte/Ihre Ehegattin geboren? Wie haben Sie Ihren Ehepartner kennengelernt?
3. Woher stammt die Familie Ihres Ehepartners?
4. Haben Sie Kinder? Wann sind sie zur Welt gekommen?
5. Wie sprach man in der Familie Ihres Ehegatten? Sprach er/sie noch andere Sprachen?
6. Wie sprechen Sie mit Ihren Kindern/Enkeln? Antworten sie Ihnen in der gleichen Sprache?
7. Wie spricht Ihr/e Ehegatte/in mit seinen/ihren Kindern/Enkeln? Antworten sie ihm/ihr in der gleichen Sprache?
8. Wie unterhalten Sie sich mit Ihrem Ehepartner?

9. Wie unterhalten sich Ihre Kinder untereinander?
10. In welcher Sprache sprechen Sie jetzt mit dem Ehepartner, den Kindern, Enkeln, Eltern, Geschwistern, Verwandten, Bekannten, Kollegen und Kolleginnen?
11. Welche Sprachen sollten Ihrer Meinung nach Ihre Kinder/Enkel sprechen?
12. Glauben Sie, dass es für Ihre Kinder/Enkel Schwierigkeiten geben könnte, diese Sprachen zu lernen?
13. Können Ihre Kinder und Enkel noch Deutsch/Polnisch?
14. Haben Sie sich überlegt, wie Sie bei Ihren Kindern/Enkeln den Erwerb dieser Sprachen unterstützen könnten?
15. Kommt es vor, dass man Deutsch/Polnisch sprechen muss? Wann?
16. Welche Familienangehörigen haben Sie in Deutschland/Polen? Welche Sprachen sprechen sie?
17. In welcher Sprache lesen Sie Zeitungen oder Zeitschriften/Bücher? Wie oft tun Sie das?
18. In welcher Sprache empfangen Sie Fernseh- oder Radiosendungen? Wie oft tun Sie das?
19. Gibt es in Ihrer Ortschaft eine Kirche? Gibt es da auch deutschsprachige Gottesdienste? Besuchen Sie eher Gottesdienste in der deutschen oder polnischen Sprache?
20. Wie unterhalten Sie sich mit dem besten Freund/der besten Freundin über Politik/gemeinsame Bekannte/im Streit/beim Erzählen von Witzen?
21. In welcher Sprache schreiben Sie meistens Privatbriefe?
22. Wann benutzen Sie Deutsch/Polnisch – heute und früher?
23. Würden Sie gerne öfter Deutsch sprechen?
24. Kamen/Kommen auch Deutsche in Ihre Ortschaft?
25. Haben Sie Ihnen vielleicht als Übersetzerin oder Übersetzer geholfen?
26. Haben Sie Ihren polnischen Nachbarn als Übersetzer/in geholfen?
27. War Ihnen die deutsche Sprache in Ihrem Leben nützlich?

Ergänzende Fragen

1. Fällt es Ihnen einfacher, eine der Sprachen, die Sie verwenden, zu sprechen? Hat sich das im Laufe Ihres Lebens verändert? Wann? Unter welchen Umständen?
2. Welche Sprachen konnten Sie in Ihrer Heimat sprechen?
3. Wie gut haben Sie vor 1945 Polnisch verstanden, gesprochen, gelesen, geschrieben? Wie schätzen Sie Ihre Deutschkenntnisse und Ihre Polnischkenntnisse gegenwärtig ein? Wenn Polnisch die Zweitsprache bildet – Fragen nach: dem Verstehen, Sprechen, Schreiben, Lesen auf Polnisch.
4. Gab es Situationen, in denen Sie eine andere (als die von Ihnen bevorzugte) Sprache verwendeten?
5. Haben Sie einen Sprachkurs besucht?
6. Waren Sie längere Zeit im Ausland? Wie lange und wo?
7. Gab es sonst noch Orte, an denen Sie länger gewohnt haben?
- 7a. Welche Leute haben dort gewohnt: Polen? Deutsche? Katholiken? Protestanten?
- 7b. Wie sprach man dort auf der Straße? In den Familien? In den Ämtern? In welcher Sprache unterhielten sie sich damals mit ihren Kindern? Mit Verwandten aus dem gleichen Ort/aus anderen Orten? Mit Freunden/Bekanntnen aus dem gleichen Ort/aus anderen Orten? Mit Fremden/Nachbarn? Mit dem Arzt/Pfarrer/mit Beamten?

Fragen zum freien Erzählen

1. Wohnen Sie gern in Ihrer Ortschaft/Region?
2. Wo würden Sie am liebsten wegziehen?
3. Können Sie sich noch an Ihre Kindheit erinnern? Welche Aufgaben hatten Kinder damals? Mussten Sie schon zu Hause mithelfen? Was hat sich im Vergleich dazu heute verändert?
4. Wie ist Ihr Tagesablauf, was machen Sie am Tag (heute und früher)?
5. Können Sie sich noch an irgendwelche Kinderstreiche, lustige Geschichten erinnern, die Sie gemacht haben?
6. Kochen Sie gerne? Können Sie mir sagen, wie Sie etwas zubereiten?
7. Wenn jemand krank ist, dann heilt man ihn heute mit unterschiedlichen Mitteln. Welche Mittel/Hausmittel verwendete man früher?
8. Können Sie sich erinnern, wie man früher eine Hochzeit gefeiert hat?
9. Können Sie sich erinnern, ob Sie nach dem 2. Weltkrieg in Polen Deutsch sprechen konnten, hat jemand das Ihnen verboten?
10. Können Sie sich noch an die Zeit vor 1945 erinnern? Möchten Sie davon erzählen?
11. Wie haben Sie das Kriegsende erlebt?
12. Wie haben Sie die Verhängung des Ausnahmezustands in Polen 1981 erlebt?
13. Engagieren Sie sich in irgendwelchen Vereinen, im Deutschen Freundschaftskreis?
14. Ist das (religiöse) Leben heute anders als früher?
15. Gibt es heute andere Sitten und Bräuche als früher?
16. Wo haben Sie früher Ihre Einkäufe gemacht?
17. Können Sie etwas von Ihrem Zuhause, von Ihrem Heimatort erzählen?
18. Was war früher besser/schlechter als heute?
19. Möchten Sie Ihren Nachfolgern eine Botschaft hinterlassen und sie auf etwas besonders aufmerksam machen?

Anhang 2

Einwilligungserklärung auf Deutsch

Einwilligungserklärung

zur Erhebung und Verarbeitung personenbezogener Daten
für Forschungszwecke

ABSCHNITT 1: Personen, Projekt und Ziel

Verantwortlicher gemäß Art. 13 I a) der europäischen Datenschutz-Grundverordnung (im Folgenden: „DS-GVO“) ist die Forschergruppe des DFG-Projektes LangGener, vertreten durch ihre zwei Projektleiter:

Prof. Dr. Björn Hansen

Universität Regensburg
Universitätsstraße 31
93053 Regensburg
Deutschland

Telefon 0941 943 3361
Telefax 0941 943 1991

Prof. Dr. Anna Zielińska

Institut für Slavische Studien
ul. Bartoszewicza 1b/17
00-337 Warszawa
Polen

Tel./Fax. +48 22 826 76 88

Ziel des Forschungsprojektes und damit auch der Zweck der Datenverarbeitung ist es, auf Sprachkontakt zurückführbare Wandelungen im Wort- und Satzbau in Deutsch-Polnisch bilingualer Rede zu erforschen.

Das Interview wird durchgeführt von:

(Bitte in Druckbuchstaben ausfüllen)

Der Betroffene im Sinne von Art. 4 Nr. 1 DS-GVO ist:

(Bitte in Druckbuchstaben ausfüllen)

ABSCHNITT 2: Verarbeitung

Die Verarbeitung von personenbezogenen Daten erfolgt auf Grundlage der europäischen DS-GVO und ihrer nationalen Ausgestaltungen. Rechtsgrundlage für die Verarbeitung ist ihre Einwilligung gem. Art. 6 I a) und Art. 9 II a) DS-GVO.

Von der Einwilligung umfasst sind persönliche Angaben wie zum Beispiel Name, Anschrift, Telefonnummer. Ferner ihre Sprachgeschichte, also unter anderem ihr Alter, Familienstand, Familiensprache. Die zu sammelnden Daten können Informationen beinhalten, die eine besondere Kategorie nach Art. 9 I DS-GVO darstellen, weil aus ihnen „die ethnische Herkunft des Betroffenen“ hervorgeht.

Die beabsichtigten Verarbeitungsschritte sind

- (1) Dokumentation und Analyse ihrer Sprachgeschichte
- (2) Anfertigen einer Tonaufnahme eines Interviews
- (3) Verschriftlichung der Tonaufnahme
- (4) Wissenschaftliche Auswertung von (1) und (3)
- (5) Veröffentlichung der Daten (1) und (3) in einer Online-Datenbank (Korpus)

ABSCHNITT 3: Einwilligungserklärung und Informationen

Hiermit willige ich ein, dass die im Rahmen des in Abschnitt 1 genannten Forschungsprojektes LangGener erhobenen personenbezogenen Daten meiner Person von Beteiligten des DFG-Projektes LangGener verarbeitet werden dürfen, um auf Sprachkontakt zurückführbare Wandelungen im Wort- und Satzbau in Deutsch-Polnisch bilingualer Rede zu erforschen.

Sofern ich besondere Kategorien von personenbezogenen Daten angebe bzw. angegeben habe, sind diese von der Einwilligungserklärung umfasst.

Ihre Einwilligung ist freiwillig. Sie können die Einwilligung ablehnen, ohne dass Ihnen dadurch irgendwelche Nachteile entstehen.

Ihre Einwilligung können Sie jederzeit gegenüber dem in Abschnitt 1 benannten Verantwortlichen widerrufen, mit der Folge, dass die Verarbeitung Ihrer personenbezogenen Daten, nach Maßgabe Ihrer Widerrufserklärung, durch diesen für die Zukunft unzulässig wird. Dies berührt die Rechtmäßigkeit der aufgrund der Einwilligung bis zum Widerruf erfolgten Verarbeitung jedoch nicht.

Relevante Definitionen der verwendeten datenschutzrechtlichen Begriffe sind in der Anlage Begriffsbestimmungen enthalten.

Zur Erreichung des Forschungszieles kann zum gegenwärtigen Zeitpunkt **keine verbindliche Aussage** darüber getroffen werden, wann ihre personenbezogenen Daten gelöscht werden.

Nach Aufnahme sollen die Daten alsbald **pseudonymisiert** werden. Das bedeutet, dass aus den Daten die direkten Bezüge auf ihre Person entfernt werden, sodass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können (Art. 4 Nr. 5 DS-GVO).

Im Rahmen der gesetzlichen Vorgaben hat der Betroffene gegenüber dem Verantwortlichen grundsätzlich Anspruch auf:

- Bestätigung, ob personenbezogene Daten des Betroffenen durch den Verantwortlichen verarbeitet werden,
- Auskunft über diese Daten und die Umstände der Verarbeitung,
- Berichtigung, soweit diese Daten unrichtig sind,
- Löschung, soweit für die Verarbeitung keine Rechtfertigung und keine Pflicht zur Aufbewahrung (mehr) besteht,
- Einschränkung der Verarbeitung in besonderen gesetzlich bestimmten Fällen und Übermittlung der personenbezogenen Daten des Betroffenen – soweit er diese bereitgestellt hat – an den Betroffenen oder einen Dritten in einem strukturierten, gängigen und maschinenlesbaren Format.

Darüber hinaus hat der Betroffene das Recht, seine Einwilligung jederzeit gegenüber dem Verantwortlichen zu widerrufen, mit der Folge, dass die Verarbeitung seiner personenbezogenen Daten, nach Maßgabe seiner Widerrufserklärung, durch diesen für die Zukunft unzulässig wird. Dies berührt die Rechtmäßigkeit der aufgrund der Einwilligung bis zum Widerruf erfolgten Verarbeitung jedoch nicht.

Schließlich sei auf das Beschwerderecht gemäß Art. 77 DS-GVO hingewiesen.

Name des Betroffenen:

(Bitte in Druckbuchstaben ausfüllen)

(Ort, Datum)

(Unterschrift)

Anlage: Begriffsbestimmung

- „Personenbezogene Daten“ sind gemäß Art. 4 Nr. 1 DSGVO alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen. Als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind. Das kann zum Beispiel die Angabe sein, wo eine Person versichert

ist, wohnt oder wie viel Geld er oder sie verdient. Auf die Nennung des Namens kommt es dabei nicht an. Es genügt, dass man herausfinden kann, um welche Person es sich handelt.

- „Besondere Kategorien“ personenbezogener Daten sind gemäß Art. 9 Abs. 1 DSGVO Daten, aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person.
- „Verarbeitung“ ist gemäß Art. 4 Nr. 2 DSGVO jeder mit oder ohne Hilfe automatisierter Verfahren ausgeführte Vorgang oder jede solche Vorgangsreihe im Zusammenhang mit personenbezogenen Daten wie das Erheben, das Erfassen, die Organisation, das Ordnen, die Speicherung, die Anpassung oder Veränderung, das Auslesen, das Abfragen, die Verwendung, die Offenlegung durch Übermittlung, Verbreitung oder eine andere Form der Bereitstellung, den Abgleich oder die Verknüpfung, die Einschränkung, das Löschen oder die Vernichtung.

Anhang 3

Einwilligungserklärung auf Polnisch

FORMULARZ ZGODY

NA PRZETWARZANIE DANYCH OSOBOWYCH I ROZPOWSZECHNIANIE
WIZERUNKU

Dane osoby udzielającej zgody	
Imię i nazwisko

Zgodnie z art. 13 ust. 1 i ust. 2 ogólnego rozporządzenia o ochronie danych osobowych z dnia 27 kwietnia 2016 r. (dalej RODO) informujemy, iż:

1. Administratorem Pani/Pana danych osobowych jest **Instytut Sławistyki Polskiej Akademii Nauk** siedziba przy **ul. Bartoszewicza 1B/17, 00-337 Warszawa, Polska (tel. +48 22 826 76 88)** zarejestrowany w Rejestrze Instytutów Naukowych prowadzonym przez Polską Akademię Nauk pod numerem RIN-I-30/98, NIP: 525-00-12-159, REGON: 000326210.
2. Administrator wyznaczył Inspektora Danych Osobowych Pana Krzysztofa Radtke, z którym można się skontaktować drogą elektroniczną: iod@ispan.waw.pl
3. Pani/Pana dane osobowe przetwarzane będą w celach naukowych na podstawie art. 6 ust. 1 lit. a RODO.
4. Odbiorcą Pani/Pana danych osobowych będą naukowcy realizujący projekt Narodowego Centrum Nauki LangGener: *Pokoleniowe zróżnicowanie języka: zmiany morfosyntaktyczne wywołane przez polsko-niemiecki kontakt językowy w mowie osób dwujęzycznych* na podstawie upoważnienia.
5. Pani/Pana dane osobowe nie będą przekazywane do państwa trzeciego/organizacji międzynarodowej.
6. Pani/Pana dane osobowe będą przechowywane do czasu cofnięcia przez Panią/Pana zgody.
7. Posiada Pani/Pan prawo dostępu do treści swoich danych oraz prawo ich sprostowania, usunięcia, ograniczenia przetwarzania, prawo wniesienia sprzeciwu, prawo do przenoszenia danych.

8. Ma Pani/Pan prawo wniesienia skargi do organu nadzorczego – Urzędu Ochrony Danych Osobowych, gdy uzna Pani/Pan, iż przetwarzanie danych osobowych Pani/Pana dotyczących narusza przepisy ogólnego rozporządzenia o ochronie danych osobowych z dnia 27 kwietnia 2016 r.
9. Podanie przez Panią/Pana danych osobowych ma charakter dobrowolny, jest jednak niezbędne do realizacji projektu LangGener.
10. Pani/Pana dane nie będą przetwarzane w sposób zautomatyzowany w tym również w formie profilowania tzn. żadne decyzje wywołujące wobec osoby skutki prawne lub w podobny sposób na nią istotnie wpływające nie będą oparte wyłącznie na automatycznym przetwarzaniu danych osobowych i nie wiązać się z taką automatycznie podejmowaną decyzją.
11. Pani/Pana dane będą pseudonimizowane.

OŚWIADCZENIE O WYRAŻENIU ZGODY NA PRZETWARZANIE DANYCH OSOBOWYCH

Oświadczam, że wyrażam zgodę na przetwarzanie przez Instytut Slawistyki Polskiej Akademii Nauk z siedzibą w Warszawie (dalej „Administrator”) moich danych osobowych w postaci mojego imienia, nazwiska, głosu, a także innych danych osobowych, w celach naukowych.

* * *

OŚWIADCZENIE O WYRAŻENIU ZGODY NA ROZPOWSZECHNIANIE WIZERUNKU

Oświadczam, że wyrażam nieodpłatnie zgodę na rozpowszechnianie przez Administratora mojego wizerunku, głosu i wypowiedzi utrwalonych na fotografiach i nagraniach audio - wideo wykonanych podczas wywiadu. Powyższa zgoda obejmuje zezwolenie dla Administratora na:

- wykonywanie fotografii, filmów i nagrań podczas wywiadu i ich wykorzystywanie na terytorium Rzeczypospolitej Polskiej oraz poza jej granicami, bez ograniczeń czasowych oraz bez konieczności każdorazowego ich zatwierdzania;
- utrwalanie, powielanie i zwielokrotnianie mojego wizerunku, głosu i wypowiedzi dowolną techniką oraz na ich wprowadzanie do pamięci komputera i zapisywanie na nośnikach danych;
- publiczne udostępnianie mojego wizerunku, głosu i wypowiedzi, w szczególności w taki sposób, aby każdy mógł mieć do nich dostęp w miejscu i w czasie przez siebie wybranym, w tym w sieci Internet, a także we wszelkich innych materiałach informacyjnych, publikacjach, w tym publikacjach wydawanych przez Administratora.

Jednocześnie oświadczam, że zrzekam się niniejszym wszelkich roszczeń istniejących i przyszłych względem Administratora z tytułu udzielenia niniejszej zgody, w tym roszczeń o wynagrodzenie za wykorzystanie mojego wizerunku, głosu i moich wypowiedzi na potrzeby określone w niniejszym oświadczeniu.

Zapoznałem/am się z powyższą treścią, w pełni ją rozumiem i dobrowolnie akceptuję.

.....
Miejscowość, data	Czytelny podpis

Anhang 4

Liste der Informantinnen und Informanten im LangGener-Korpus (deutsch-polnisches Subkorpus)

Generation Polen¹

AL_PAR, ein Mann, wohnhaft in A., Woiwodschaft Ermland-Masuren, Kreis Kętrzyn; geboren 1933 in A., Provinz Ostpreußen, Regierungsbezirk Allenstein (heute A., Woiwodschaft Ermland-Masuren, Kreis Kętrzyn)

BQ_RAC, eine Frau, wohnhaft in Racibórz, Woiwodschaft Schlesien, Kreis Racibórz; geboren 1941 in A., Provinz Niederschlesien, Kreis Namslau (heute A., Woiwodschaft Opole, Kreis Namysłów)

BT_KLO, ein Mann, wohnhaft in A., Woiwodschaft Opole, Kreis Kędzierzyn-Koźle; geboren 1935 in A., Provinz Oberschlesien, Kreis Cosel (heute A., Woiwodschaft Opole, Kreis Kędzierzyn-Koźle)

CS_GIZ, eine Frau, wohnhaft in Giżycko, Woiwodschaft Ermland-Masuren, Kreis Giżycko; geboren 1942 in Angerburg, Provinz Ostpreußen, Regierungsbezirk Allenstein (heute Węgorzewo, Woiwodschaft Ermland-Masuren, Kreis Giżycko)

EM_UKT, ein Mann, wohnhaft in A., Woiwodschaft Ermland-Masuren, Kreis Pisz; geboren 1940 in A., Provinz Ostpreußen, Regierungsbezirk Allenstein (heute Woiwodschaft Ermland-Masuren, Kreis Pisz)

EU_KET, eine Frau, wohnhaft in Kętrzyn, Woiwodschaft Ermland-Masuren, Kreis Kętrzyn; geboren 1939 in A., Provinz Ostpreußen, Regierungsbezirk Allenstein (heute A., Woiwodschaft Ermland-Masuren, Kreis Kętrzyn)

FL_BON, eine Frau, wohnhaft in A., Woiwodschaft Westpommern, Kreis Koszalin; geboren 1940 in A., Provinz Pommern, Regierungsbezirk Köslin (heute A., Woiwodschaft Zachodniopomorskie, Kreis Koszalin)

FT_ELB, eine Frau, wohnhaft in Elbląg, Woiwodschaft Ermland-Masuren, Kreis Elbląg; geboren 1934 in A., Provinz Westpreußen, Regierungsbezirk Danzig, Kreis Elbing (heute A., Woiwodschaft Ermland-Masuren, Kreis Elbląg)

HD_SWI, eine Frau, wohnhaft in A., Woiwodschaft Westpommern, Kreis Świdwin; geboren 1934 in A., Provinz Pommern, Köslin (heute A., Woiwodschaft Zachodniopomorskie, Kreis Białogard)

¹ Kleine Ortschaften werden aus Datenschutzgründen nur in Abkürzung genannt.

IL_PIL, eine Frau, wohnhaft in Piła, Woiwodschaft Großpolen, Kreis Piła; geboren 1937 in Schneidemühl, Grenzmark Posen-Westpreußen (heute Piła, Woiwodschaft Großpolen, Kreis Piła)

IT_ELB, eine Frau, wohnhaft in Elbląg, Woiwodschaft Ermland-Masuren, Kreis Elbląg; geboren 1936 in A., Provinz Ostpreußen, Regierungsbezirk Allenstein (heute A., Woiwodschaft Ermland-Masuren, Kreis Elbląg)

IT_GLO, eine Frau, wohnhaft in Głogówek, Woiwodschaft Opole, Kreis Prudnik; geboren 1938 in Oberglogau, Provinz Oberschlesien, Kreis Neustadt O.S. (heute Głogówek, Woiwodschaft Opole, Kreis Prudnik)

JT_BAR, eine Frau, wohnhaft in Barwice, Woiwodschaft Westpommern, Kreis Szczecinek; geboren 1942 in Bärwalde, Provinz Pommern, Regierungsbezirk Köslin (heute Barwice, Woiwodschaft Westpommern, Kreis Szczecinek)

JT_STA, eine Frau, wohnhaft in Stargard, Woiwodschaft Westpommern, Kreis Stargard; geboren 1939 in Provinz Pommern, Kreis Rügen (heute Bundesland Mecklenburg-Vorpommern, Landkreis Rügen)

KL_GDA, eine Frau, wohnhaft in Gdańsk/Sopot, Woiwodschaft Pommern; geboren 1935 in Danzig, Provinz Westpreußen, Regierungsbezirk Danzig (heute Gdańsk, Woiwodschaft Pommern)

KL_PIA, ein Mann, wohnhaft in A., Woiwodschaft Ermland-Masuren, Kreis Pisz; geboren 1942 in A., Provinz Ostpreußen, Regierungsbezirk Allenstein (heute A., Woiwodschaft Ermland-Masuren, Kreis Pisz)

KQ_GOG, eine Frau, wohnhaft in Gogolin, Woiwodschaft Opole, Kreis Krapkowice; geboren 1933 in Gogolin, Provinz Oberschlesien, Kreis Groß Strehlitz (heute Gogolin, Woiwodschaft Opole, Kreis Krapkowice)

LA_ELB, eine Frau, wohnhaft in Elbląg, Woiwodschaft Ermland-Masuren, Kreis Elbląg; geboren 1937 in Danzig, Provinz Westpreußen, Regierungsbezirk Danzig (heute Gdańsk, Woiwodschaft Pommern)

LA_MIK, eine Frau, wohnhaft in A., Woiwodschaft Ermland-Masuren, Kreis Elk; geboren 1932 in A., Provinz Ostpreußen, Regierungsbezirk Allenstein (heute Woiwodschaft Ermland-Masuren, Kreis Elk)

LU_KLO, ein Mann, wohnhaft in A., Woiwodschaft Opole, Kreis Kędzierzyn-Koźle; geboren 1940 in A., Provinz Schlesien (ab 1941 Provinz Oberschlesien), Kreis Cosel (heute A., Woiwodschaft Opole, Kreis Kędzierzyn-Koźle)

NH_PIL, eine Frau, wohnhaft in Piła, Woiwodschaft Großpolen, Kreis Piła; geboren 1935 in Schneidemühl, Grenzmark Posen-Westpreußen (heute Piła, Woiwodschaft Großpolen, Kreis Piła)

PQ_GDY, eine Frau, wohnhaft in Gdynia, Woiwodschaft Pommern; geboren 1932 in A., Provinz Ostpreußen, Regierungsbezirk Allenstein (heute A., Woiwodschaft Ermland-Masuren, Kreis Pisz)

QK_BOB, ein Mann, wohnhaft in A., Woiwodschaft Westpommern, Kreis Koszalin; geboren 1944 in A., Provinz Pommern, Regierungsbezirk Köslin, (heute A., Woiwodschaft Westpommern, Kreis Koszalin)

QN_BRO, ein Mann, wohnhaft in A., Woiwodschaft Opole, Kreis Krapkowice; geboren 1942 in Beuthen, Provinz Oberschlesien, Kreis Beuthen O.S. (heute Bytom, Woiwodschaft Schlesien, Kreis Bytom)

SB_KLA, eine Frau, wohnhaft in A., Woiwodschaft Westpommern, Kreis Koszalin; geboren 1938 in A., Provinz Pommern, Regierungsbezirk Köslin (heute A., Woiwodschaft Zachodniopomorskie, Kreis Koszalin)

SL_OPO, eine Frau, wohnhaft in Opole, Woiwodschaft Opole, Stadtkreis Opole; geboren 1939 in Oppeln, Provinz Schlesien (ab 1941 Provinz Oberschlesien), Kreis Oppeln (heute Opole, Woiwodschaft Opole, Stadtkreis Opole)

SO_NIE, eine Frau, wohnhaft in A., Woiwodschaft Westpommern, Kreis Koszalin; geboren 1934 in Dortmund, Provinz Westfalen, Bezirk Arnberg (heute Dortmund, Bundesland Nordrhein-Westfalen, Bezirk Arnberg)

SV_NAK, ein Mann, wohnhaft in A., Woiwodschaft Opole, Landkreis Opole; geboren 1934 in A., Provinz Oberschlesien, Kreis Groß Strehlitz (heute A., Woiwodschaft Opole, Kreis Strzelce Opolskie)

UM_JAN, eine Frau, wohnhaft in A., Woiwodschaft Ermland-Masuren, Kreis Kętrzyn; geboren 1933 in A., Provinz Ostpreußen, Regierungsbezirk Allenstein (heute A., Woiwodschaft Ermland-Masuren, Kreis Kętrzyn)

XL_PIL, ein Mann, wohnhaft in Piła, Woiwodschaft Großpolen, Kreis Piła; geboren 1937 in Schneidemühl, Grenzmark Posen-Westpreußen (heute Piła, Woiwodschaft Großpolen, Kreis Piła)

Generation Deutschland

BK_AUG, eine Frau, wohnhaft in Augsburg, Bundesland Bayern, Landkreis Augsburg; geboren 1970 in Nysa, Woiwodschaft Opole, Kreis Nysa (heute Nysa, Woiwodschaft Opole, Kreis Nysa); ausgereist 1988

BK_HAM, ein Mann, wohnhaft in Hamburg, Bundesland Hamburg; geboren 1961 in Nysa, Woiwodschaft Opole, Kreis Nysa (heute Nysa, Woiwodschaft Opole, Kreis Nysa); ausgereist 1988

BL_AUG, ein Mann, wohnhaft in Augsburg, Bundesland Bayern, Landkreis Augsburg; geboren 1967 in Tychy, Woiwodschaft Katowice, Kreis Tychy (heute Tychy, Woiwodschaft Schlesien, Kreis Tychy); ausgereist 1988

BM_HAM, ein Mann, wohnhaft in Hamburg, Bundesland Hamburg; geboren 1971 in Piekary Śląskie, Woiwodschaft Katowice, Kreis Tarnogóra (heute Piekary Śląskie, Woiwodschaft Schlesien); ausgereist 1988

BN_WUP, ein Mann, wohnhaft in Wuppertal, Bundesland Nordrhein-Westfalen; geboren 1964 in A., Woiwodschaft Opole, Kreis Prudnik (heute A., Woiwodschaft Opole, Kreis Prudnik); ausgereist 1990

CH_BER, eine Frau, wohnhaft in Berlin, Bundesland Berlin; geboren 1967 in Szczecin, Woiwodschaft Szczecin (heute Szczecin, Woiwodschaft Westpommern); ausgereist 2011

CL_BIE, eine Frau, wohnhaft in Bielefeld, Bundesland Nordrhein-Westfalen; geboren 1965 in Głogówek, Woiwodschaft Opole, Kreis Prudnik (heute Głogówek, Woiwodschaft Opole, Kreis Prudnik); ausgereist 1979

FL_HAM, eine Frau, wohnhaft in Hamburg, Bundesland Hamburg; geboren 1957 in Kłodawa, Woiwodschaft Konin, Kreis Koło (heute Kłodawa, Woiwodschaft Großpolen); ausgereist 1980

HM_STE, ein Mann, wohnhaft in Steinen, Bundesland Baden-Württemberg, Landkreis Lörrach; geboren 1965 in Głogówek, Woiwodschaft Opole, Kreis Prudnik (heute Głogówek, Woiwodschaft Opole, Kreis Prudnik); ausgereist 1988

JS_STR, eine Frau, wohnhaft in Straubing, Bundesland Bayern, Landkreis Straubing-Bogen; geboren 1962 in Bieruń, Woiwodschaft Katowice, Kreis Tychy (heute Bieruń, Woiwodschaft Schlesien, Kreis Bieruń-Łędziny); ausgereist 1988

KC_HAM, eine Frau, wohnhaft in Hamburg, Bundesland Hamburg; geboren 1954 in Czersk, Woiwodschaft Bydgoszcz, Kreis Chojnice (heute Czersk, Woiwodschaft Pomorskie, Kreis Chojnice); ausgereist 1989

KI_VIM, ein Mann, wohnhaft in Bühl, Bundesland Baden-Württemberg, Landkreis Rastatt; geboren 1968 in Koźle, Woiwodschaft Opole, Kreis Koźle (heute Koźle, Woiwodschaft Opole, Kreis Kędzierzyn-Koźle); ausgereist 1991

KL_BIE, ein Mann, wohnhaft in Bielefeld, Bundesland Nordrhein-Westfalen; geboren 1962 in Głogówek, Woiwodschaft Opole, Kreis Prudnik (heute Głogówek, Woiwodschaft Opole, Kreis Prudnik); ausgereist 1988

KN_HAT, eine Frau, wohnhaft in Hattingen, Bundesland Nordrhein-Westfalen, Ennepe-Ruhr-Kreis; geboren 1965 in Głogówek, Woiwodschaft Opole, Kreis Prudnik (heute Głogówek, Woiwodschaft Opole, Kreis Prudnik); ausgereist 1988

LC_BUR, eine Frau, wohnhaft in Regensburg, Bundesland Bayern; geboren 1971 in Świętochłowice, Woiwodschaft Katowice, Kreis Świętochłowice (heute Woiwodschaft Schlesien, Kreis Świętochłowice); ausgereist 1993

LO_LAM, eine Frau, wohnhaft in A., Bundesland Rheinland-Pfalz, Landkreis Bad Dürkheim; geboren 1970 in Koźle, Woiwodschaft Opole, Kreis Koźle (heute Koźle, Woiwodschaft Opole, Kreis Kędzierzyn-Koźle); ausgereist 1990

NA_HAM, eine Frau, wohnhaft in Hamburg, Bundesland Hamburg; geboren 1965 in Koźuchów, Woiwodschaft Zielona Góra, Kreis Koźuchów, später Nowa Sól (heute Koźuchów, Woiwodschaft Lubuskie, Kreis Nowa Sól); ausgereist 1983

ND_REG, ein Mann, wohnhaft in Regensburg, Bundesland Bayern; geboren 1964 in Nysa, Woiwodschaft Opole, Kreis Nysa (heute Nysa, Woiwodschaft Opole, Kreis Nysa); ausgereist 1988

NE_BUR, eine Frau, wohnhaft in Regensburg, Bundesland Bayern; geboren 1972 in Świętochłowice, Woiwodschaft Katowice, Kreis Świętochłowice (heute Świętochłowice, Woiwodschaft Schlesien, Kreis Świętochłowice); ausgereist 1993

NF_PAD, eine Frau, wohnhaft in Paderborn, Bundesland Nordrhein-Westfalen, Kreis Paderborn; geboren 1968 in Krapkowice, Woiwodschaft Opole, Kreis Krapkowice (heute Krapkowice, Woiwodschaft Opole, Kreis Krapkowice); ausgereist 1990

NI_AUG, ein Mann, wohnhaft in Augsburg, Bundesland Bayern, Landkreis Augsburg; geboren 1965 in A., Woiwodschaft Opole, Kreis Nysa (heute A., Woiwodschaft Opole, Kreis Nysa); ausgereist 1988

NL_REG, ein Mann, wohnhaft in Regensburg, Bundesland Bayern; geboren 1967 in Kłodzko, Woiwodschaft Wrocław, Kreis Kłodzko (heute Kłodzko, Woiwodschaft Dolnośląskie Kreis Kłodzko); ausgereist 1980

NT_LUB, ein Mann, wohnhaft in Lübeck, Bundesland Schleswig-Holstein; geboren 1961 in A., Woiwodschaft Opole, Kreis Koźle (heute A., Woiwodschaft Opole, Kreis Kędzierzyn-Koźle); ausgereist 1981

QX_BER, ein Mann, wohnhaft in Berlin, Bundesland Berlin; geboren 1989 in Sulechów, Woiwodschaft Zielona Góra, (heute Sulechów, Woiwodschaft Lubuskie, Kreis Zielona Góra); ausgereist 2012

SO_LAM, ein Mann, wohnhaft in A., Bundesland Rheinland-Pfalz, Landkreis Bad Dürkheim; geboren 1964 in A., Woiwodschaft Opole, Kreis Koźle (heute Woiwodschaft Opole, Kreis Kędzierzyn-Koźle); ausgereist 1991

UC_LUC, ein Mann, wohnhaft in Luckenwalde, Bundesland Brandenburg, Landkreis Teltow-Fläming; geboren 1966 in A., Woiwodschaft Szczecin, Kreis Chojna (heute Woiwodschaft Zachodniopomorskie, Kreis Gryfino); ausgereist 1990

XL_MUN, ein Mann, wohnhaft in München, Bundesland Bayern; geboren 1964 in Prudnik, Woiwodschaft Opole, Kreis Prudnik (heute Prudnik, Woiwodschaft Opole, Kreis Prudnik); ausgereist 1987

XX_HAM, ein Mann, wohnhaft in Hamburg, Bundesland Hamburg; geboren 1951 in Siemianowice Śląskie, Woiwodschaft Katowice, Kreis Tarnogóra (heute Siemianowice Śląskie, Woiwodschaft Schlesien); ausgereist 1989

Anhang 5

Liste der Respondentinnen und Respondenten im deutsch-tschechischen Korpus *Jazyky v migraci* (Sprachen in Migration)

Spätaussiedlerinnen und Spätaussiedler

BJ_MUN, eine Frau, wohnhaft in München, geboren 1956 im Okres Jablonec nad Nisou, ausgereist 1968

CI_BUB, eine Frau, wohnhaft im Landkreis Günzburg, geboren 1948 im Okres Olomouc, ausgereist 1964

II_ROD, ein Mann, wohnhaft im Landkreis Offenbach, geboren 1955 im Okres Liberec, ausgereist 1967

IL_REG, eine Frau, wohnhaft in Regensburg, geboren 1957 im Okres Klatovy, ausgereist 1968

KL_HEI, ein Mann, wohnhaft in Heidenheim, geboren 1951 im Okres Náchod, ausgereist 1967

LS_ZEI, eine Frau, wohnhaft in Regensburg, geboren 1958 in Karlsbad, ausgereist 1967

QC_WUN, ein Mann, wohnhaft in Wunsiedel, geboren 1952 im Okres Cheb, ausgereist 1967

IL_BLA, eine Frau, wohnhaft im Landkreis Cham, geboren 1958 im Okres Sokolov, ausgereist 1967

BL_BLA, ein Mann, wohnhaft im Landkreis Cham, geboren 1952 im Okres Děčín, ausgereist 1966

TU_HAG, eine Frau, wohnhaft im Landkreis Regensburg, geboren 1953 im Okres Sokolov, ausgereist 1968

Migrantinnen und Migranten

KC_ISM, ein Mann, wohnhaft im Landkreis München, geboren 1951 in Karlsbad, ausgereist 1969

KG_MUN, eine Frau, wohnhaft in München, geboren 1948 im Okres České Budějovice, ausgereist 1968

NT_GAU, eine Frau, wohnhaft in München, geboren 1953 in Ostrava, ausgereist 1968

QC_MUN, ein Mann, wohnhaft in München, geboren 1953 in Prag, ausgereist 1968

ER_HER, eine Frau, wohnhaft im Landkreis Starnberg, geboren 1951 in Prag, ausgereist 1970

KN_MUN, eine Frau, wohnhaft in München, geboren 1951 in Prag, ausgereist 1969

HD_MUN, ein Mann, wohnhaft in München, geboren 1968 in Prag, ausgereist 1986

LQ_UNT, ein Mann, wohnhaft im Landkreis München, geboren 1941 in Prag, ausgereist 1969

MQ_UNT, eine Frau, wohnhaft im Landkreis München, geboren 1945 in Prag, ausgereist 1969

WG_MUN, ein Mann, wohnhaft in München, geboren 1945 in Westböhmen, ausgereist 1969

HANSEN • ZIELIŃSKA (Hg.)

Soziolinguistik trifft Korpuslinguistik

Die kollektive Monographie ist ein Beitrag zur empirisch orientierten Erforschung deutsch-polnischer bzw. -tschechischer Zweisprachigkeit, der sozio- und korpuslinguistische Perspektiven zusammenführt. Sie ist im Rahmen des deutsch-polnischen Projekts *Language across generations: contact induced change in morpho-syntax in German-Polish bilingual speech* entstanden. Vorgestellt wird das deutsch-polnische Korpus *LangGener*, das Interviews umfasst, die von Bilingualen aus zwei nach Spracherwerbskriterien definierten Generationen aufgenommen wurden. Das in einem Unterprojekt erstellte deutsch-tschechische Korpus basiert auf Interviews mit zwei verschiedenen Migrationsgruppen. Ein korpuslinguistisches Novum ist die Annotation von sprachbiographischen Fakten und von Sprachkontaktphänomenen wie Replikationen. Der Band stellt zudem eine praktische Anleitung für die Erstellung soziolinguistischer Korpora zur Mehrsprachigkeit dar.

