

Roland Kehrein und Lars Vorberger

6 Dialekt- und Variationskorpora

Abstract: Dialekt- und Variationskorpora enthalten Sprachdaten zu bestimmten Varietäten des Deutschen und/oder bilden sprachliche Variation in der areal-horizontalen oder der situativ-vertikalen Dimension ab. Somit stellen sie für die (moderne) Regionalsprachenforschung eine wichtige Datengrundlage dar. Im vorliegenden Beitrag erfolgt nach einer kurzen begrifflichen Erläuterung von Dialekt- und Variationskorpora sowie deren Spezifika hinsichtlich ihrer Zusammenstellung (Datenerhebung) eine schlaglichtartige chronologische und systematische Präsentation von Dialektatlaskorpora, variationslinguistischen Tonkorpora und modernen variationslinguistischen Korpora. Über deren Inhalt hinaus werden jeweils auch die Formen ihrer Erschließung sowie ihre Zugänglichkeit im Rahmen von Online-Plattformen beschrieben.

Keywords: Sprachvariation, Dialekt, Dialektatlas, Deutsch heute, Pfeffer, SiN, REDE, Regionalsprache, Sprachatlas, Wenker, Zwirner

1 Einleitung

Im vorliegenden Beitrag wird es um Dialekt- und Variationskorpora gehen, Korpora von Sprachdaten also, in denen sprachliche Variation und/oder sprachliche Varietäten erfasst sind. Die Frage, was ein Korpus ist, wird im vorliegenden Band ausführlich behandelt. Aus diesem Grund werden wir uns nicht intensiver mit dem Korpusbegriff auseinandersetzen. Wir legen eine Minimaldefinition zugrunde, nach der ein Korpus die folgenden Eigenschaften hat:

- Es handelt sich um eine „geschlossen verfügbare Sammlung von Texten“ (Lehmann 2007: 17).
- Diese Texte gehören einer bestimmten Kategorie an.
- Es kann sich um geschriebene Sprache und um gesprochene Sprache handeln, wobei letztere in irgendeiner Form transkribiert sein muss (vgl. Friginal & Hardy 2014: 20).

Roland Kehrein, Forschungszentrum Deutscher Sprachatlas, Pilgrimstein 16, D-35032 Marburg, E-Mail: kehrein@uni-marburg.de

Lars Vorberger, Forschungszentrum Deutscher Sprachatlas, Pilgrimstein 16, D-35032 Marburg, E-Mail: lars.vorberger@deutscher-sprachatlas.de

Korpora sind in vielen Fällen zusätzlich über Metadaten sowie durch Annotationen erschlossen und können dadurch mit Recherchewerkzeugen durchsucht werden. Da diese Eigenschaften aber nicht auf alle in diesem Beitrag präsentierten Korpora zutreffen, schließen wir sie nicht in unsere Definition mit ein.

Was versteht man auf der anderen Seite nun unter Variation von Sprache? Diese lässt sich auf mehreren Dimensionen beobachten und beschreiben. Unterschieden werden (vgl. Coseriu 1992: 280):

- Sprachvariation in Abhängigkeit vom geographischen Raum: diatopische Dimension,
- Sprachvariation in Abhängigkeit von der Zeit: diachrone Dimension,
- Sprachvariation in Abhängigkeit von der Kommunikationssituation: diaphasische Dimension und
- Sprachvariation in Abhängigkeit von sozialen Faktoren wie Alter/Generation, Bildung, Geschlecht usw.: diastratische Dimension.

Die vor allem in jüngster Zeit für das Deutsche intensiv untersuchte Sprachvariation auf der vertikalen Dimension zwischen Standardsprache und Dialekt (vgl. etwa Lenz 2003; Lameli 2004; Christen et al. 2010; Kehrein 2012; Rocholl 2015; Vorberger 2017, um nur die rezent vorgelegten Buchpublikationen zu nennen) integriert Aspekte der diaphasischen und diastratischen Dimension, da sie von situativen Faktoren abhängt, zu denen beispielsweise auch Eigenschaften des Gesprächspartners gehören. Dialekt- und Variationskorpora sind daher entsprechend bestimmbar als abgeschlossene Sammlungen von Texten gesprochener Sprache, die zu Dokumentationszwecken oder für bestimmte Forschungsziele (z. B. Erstellung von Sprachatlanten, Untersuchung intersituativer Sprachvariation) erhoben wurden. In den meisten Fällen handelt es sich nicht um Zusammenstellungen vorhandener Texte, sondern um eigens zu den genannten Zwecken angefertigte Neuerhebungen.¹ Die wichtigste Variationsdimension im Deutschen – dies ist aus der Geschichte des Gesamtsprachsystems Deutsch zu erklären – ist die diatopische Dimension (vgl. auch Bellmann 1994: 2). Dialekt- und Variationskorpora enthalten daher immer Sprachdaten aus verschiedenen Regionen und Orten eines definierten Untersuchungsgebietes des deutschen Sprachraums. Um die Vergleichbarkeit der Daten zu gewährleisten, werden bei

1 Ausnahmen bilden Sammlungen von Sprachaufzeichnungen, die ursprünglich nicht für sprachwissenschaftliche Zwecke angefertigt wurden. Als Beispiele lassen sich Zusammenstellungen (i) von Mitschnitten von Gemeinderatssitzungen (vgl. Lameli 2004) sowie (ii) von Notrufgesprächen (vgl. Kehrein 2006; Christen, Hove & Petkova 2015) nennen, die dann jeweils hinsichtlich bestimmter Fragestellungen ausgewertet wurden. Diese Korpora sind aus verschiedenen Gründen in der Regel aber nicht zugänglich, weshalb sie im vorliegenden Beitrag keine weitere Berücksichtigung finden.

der Datensammlung in der Regel klare Kriterien an die Informantenauswahl und an die Erhebungsmethoden angelegt (vgl. auch Hunston 2008). Bei der Informantenauswahl wird zunächst einmal ein besonderes Augenmerk auf die Orts-/Regionsfestigkeit gelegt. Das bedeutet, dass die Informanten sowie möglichst auch ihre Eltern und Großeltern aus dem Ort/der Region stammen und sie keine längeren Aufenthalte in anderen Regionen gehabt haben sollten. Darüber hinaus werden Informanten beispielsweise nach ihrem Alter bzw. ihrer Generationenzugehörigkeit, ihrem (früheren) Beruf und/oder ihrem Bildungsniveau ausgewählt. Die außersprachliche Variable Geschlecht wird bei großräumig angelegten Projekten in der Regel nicht systematisch variiert. Dies hat zum einen forschungspraktische Gründe, denn die Informantenzahl würde sich jeweils verdoppeln. Zum anderen ist noch nicht abschließend geklärt, inwieweit das Geschlecht einen Einfluss auf das variative Sprachverhalten und die Dialektkompetenz hat.²

Was die Erhebungsmethoden angeht, so lässt sich hinsichtlich Dialekt- und Variationskorpora auf einer übergeordneten Ebene zunächst einmal die direkte von der indirekten Erhebung unterscheiden. Während bei ersterer die Sprachdaten (meist vor Ort) im unmittelbaren Kontakt mit dem Informanten erfasst werden (Sprachaufzeichnung oder direkte Transkription), arbeitet man bei der zweiten Methode mit Fragebogen, die von den Informanten schriftlich auszufüllen sind. Über diese Möglichkeiten der Datenerfassung hinaus lässt sich differenzieren, welcher Bereich sprachlicher Variation erhoben werden soll. Traditionell ging es um die Dialekte, und zwar um eine möglichst „bodenständige“ Form der Dialekte. Es wurde die Dialektkompetenz der Informanten in direkten oder indirekten Befragungen erhoben. Solche Befragungen können prinzipiell alle linguistischen Systemebenen zum Gegenstand haben. Sie basieren auf einem standardisierten Set an Abfragewörtern und -sätzen, sodass für alle berücksichtigten Orte und Sprecher vergleichbares Sprachmaterial vorliegt. In jüngeren Projekten werden per Abfrage und durch lautes Vorlesen auch Daten zur Standardkompetenz der Informanten, ihr individuell bestes Hochdeutsch, erhoben.³ Außer der Kompetenzabfrage können Dialekt- und

² Zu dem Ergebnis, dass der Faktor Geschlecht keine Rolle spielt, gelangt beispielsweise Auer (1990) in seiner Untersuchung zur Sprachvariation in der Konstanzer Stadtsprache. Während im Falle von Dialektatlanten vielfach angenommen wird, der geeignetste und typische Informant sei „non mobile, old, rural, male“ (NORM; vgl. Chambers & Trudgill 1998), geben beispielsweise Jaberg & Jud (1928) an, dass Frauen über eine besonders ausgeprägte und differenziertere Dialektkenntnis verfügten (vgl. Jaberg & Jud 1928: 189–193).

³ Als Vorbild für die Erhebung dieses Gegenstands kann Wilhelm Viëtor mit seiner Reihe „Beiträge zur Statistik der Aussprache des Schriftdeutschen“ (vgl. z. B. Viëtor 1888) gelten. In jüngerer Zeit hat Werner König sowohl mit Informanten des *Sprachatlas von Bayerisch-Schwaben*

Variationskorpora auch Texte frei formulierter Monologe oder kurze (Nach-)Erzählungen enthalten (z. B. die von Fleischer & Gadmer 2002 herausgegebenen Schweizer Aufnahmen des Züricher Phonogrammarchivs). Auch in diesen Fällen zielt die Erhebung meist auf den Dialekt der Informanten. In jüngeren Projekten schließlich werden noch Daten zum variativen Sprachgebrauch der Informanten erhoben. Aus Gründen der Vergleichbarkeit handelt es sich um „freie“ Gespräche, die in Situationen aufgezeichnet werden, bei denen möglichst viele außersprachliche Variablen konstant gehalten werden können. Dazu gehört vor allem das (sprachbiographische) Interview, in dem ein Explorator als teilnehmender Beobachter die Gesprächsbeiträge der Informanten durch möglichst offene Fragen initiiert, dazu gehören aber auch Situationen, in denen zwei oder mehr Informanten in einer bestimmten Weise miteinander interagieren (z. B. sog. Freundesgespräche oder auch Wegbeschreibungen in der sog. Map Task).

Wir werden im Folgenden Dialekt- und Variationskorpora des Deutschen vorstellen. Dabei werden wir analog zur Chronologie ihrer jeweils ersten Erhebung mit Datensammlungen für Dialekt- bzw. Sprachatlanten beginnen, dann Tonkorpora zu deutschen Varietäten beschreiben oder kurz tabellarisch erfassen (s. Tab. 6.1), bevor drei jüngere Projekte vorgestellt werden, in denen mehrere Variationsdimensionen berücksichtigt werden. Nach der Präsentation der Korpora wird es um ihre Aufbereitung und, was eng damit verknüpft ist, um ihre Erschließung (z. B. über Register) und Zugänglichkeit (z. B. über Datenbank-Frontends) gehen. Exemplarisch werden wir hier die *Datenbank für gesprochenes Deutsch* (DGD2) und die Internetplattform *Regionalsprache.de* (REDE) vorstellen, auf denen jeweils in verschiedenen Korpora recherchiert werden kann.

2 Dialektatlaskorpora

Über die Datensammlungen für Dialektatlanten wurden bereits lange bevor man überhaupt an Korpuslinguistik gedacht hat, Korpora als abgeschlossene Sammlungen von Texten gesprochener Sprache zu Dokumentationszwecken und für bestimmte Forschungsziele erstellt. Allerdings waren diese Datensammlungen in Sprachatlasprojekten nicht als Korpuserstellung gedacht. Daher sind sie in Form der Rohdaten in der Regel auch nicht zugänglich und für weitere Analysen verwendbar. Die einzige, planmäßige (!) Aufbereitung dieser

(König 1996–2009) (vgl. dazu Wecker-Kleiner 2009) als auch für den *Atlas zur Aussprache des Schriftdeutschen in der Bundesrepublik Deutschland* (König 1989) solche standardorientierten Sprachdaten erhoben. Berücksichtigung findet das individuell beste Hochdeutsch auch in den in Abschnitt 4 beschriebenen Projekten zur Sprachvariation auf der vertikalen Dimension.

Korpora besteht in der Sprachkartenerstellung für den jeweiligen Atlas. Ein häufig nicht unbeträchtlicher Teil der gesammelten Daten bleibt somit unausgewertet (vgl. zu dieser Problematik auch Kunst & Barbiers 2010: 401–402).

2.1 Sprachatlas des Deutschen Reichs

Mit der von Georg Wenker im Zeitraum zwischen 1876 und 1887 durchgeführten Sprachdatenerhebung für seinen *Sprachatlas des Deutschen Reichs* (Wenker 1889–1923) liegt das bis heute umfangreichste Korpus zu den Dialekten einer Einzelsprache vor. Für ihn stand dabei nicht ein dokumentarisches Interesse im Vordergrund, sondern ein wissenschaftliches. Entsprechend bildete die Datensammlung für ihn die Grundlage für die Erstellung und die Auswertung von Sprachkarten. Wenkers erklärtes Ziel war es, eine über große Distanzen hinweg vergleichende Dialektkunde zu betreiben (vgl. Wenker 1886: 190). Dabei war ihm vollkommen klar, dass eine detaillierte Totalerhebung aller Dialekte unmöglich war: Er beschränkte sich daher auf die Übertragung von 40 Sätzen in den Dialekt seiner Erhebungsorte. Sein Motto war, „lieber *Weniges* aus möglichst *allen*, als *Vieles* aus einer *ungenügenden* Zahl an Ortschaften einzusammeln“ (Wenker 1881: VIII; im Orig. Hervorhebung durch Sperrung). Die Datenerhebung erfolgte indirekt durch den Versand von Fragebogen an alle Schulorte des damaligen Deutschen Reichs. Die Lehrer wurden gebeten, gemeinsam mit den Schülern, „die obenstehenden Fragen gewissenhaft zu beantworten sowie eine Uebersetzung der einliegenden hochdeutschen Sätzchen in die ortsübliche Mundart umstehend einzutragen“ (vgl. Fleischer 2017: 188). Insgesamt lagen Wenker nach Abschluss seiner Erhebung 48.505 (deutsche und fremdsprachliche) Fragebogen vor, von denen allerdings knapp 2.500 als „unbrauchbar“ bezeichnet wurden (vgl. Wenker 2013: 2). Für den Sprachatlas wurden 46.011 Fragebogen aus 42.496 Erhebungsorten verarbeitet (vgl. Wenker 2013: 9). Bei den Fragebogen handelt es sich um eine geschlossene Sammlung von Texten einer Kategorie. Die Texte sind als Transkriptionen gesprochener Sprache durch die Informanten, Lehrer und Schüler, zu bewerten. Derzeit liegen die Transkriptionen lediglich in handschriftlicher Form vor. Obwohl in dem Projekt *Digitaler Wenker-Atlas* (DiWA) zwischen 2001 und 2009 alle Wenkerkarten und die Original-Erhebungsbogen digitalisiert und im Internet zur freien Verfügung bereitgestellt wurden,⁴ sind bisher nur die kartierten Teile des Korpus nach Kartenthemen durchsuchbar. Die Fragebogen sind lediglich als Bilddateien zugänglich, eine Volltextdigitalisierung steht dagegen noch

⁴ Seit 2010 sind alle Karten des *Digitalen Wenker-Atlas* über die Forschungsplattform <https://www.regionalsprache.de> (letzter Zugriff: 31. 7. 2017) abrufbar.

aus. Die Karten sind darüber hinaus durch ein morphologisches (vgl. Rabanus 2005), ein phonetisch-phonologisches (vgl. Lameli 2009) und ein historisches Register erschlossen (siehe auch Abschnitt 5).

Die erste Auswertung des Korpus haben Wenker und seine Kollegen selbst vorgenommen, indem sie auf seiner Basis einen Sprachatlas mit 1.668 Karten(blättern) zu 576 Kartenthemen erstellt haben. Darüber hinaus hat Wenker „eine zum Buch gebundene 59-seitige Einleitung zum Sprachatlas des Deutschen Reichs, 431 erläuternde Texthefte im Umfang von ca. 2600 Seiten zu 356 wortbezogenen Karten des Atlases sowie ein separates Textheft zu den Sprachverhältnissen im Nordosten des Untersuchungsgebietes“ (Wenker 2013: XVII) verfasst. Weitere Auswertungen im Sinne von raumbezogenen Korpusanalysen, die sich unter anderem auf die über den Sprachatlas erschlossenen Korpusteile stützen, sind Ferdinand Wredes Dialekteinteilung (vgl. Wrede 1937), Peter Wiesingers *Phonetisch-phonologische Untersuchungen zur Vokalentwicklung in den deutschen Dialekten* (Wiesinger 1970) sowie sein Beitrag „Die Einteilung der deutschen Dialekte“ (Wiesinger 1983). Echte raumstatistische Auswertungen des Wenkermaterials hat – für das Gebiet der Bundesrepublik Deutschland – erstmals Alfred Lameli (v. a. Lameli 2013) vorgelegt. Als Grundlage für diese Analysen hat er für ein Set von 66 lautlichen und morphologischen Variationsphänomenen die Varianten für die 439 Landkreise Deutschlands in einer Datenbank erfasst. Auf dieser Datenbasis wurden nicht nur raumstatistische linguistische Analysen durchgeführt, sondern auch die Sprachdaten in ihrem Verhältnis zu außersprachlichen Daten betrachtet (vgl. dazu Heblich; Lameli & Riener 2015; Falck, Lameli & Ruhose 2016). Jürg Fleischer führt außerdem derzeit auf Basis der Wenkerbogen ein Projekt *Morphosyntaktische Auswertung von Wenkersätzen* durch, für das erste Ergebnisse bereits publiziert wurden (vgl. etwa Fleischer 2012).

2.2 Mittelrheinischer Sprachatlas (MRhSA)

Von 1978 bis 1988 wurden Sprachdaten für den von Günter Bellmann initiierten Mittelrheinischen Sprachatlas (Bellmann, Herrgen & Schmidt 1994–2002) erhoben.⁵ Das Untersuchungsgebiet umfasst das linksrheinische Rheinland-Pfalz und das Saarland, dialektgeographisch also Teile des Rhein- und des Moselfränkischen. Bei dem Atlas handelt es sich um den ersten bidimensionalen Dialektatlas für das Deutsche. Es wurden Sprachaufnahmen mit älteren Landwirten (um 75 Jahre) – als Datenserie 1 – und mit Handwerkern der mittleren

⁵ Vgl. zum MRhSA zusammenfassend auch Girth (2015).

Generation (um 35 Jahre) – als Datenserie 2 – erhoben, die für die Ausübung ihres Berufes täglich von ihrem Wohnort in eine andere Ortschaft im Nahverkehrsbereich pendeln (vgl. Bellmann 1994: 40). Bellmann formuliert die Ziele des MRhSA folgendermaßen:

Er will einmal für ein Teilgebiet der deutschen Sprachfläche – als Laut- und Formenatlas – in der üblichen Weise die Verteilung der Dialektmerkmale auf der lokal-sprachlandschaftlichen (horizontalen) Ebene dokumentieren. Indem er dies tut – darin besteht sein zweites Ziel –, richtet er seinen Blick auch auf die zwischen Dialekt und Standardsprache sich erstreckende vertikale Dimension, deren Reflexe von den Sprachatlanten des Deutschen sonst, wenn überhaupt, dann eher als Störfaktoren wahrgenommen worden sind. (Bellmann 1994: 1)

Die Sprachdaten wurden in direkter Befragung mit jeweils mehreren Informanten pro Ort (549 Orte für Datenserie 1 und 292 Orte für Datenserie 2) erhoben und dabei sowohl auf Tonbändern aufgezeichnet als auch unmittelbar als IPA-Transkriptionen in Formulare eingetragen. Das Korpus besteht insgesamt aus 1.100 Abfragewörtern für die Datenserie 1 und 440 Abfragewörtern für die Datenserie 2.

Im Unterschied zu vielen anderen Dialektatlanten ist das für den Atlas erhobene Korpus im Rahmen des REDE-Projekts durch orthographische Transkriptionen der Erhebungskontexte und eine Ton-Text-Synchronisierung⁶ (Alignierung) aufbereitet worden und somit zugänglich und weiter analysierbar (siehe auch Abschnitt 5). Insgesamt sind 714 transkribierte und nach Freitext durchsuchbare Sprachaufnahmen verfügbar.

2.3 Bayerische Dialektdatenbank (BayDat)

Den Bayerischen Sprachatlas (BSA) charakterisiert Horst Haider Munske (2015) unter anderem als „(a) ein dialektologisches Großprojekt mit sechs Standorten in Bayern [...], (b) eine Sammelpublikation von über 50 großformatigen Sprachatlas-Bänden [...] und (c) eine einheitliche, systematische, wissenschaftlich fundierte Exploration der Dialekte im Freistaat Bayern, deren Ergebnisse in enger Transkription im Internet (unter BayDat) zugänglich sind“ (Munske 2015: 1). Im vorliegenden Beitrag geht es also vor allem um den zuletzt genannten Aspekt. Die sechs Teilprojekte des Bayerischen Sprachatlas sind:

⁶ Die Synchronisierung von Ton und Transkription stellt für Korpora gesprochener Sprache gegenüber der reinen Transkription für viele Fragestellungen eine erhebliche Steigerung des möglichen Erkenntnisgewinns dar (vgl. auch Hunston 2008: 159–160).

- Sprachatlas von Bayerisch-Schwaben (SBS; bearbeitet in Augsburg),
- Sprachatlas von Unterfranken (SUF; bearbeitet in Würzburg),
- Sprachatlas von Mittelfranken (SMF; bearbeitet in Erlangen),
- Sprachatlas von Niederbayern (SNiB; bearbeitet in Passau),
- Sprachatlas von Nordostbayern (SNOB; bearbeitet in Bayreuth) und
- Sprachatlas von Oberbayern (SOB; bearbeitet in Passau).

Zwischen 1981 und 1998 wurden in allen Teilprojekten in insgesamt 1.613 Erhebungsorten ältere, ortsfeste und dialektkompetente Informanten in einer direkten Befragung aufgenommen. Die Fragebücher umfassten jeweils bis zu 2.800 Fragen. Die Antworten der Informanten wurden direkt in Teuthonista (vgl. Teuchert 1924/1925) phonetisch transkribiert, teilweise erfolgten zusätzlich Tonbandaufnahmen. Die Transkriptionen wurden nachträglich in den Teilprojekten digital erfasst.

Am Würzburger Standort des Bayerischen Sprachatlas erfolgte dann die Integration der digital erfassten Transkriptionen aller Standorte in eine gemeinsame Datenbank, die *Bayerische Dialektdatenbank* (BayDat; vgl. <http://www.baydat.uni-wuerzburg.de:8080/cocoon/baydat/> [letzter Zugriff: 31. 7. 2017]). Diese Datenbank ermöglicht es den Benutzern, über Suchanfragen „eigene Auswertungen vorzunehmen, z. B. Belegorte, sachliche oder sprachliche Themen auszuwählen und eigene Sprachkarten zu entwerfen“ (Munske 2015: 11). Die Ausgabe der Suchanfragen erfolgt jeweils in Form der Teuthonista-Transkription im HTML- oder im PDF-Format. Auch hier ist ein deutlicher Mehrwert gegenüber den rein in Atlasbänden publizierten Dialektatlanten erreicht worden, indem der große Teil der Daten, die nicht auf Karten abgebildet werden konnten, zugänglich und auswertbar ist. Die Daten der BayDat sind auch über die Internetplattform *Regionalsprache.de* in Form von über 4.000 Karten abrufbar. Es handelt sich dabei entsprechend der Datengrundlage um Vollformenkarten, auf denen die Transkription am Ort wiedergegeben ist.

Die Daten, die im Rahmen des Bayerischen Sprachatlas erhoben wurden, sind – wie in allen Sprachatlasprojekten – zunächst einmal für die Kartenerstellung ausgewertet worden. Darüber hinaus wurden mit den Daten des Sprachatlas von Bayerisch-Schwaben im Rahmen des DFG-Projekts *Neue Dialektometrie mit Methoden der stochastischen Bildanalyse* raumstatistische Analysen durchgeführt (vgl. zusammenfassend Pröll et al. 2015).

3 Variationslinguistische Tonkorpora

3.1 Variationslinguistische Tonkorpora am IDS

Das größte Korpus am IDS und zugleich das größte variationslinguistische Tonkorpus für den deutschen Sprachraum ist das *Zwirner-Korpus* (ZW). Eberhard Zwirner verfolgte das Ziel, von allen deutschen Dialekten Tonaufnahmen zu machen und diese wissenschaftlich auszuwerten (vgl. Bethge 1976: 25). Neben der vollständigen Erfassung der deutschen Dialekte war Zwirner ebenso an der sprachlichen Dynamik und weiteren Formen des regionalen Sprechens interessiert, was sich in der Anlage seiner Erhebung äußert (s. Fußnote 8, vgl. auch Schmidt & Herrgen 2011: 119–121). Für das Vorhaben wurden zwischen 1951 und 1972 an ca. 1.000 Orten in den alten Bundesländern, im Elsass, den Niederlanden, Liechtenstein und Vorarlberg Sprachaufnahmen durchgeführt. Das Gebiet wurde in Planquadrate (10 Längen- und 15 Breitenminuten, vgl. Zwirner & Bethge 1958: 15) aufgeteilt und in jedem der Quadrate wurde ein möglichst vom Verkehr abgeschiedenes Dorf bzw. in manchen Fällen auch eine kleinere Stadt ausgewählt (vgl. Zwirner & Bethge 1958: 16).⁷ In jedem Untersuchungsort wurden sechs Sprecher aufgezeichnet: drei ortsfeste einheimische Sprecher und drei Heimatvertriebene. Bei den autochthonen Sprechern wurden jeweils ein Sprecher im Alter von über 60 Jahren aus der „ländlichen Bevölkerung“ (Zwirner & Bethge 1958: 17), ein Sprecher der mittleren Generation und ein Sprecher im Alter um die 20 Jahre ausgewählt.⁸ Bei den Heimatvertriebenen wurden die Sprecher der am häufigsten vertretenen landsmannschaftlichen Gruppe berücksichtigt (vgl. dazu Zwirner & Bethge 1958: 16–17). Durchgeführt wurden die Aufnahmen durch Germanisten der Landesuniversitäten, Wörterbuchleiter und Volkskundler (vgl. zur genauen Übersicht Zwirner & Bethge 1958: 13–14), die von technischen Mitarbeitern unterstützt wurden. Aufgenommen wurden Erzählungen der Gewährspersonen, teilweise auch Dialoge, die Abfrage der Wochentage und Zahlen (Eins bis Zehn) im Dialekt und teilweise auch die Abfrage der Wenkersätze im Dialekt (vgl. Zwirner & Bethge 1958: 18–19). Bei den Aufnahmen zielte Zwirner auf eine möglichst natürliche Gesprächssituation

⁷ Um Herford verdichten sich die Aufnahmeorte, da auf Wunsch des Westfälischen Heimatbunds im damaligen Landkreis Herford alle Orte aufgenommen wurden (insges. 180). Für eine weitere Verdichtung von Tonaufnahmen in Anlehnung an die Zwirner-Erhebung sorgte Arno Ruoff mit Aufnahmen im Südwesten des deutschen Sprachraums (vgl. Bethge 1976: 27, sowie Tab. 6.1).

⁸ Durch diese Auswahl versprach sich Zwirner, „sprachliche Veränderungen, insbesondere den Übergang zum Hochdeutschen [...] zu erfassen“ (Zwirner & Bethge 1958: 17).

und die Verwendung der Sprache, „in der der Sprecher sich in seinem Haus oder an seinem Arbeitsplatz zu unterhalten pflegt“ (Zwirner & Bethge 1958: 19). Mitunter wurden die Sprecher, soweit es den Aufnahmeleitern möglich war, auch im jeweiligen Dialekt befragt (vgl. Zwirner & Bethge 1958: 15). Nach den Aufzeichnungen wurde ein Aufnahmeprotokoll mit zahlreichen Metadaten (u. a. auch eine Einordnung der Sprachschicht und der Sprechsituation) angefertigt (vgl. Zwirner & Bethge 1958: 21–24). Auf diese Weise entstanden insgesamt 5.796 Sprachaufnahmen im Umfang von 1.077 Stunden. Durch die Gegensätzlichkeit der „objektiven sprachgebrauchssteuernden Faktoren und [des] geforderte[n] Sprachverhalten[s]“ (Schmidt & Herrgen 2011: 121) lässt sich davon ausgehen, dass die Sprecher bei den Aufnahmen variiert haben bzw. variieren mussten. Dies hat wiederum zur Folge, dass das Zwirner-Korpus „für die regionalsprachliche Sprachdynamikforschung [eine] sehr ergiebigen Datenreihe“ (Schmidt & Herrgen 2011: 121) darstellt.⁹

Auch für das Gebiet der ehemaligen DDR liegt ein umfassendes variationslinguistisches Tonkorpus vor: „Deutsche Mundarten DDR“ (DR, oder kurz: DDR-Korpus). Der Begründer, Hans-Joachim Schädlich, verfolgte in Zusammenarbeit mit dem Institut für deutsche Sprache und Literatur der Akademie der Wissenschaft der DDR das Ziel der systematischen Erfassung der Dialekte sowie der „landschaftlich bestimmten Formen der Umgangssprache“ (Schädlich & Eras 1964: 376) in der damaligen DDR. Für dieses Vorhaben wurde das Gebiet in Planquadrate (10 Längen- und 15 Breitenminuten, vgl. Schädlich & Eras 1964: 377) aufgeteilt und in jedem dieser Planquadrate ein Ortspunkt ausgewählt (insgesamt 437 Ortspunkte, vgl. Schädlich & Eras 1965: 24). Bei der Auswahl, die den Aufnahmeleitern (Mitarbeiter der Wörterbücher, vgl. Schädlich & Eras 1964: 376–377) oblag, wurden Orte berücksichtigt, „bei denen mit günstigen sprachlichen Verhältnissen gerechnet werden konnte“ (Schädlich & Eras 1964: 377). Aufgezeichnet wurden pro Ort drei ortsfeste Sprecher der bodenständigen Bevölkerung aus drei Generationen (vgl. Schädlich & Eras 1964: 377). Dabei wurden verschiedene Situationen erhoben. Zum einen mussten die Informanten Vergleichstexte in den Dialekt übertragen. Diese bestanden aus den jeweiligen Wortlisten und festen Texten der Mundartwörterbücher sowie einer Variation der Wenkersätze. Die Texte wurden den Sprechern zur Vorbereitung vorab zugeschickt, teilweise wurde auch eine schriftliche Erhebung durchgeführt (vgl. dazu Schädlich & Eras 1964: 377). Zusätzlich wurden freie – aber gelenkte – Gespräche und Erzählungen in einer intendierten freien und ungezwungenen Gesprächsatmosphäre (vgl. Schädlich & Eras 1964: 376–377) aufgenommen. Nach den Aufnahmen wurde durch den Leiter ein ausführliches Auf-

⁹ Die Archivierung und Zugänglichkeit der Korpora wird in Abschnitt 5 beschrieben.

nahmeprotokoll angefertigt (vgl. Schädlich & Eras 1964: 381). Von 1960 bis 1964 wurden insgesamt 1.642 Sprachaufnahmen im Umfang von ca. 385 Stunden durchgeführt. Durch die Anlage der Erhebung kann auch der „Frage nach den sprachlichen Verhältnissen zwischen den Generationen einerseits und dem Verhältnis der Mundart und der Umgangssprache andererseits“ (Schädlich & Eras 1964: 382) nachgegangen werden.¹⁰ Mit dem DDR-Korpus liegt somit für das Gebiet der ehemaligen DDR ein mit dem Zwirner-Korpus vergleichbares Korpus vor.

Mit den *Tonaufnahmen der Vertriebenenmundarten* (TAVM-Korpus, auch *Deutsche Mundarten: ehemalige Ostgebiete* [OS] genannt) existiert ein weiteres umfangreiches Variationskorpus für den deutschen Sprachraum. Zwar lagen mit der Erhebung Zwirners und der Ostdeutschen Dialektgeographie (siehe Abschnitt 3.2) schon erste Aufnahmen und Vorarbeiten zu den Dialekten der Vertriebenen vor, doch sah Bellmann (1964: 64) damals die „eingehende Erfassung der sterbenden ostdeutschen Dialekte [...] in zwölfter Stunde“ geboten. Konkretes Ziel des Vorhabens war es, die bedrohten standardfernsten Sprechlagen (d. h. die Dialekte bzw. die „Umgangssprache der untersten Stufe“ [Bellmann 1964: 66]) der Sprecher aus den ehemaligen deutschen Ostgebieten systematisch zu erfassen. Somit lässt sich von einer „archaisierenden Gesamtzielsetzung“ (Bellmann 1970: 10) sprechen. Durchgeführt wurde das Vorhaben in Zusammenarbeit verschiedener Institutionen (u. a. Deutscher Sprachatlas, Deutsches Spracharchiv, vgl. Bellmann 1964: 64). Dazu wurden zwischen 1962 und 1965 an den damaligen Wohnorten (ca. 900 Orte) insgesamt 981 Sprachaufnahmen im Gesamtumfang von 462 Stunden durchgeführt. Es wurden Sprecher ausgewählt, die über 60 Jahre alt waren, aus bäuerlichen oder ähnlichen Schichten kamen und vor ihrer Umsiedlung ortsfest waren (vgl. dazu und zur allgemein schwierigen Sprecherauswahl und -suche Bellmann 1964: 67–70). Mit ihnen wurden folgende Aufnahmen durchgeführt: Abfrage der leicht modifizierten Wenkersätze (vgl. Bellmann 1970: 24–27), der Wochentage und Zahlen (Eins–Fünfzehn) im Dialekt und ein freier Text, der in den meisten Fällen einem Monolog mit thematischer Vorgabe entspricht (vgl. Bellmann 1964: 73–76). Schmidt & Herrgen (2011: 123) weisen dem TAVM-Korpus einerseits einen „extrem hohen dokumentarischen Wert [...] für die Lauterscheinungen der untergegangenen Varietäten“ zu. Andererseits eignet sich das Korpus auch für verschiedene sprachdynamische Analysen (vgl. Schmidt & Herrgen 2011: 123).¹¹

10 Im Arbeitsgebiet des Thüringischen Wörterbuchs wurde zusätzlich explizit die „geläufige Form der Umgangssprache aufgezeichnet“ (Schädlich & Eras 1964: 377).

11 Bellmann (1964: 70–72) selbst beschreibt sprachdynamische Prozesse zur Zeit der Aufnahmen. So wurden bspw. stigmatisierte Sprechweisen vermieden oder so hat bspw. für die Oberschlesier das Sorbische die Funktion des Dialekts übernommen.

Mit den drei variationslinguistischen Tonkorpora (Zwirner, DDR, TAVM) ist für den Zeitraum zwischen 1955 und 1965 für den deutschen Sprachraum „in ausgezeichneter Qualität die Dialektperformanz bzw. die stark regional geprägte Alltagssprache dreier Altersgruppen mit bis zu sechs Sprechern aus verschiedenen sozialen Schichten dokumentiert“ (Schmidt & Herrgen 2011: 123). Dies gilt weitgehend auch für die Dialektkompetenz.

Zusätzlich liegt für den Zeitraum mit dem *Pfeffer-Korpus* auch ein Tonkorpus der Umgangssprache vor. Alan Pfeffer vom Institut für Grunddeutsch der Universität Pittsburgh setzte sich zum Ziel, die gesprochene Umgangssprache der deutschsprachigen Länder zu Anfang der 1960er Jahre zu dokumentieren (vgl. Pfeffer & Lohnes 1984: 9; oft ist auch die Rede von gesprochener Alltagssprache, städtischer oder gehobener Umgangssprache, vgl. Pfeffer 1975: 28; zum Begriff Grunddeutsch, vgl. Pfeffer 1975: 5–10). In Zusammenarbeit mit verschiedenen Institutionen in den entsprechenden Ländern (BRD, DDR, Österreich, Schweiz, vgl. Pfeffer & Lohnes 1984: 14) entstanden 1961 insgesamt 401 Sprachaufnahmen in einem Umfang von ca. 79 Stunden. Die Aufzeichnungen wurden in 56 Städten durchgeführt, wobei sich die Anzahl der Aufnahmen pro Stadt nach der Bevölkerungsstatistik richtete (vgl. Pfeffer & Lohnes 1984: 13–14). Erhoben wurden 185 Sprecherinnen und 218 Sprecher (insgesamt 403 Sprecher), deren Auswahl nach Bildung, Beruf, Alter und Größe des Wohnorts variiert wurde (vgl. Pfeffer & Lohnes 1984: 17; Pfeffer 1975: 28–29). Von den Sprechern wurden hauptsächlich ungezwungene dialogische Gespräche, teilweise auch Monologe mit Einleitung durch den Aufnahmeleiter und Gruppengespräche aufgezeichnet (vgl. Pfeffer 1975: 11, 29). Dafür wurden 25 Themen systematisch über die Aufnahmen verteilt (vgl. dazu Pfeffer & Lohnes 1984: 17–18). Aufgrund des Erkenntnisinteresses „gesprochene Alltagssprache“ wurden zu stark dialektale Sprachproben aus dem Korpus ausgeschlossen (vgl. Pfeffer 1975: 11). Insgesamt zeichnet sich das Korpus gemäß dem Ziel bzw. der Anlage der Erhebung durch eine relativ große sprachliche Variation aus (vgl. Spiekermann 2008: 107), wird aber meist als standardnäheres Korpus bewertet (vgl. Lenz 2007: 179).

Die beschriebenen variationslinguistischen Tonkorpora sind von großem Nutzen für die Validierung von Ergebnissen und Daten (vgl. Schmidt & Herrgen 2011: 115, 123), wurden bisher aber selten umfassend systematisch ausgewertet. Zu nennen sind bei den Auswertungen der Korpora die Phonai-Monographien zum Zwirner-Korpus (vgl. <http://pub.ids-mannheim.de/abgeschlossen/phonai/> [letzter Zugriff: 31.7. 2017]), mit denen für 30 Ortspunkte phonetische Transkriptionen und phonologische Analysen vorliegen. Für den ersten Band des *Schlesischen Sprachatlases* (Schmitt 1965–1967) hat Bellmann die entsprechenden Tonaufnahmen der Vertriebenenmundarten ausgewertet. Der *Datenbank regionaler Umgangssprachen des Deutschen* (DRUGS) liegen u. a. Tonauf-

Tab. 6.1: Weitere variationslinguistische Tonkorpora in der DGD2.

Korpus	Böblingen-Korpus (BB)	Südwestdeutschland-Vorarlberg-Korpus (SV)*	Schwarzwald-Korpus (SW)*	König-Korpus (KN)
Sprechlage	v. a. intendierter Dialekt	intendierter Dialekt	v. a. intendierter Dialekt	Vorlese-aussprache
Sprecher	Sprecher verschiedenen Alters	Gewährspersonen des <i>Vorarlberger Sprach-atlas</i> (VALTS; vgl. Gabriel 1985)	Sprecher ab 5 Jahren	43 Akademiker (17–27 Jahre)
Region	Kreis Böblingen	Südwestdeutschland, Liechtenstein, Vorarlberg	drei Weiler im Schwarzwald	43 Ortspunkte in den alten Bundesländern ¹²
Aufnahme-datum	1965–1967	1966–1970	1964/1974	1975–1976
Aufnahme-situation	Erzählungen, Gespräche & Vorlese-aussprache	Erzählungen, Wochentage, Zahlen	Erzählungen, Wochentage, Zahlen & Vorlese-aussprache	Vorlese-aussprache (Ausschnitt des Grundgesetzes)
Verfügbar-keit	73 Aufnahmen (insges. 42,5h) in DGD2	242 Aufnahmen (insges. 72h) in DGD2	126 Aufnahmen (insges. 37,5h) in DGD2	43 Aufnahmen (insges. 42,5h) und Transkripte in DGD2

* angelehnt an die Zwirner-Erhebung

nahmen des Pfeffer-Korpus zugrunde (vgl. Lauf 1994: 6). Als raumbezogene Korpusanalyse ist Spiekermann (2008) zu nennen, der in seiner Untersuchung zur Sprache in Baden-Württemberg die entsprechenden Tonaufnahmen des Pfeffer-Korpus ausgewertet hat (vgl. Spiekermann 2008: 105–107). Das DDR-Korpus wurde u. a. für den *Sprachatlas für Rügen und die vorpommersche Küste* (Herrmann-Winter 2013) analysiert. Zudem liegen zahlreiche einzelthematische Untersuchungen vor, die sich auf die genannten Korpora stützen. Als Beispiele können Rowley (1997), Fleischer (2002), Freywald (2010), Lenz (2013), Werth (2014) oder Kasper & Werth (2015) genannt werden, die vorwiegend syntaktische Auswertungen mit den Daten des Zwirner- und/oder des Pfeffer-Korpus vornehmen.

¹² Der *Atlas zur Aussprache des Schriftdeutschen in der Bundesrepublik Deutschland* (König 1989) basiert auf Aufnahmen mit Sprechern aus 44 Orten, von denen 43 Aufnahmen im Korpus enthalten sind.

3.2 Weitere variationslinguistische Korpora

Für den deutschen Sprachraum liegen weitere variationslinguistische Tonkorpora vor, die hier in Auswahl gesondert behandelt werden, da sie nicht Teil des IDS/der DGD2 sind. Hierzu zählt die „Ostdeutsche Dialektgeographie“. Walther Mitzka und Ludwig Erich Schmitt verfolgten das Ziel der dialektgeographischen Erfassung der ehemaligen deutschen Ostgebiete (siehe Abschnitt 3.1). Im Rahmen des Vorhabens wurden von 1954 bis 1958 mit Sprechern aus diesen Gebieten an ihren damaligen Wohnorten Tonaufnahmen durchgeführt. Aufnahmeleiter waren dabei Mitarbeiter des Deutschen Sprachatlas, der Wörterbuchprojekte, der Landsmannschaften, Lehrer und interessierte Laien (vgl. Göschel 1977: 15). Aufgezeichnet wurden die Dialektabfrage der Wenkersätze, der Wochentage und Zahlen sowie Erzählungen und teilweise Dialoge (vgl. Göschel 1977: 14). Die Angaben zur Anzahl der Aufnahmen differieren. Es wird angenommen, dass ca. 1.050–1.100 Aufnahmen angefertigt wurden (vgl. Göschel 1977: 15). Davon sind 729 Wenkersatz-Aufnahmen im REDE-System zugänglich (siehe Abschnitt 5). Göschel (1977: 15) merkt zudem an, dass durch die Entstehung des Korpus die „Qualität der Aufnahmen und [die] räumliche Verteilung sehr uneinheitlich“ sind.

In Vorbereitung des Hessischen Dialektzensus wurde 1982 die Mundart-Aktion „Ich sag’s hessisch“ von den hessischen Sparkassen, der Landesbausparkasse und der Hessen-Nassauischen Versicherung in Zusammenarbeit mit dem Hessen-Nassauischen Wörterbuch (unter Leitung von Heinrich Dingeldein) durchgeführt. Daraus entstanden ist das Korpus *Tonaufnahmen der hessischen Mundarten* (TAHM). Die Sprecher sollten eine Kassette in Selbstaufnahme besprechen. Auf der A-Seite sollten die Wenkersätze in den Dialekt übersetzt werden, die B-Seite stand für freie Texte, Lieder und Inszenierungen zur Verfügung. Insgesamt liegen 1.200 Tonaufnahmen hauptsächlich von Schülerinnen und Schülern, aber auch von Sprechern der mittleren und älteren Generation vor. Davon sind 508 Wenkersatz-Aufnahmen im REDE-System zugänglich (siehe Abschnitt 5). Durch die Anlage der Erhebung sind zum Teil nur eingeschränkt Metadaten vorhanden. Zudem ist bei der Aufnahme der Dialekt zwar intendiert, doch ist vorwiegend von Regiolektaufnahmen auszugehen (persönliche Auskunft von H. Dingeldein; vgl. zu diesem Abschnitt auch <https://regionalsprache.de/tonkorpora.aspx#TAHM> [letzter Zugriff: 31. 7. 2017]). In Anlehnung an das TAHM-Korpus wurde von 2003 bis 2004 unter Leitung von Stefan Arend das Projekt *Tonarchiv osthessischer Mundarten* (ToM) durchgeführt, das zum gleichnamigen Korpus führte. Auch hier führten hauptsächlich Sprecherinnen und Sprecher der älteren und mittleren Generation Selbstaufnahmen (Wenkersätze, Erzählungen und Gedichte) in 23 Gemeinden und Ortsteilen des Landkreises Fulda durch. Insgesamt liegen 400 Tonaufnahmen

vor, von denen 179 Wenkersatz-Aufzeichnungen im REDE-System zugänglich sind. Alle Aufnahmen sind auch über das Medienzentrum Fulda (<http://tom.medienzentrum-fulda.de> [letzter Zugriff: 31. 7. 2017]) abrufbar (vgl. hierzu <https://regionalsprache.de/tonkorpora.aspx#TOM> [letzter Zugriff: 31. 7. 2017] sowie Arend 2002/2003).¹³

Im Kontext variationslinguistischer Tonkorpora ist auch das aktuelle Forschungsprojekt *Sprachalltag II: Sprachatlas – Digitalisierung – Nachhaltigkeit* an der Tübinger Arbeitsstelle Sprache in Südwestdeutschland – Arno-Ruoff-Archiv zu erwähnen. Eines der Ziele des Projekts besteht in der Digitalisierung der ca. 2.000 Tonaufnahmen des Arno-Ruoff-Archivs und der dazu erstellten Transkriptionen im Umfang von etwa 20.000 Seiten (vgl. dazu <http://www.wiso.uni-tuebingen.de/faecher/empirische-kulturwissenschaft/ta-sprache/forschung/sprachalltag-ii-sprachatlas-digitalisierung-nachhaltigkeit.html> [letzter Zugriff: 31. 7. 2017]).

4 Moderne variationslinguistische Korpora

Die Korpora, die unter diesem Punkt zusammengefasst werden, zeichnen sich aus durch (a) moderne Erhebungsmethoden und (b) eine umfassende Perspektivierung sprachlicher Variation, die auch eine Gesamtbetrachtung der modernen Regionalsprachen ermöglicht (vgl. Schmidt & Herrgen 2011). Hinsichtlich der Dimensionen sprachlicher Variation werden sowohl die diatopische (horizontale) als auch die vertikale und die diachrone Dimension betrachtet.

Ein modernes variationslinguistisches Korpus (*Deutsch heute*) ist im Rahmen des IDS-Projekts *Variation des gesprochenen Deutsch* entstanden.¹⁴ Ziel des Projekts – in Anschluss an das Pfeffer- und das König-Korpus (siehe Abschnitt 3.1) – war die „Beschreibung der sprachlichen Variation in [formellen] Situationen“ (Kleiner 2015: 491) und letztlich die Beantwortung der Frage, wie die Standardsprache verwendet wird. Dazu wurden von 2006 bis 2009 an 192 Ortspunkten im gesamten deutschsprachigen Raum Aufnahmen durchgeführt (Deutschland 146, Österreich 25, Schweiz 13, Italien 3, Luxemburg 2, Belgien 2, Liechtenstein 1; vgl. Kleiner 2015: 492 und <http://prowiki.ids-mannheim.de/bin/view/AADG/OrtsListe> [letzter Zugriff: 31. 7. 2017]). Aufgezeich-

¹³ Für die beiden Korpora TAHM und ToM gilt, dass die Aufnahmen durch die Entstehung der Korpora, deren Vorteile auf praktischer Seite liegen, in Bezug auf akustische Qualität und sprachliche Realisierung recht heterogen sind (vgl. <https://regionalsprache.de/tonkorpora.aspx> [letzter Zugriff: 31. 7. 2017]).

¹⁴ Vgl. zum *Deutsch heute*-Korpus zusammenfassend auch Kleiner (2015).

net wurden 345 weibliche und 326 männliche Sprecher mit höherer Schulbildung im Alter von 16 bis 20 Jahren (d. h. Oberstufenschüler) sowie als Vergleichsgruppe 81 weibliche und 77 männliche Sprecher im Alter von 50 bis 60 Jahren. Folgende Situationen wurden erhoben: Vorleseausssprache („Nordwind und Sonne“ [Fabel], populärwissenschaftlicher Text, konstruierte Texte), Übersetzung (Englisch > Deutsch), Bildbeschreibung, Wortliste (Standard), sprachbiographisches Interview und Wegbeschreibung (Map Task). Insgesamt liegen 829 Aufnahmen im Umfang von ca. 1.244 Stunden vor. Diese wurden orthographisch transkribiert und (teilweise automatisch) aligniert. Ausschnitte der Aufnahmen wurden des Weiteren phonetisch transkribiert. Im Rahmen des *Forschungs- und Lehrkorpus gesprochenes Deutsch* (FOLK) als Teil der DGD2 sind einige Interviews und Map Tasks des Korpus bereits zugänglich. Mittelfristig ist geplant, das gesamte Korpus in die DGD2 aufzunehmen. Ergebnisse des Projekts werden u. a. online als *Atlas zur Aussprache des deutschen Gebrauchsstandards* (AADG) (Kleiner 2011 ff.) publiziert.

Eine weitere umfassende Analyse der sprachlichen Variation unternimmt das SiN-Projekt (*Sprachvariation in Norddeutschland*),¹⁵ das von 2007 bis 2013 von der Deutschen Forschungsgemeinschaft gefördert wurde. Ziel des Projekts war die „Dokumentation und Interpretation der aktuellen Sprachsituation im Norden Deutschlands“ mit einem „integrative[n], multiperspektivische[n] Ansatz“ (Elementaler et al. 2015: 397). Dazu wurde von 2007 bis 2010 das SiN-Korpus erstellt. An 36 Ortspunkten im niederdeutschen Sprachraum (jeweils zwei Orte in den traditionell abgegrenzten niederdeutschen Dialektgebieten) wurde die Erhebung durchgeführt (vgl. Elementaler et al. 2015: 399–400). Dabei wurden ländlich geprägte Gemeinden und Ortsteile mit 2.000–8.000 Einwohnern gewählt. Aufgenommen wurden Sprecherinnen im Alter von 40 bis 60 Jahren.¹⁶ Pro Erhebungsort wurden vier Sprecherinnen aufgenommen (insgesamt 144 Sprecherinnen), von denen jeweils zwei dialektkompetent sein sollten und zwei kein Niederdeutsch beherrschen sollten. Mit ihnen wurden folgende Aufnahmen durchgeführt: Dialektabfrage der Wenkersätze, (dialektale) Erzählung, informelles Tischgespräch, formelles, leitfadengestütztes Interview, Vorleseausssprache („Nordwind und Sonne“, Zeitungsartikel). Außerdem wurden mit allen Informantinnen ein Salienz-, Situativitäts-, Normativitäts- und Arealitätstest durchgeführt. Pro Sprecherin dauerten die Aufnahmen ca. 3 Stunden, sodass ein Korpus im Umfang von ca. 432 Stunden vorliegt. Die festen Texte sowie Ausschnitte (2.500 Wörter) von Tischgesprächen und

¹⁵ Vgl. zu SiN zusammenfassend auch Elementaler et al. (2015).

¹⁶ Zur komplementären Auswahl in Bezug zum REDE-Projekt, vgl. Elementaler et al. (2015: 399) sowie Ganswindt, Kehrein & Lameli (2015: 431).

Interviews wurden normorthographisch bzw. nach Niederdeutschkonventionen transkribiert und aligniert. Ausschnittweise wurden auch phonetische Transkriptionen angefertigt. Einige der Tischgespräche sind in der DGD2 (FOLK) zugänglich. Das Gesamtkorpus wird am Hamburger Zentrum für Sprachkorpora (HZSK) dauerhaft gesichert und „soll mittelfristig der wissenschaftlichen Forschung zur Verfügung gestellt werden“ (Elmentaler et al. 2015: 411). Die Ergebnisse liegen vor und werden sukzessive publiziert (vgl. Elmentaler et al. 2015: 411–424).

Das Forschungsprojekt *Regionalsprache.de* (REDE) wird seit 2008 am Forschungszentrum Deutscher Sprachatlas der Universität Marburg durchgeführt.¹⁷ Gefördert wird das Projekt, dessen Laufzeit 19 Jahre beträgt, durch die Akademie der Wissenschaften und der Literatur Mainz. Das übergeordnete Ziel des Projekts ist die umfassende und systematische Erforschung der modernen Regionalsprachen des Deutschen. Für dieses übergeordnete Ziel sind zwei konkretere Teilziele formuliert: (1) Aufbau eines forschungszentrierten Informationssystems zu den modernen Regionalsprachen des Deutschen (siehe dazu Abschnitt 5) sowie (2) Ersterhebung und Analyse der variationslinguistischen Struktur und Dynamik der modernen Regionalsprachen des Deutschen. Für das zweite Teilziel wurde unter anderem ein umfangreiches modernes variationslinguistisches Korpus (REDE-Korpus) erstellt. Von 2008 bis 2015 wurden Sprachaufnahmen an 150 Orten (Städte und Gemeinden) im gesamten Gebiet der Bundesrepublik Deutschland durchgeführt (vgl. <https://regionalsprache.de/empirie-stand-der-bearbeitung.aspx> [letzter Zugriff: 31. 7. 2017]). Aufgenommen wurden jeweils mindestens vier ortsfeste Sprecher aus drei Generationen. Die Sprecher der älteren Generation waren über 65 Jahre alt und sind (meist) einer handwerklichen oder landwirtschaftlichen Tätigkeit nachgegangen. In der mittleren Generation wurden 45- bis 55-jährige Polizisten aufgezeichnet und schließlich in der jüngsten Generation 18- bis 23-jährige Abiturienten aus dem jeweiligen Ort (insgesamt über 700 Sprecher, da an manchen Orten aus forschungspraktischen Gründen mehr als die vorgesehenen Repräsentanten einer Generation aufgezeichnet wurden). Die Erhebungssituationen des REDE-Projektes versuchen, verschiedene Grade der Formalität und somit verschiedene Grade der sprecherseitigen Orientierung an der Standardsprache zu evozieren. Folgende Aufnahmen wurden durchgeführt: Abfrage der Wenkersätze im Dialekt und im individuell besten Hochdeutsch, Vorleseausprache („Nordwind und Sonne“), formelles sprachbiographisches Interview sowie informelles Gespräch mit einem vertrauten Gesprächspartner ohne Anwesenheit Dritter

17 Vgl. zum REDE-Projekt zusammenfassend auch Ganswindt, Kehrein & Lameli (2015).

(sog. Freundesgespräch). Die Aufnahmen dauerten jeweils ca. 2,5 Stunden, sodass bei einer Gesamtzahl von 4.026 Aufnahmen ein Korpus im Umfang von ca. 1.633 Stunden vorliegt. Die Aufbereitung der Daten erfolgt sukzessive. Die standardisierten Texte werden vollständig und die freien Situationen auszugsweise normorthographisch und feinphonetisch transkribiert sowie mit dem Tonsignal synchronisiert. Die aufbereiteten Daten und Forschungsergebnisse werden sukzessive im REDE-System (siehe Abschnitt 5) veröffentlicht. Momentan sind 2.014 Aufnahmen und Transkripte zugänglich.

5 Zugänge zu Dialekt- und Variationskorpora

Die vier in Abschnitt 3.1 beschriebenen variationslinguistischen Tonkorpora des IDS (Zwirner, DDR, TAVM, Pfeffer) sind ins *Deutsche Spracharchiv* (DSAv) und später ins *Archiv für gesprochenes Deutsch* (AGD) übergegangen und liegen vollständig digitalisiert in der *Datenbank für gesprochenes Deutsch* (DGD2) vor. Diese ist seit 2012 das Nachfolgesystem der DGD1. Aufgrund der zentralen Bedeutung der Korpora und der Datenbank für die Variationslinguistik sollen der Zugang und die Nutzung kurz umrissen werden (vgl. zum gesamten Abschnitt Schmidt, Dickgießer & Gasch 2013).¹⁸ Die DGD2 ist ein Korpusmanagementsystem, in dem Teilbestände des ADG (Sprachaufnahmen, Transkripte, Metadaten) zur Verfügung gestellt werden (zu den anderen Korpora der DGD2 siehe Abschnitt 3.1, Boas & Fingerhuth in diesem Band sowie Schmidt in diesem Band und Internetseite der DGD). Die Sprachaufnahmen liegen digitalisiert und teilweise tontechnisch bearbeitet im WAVE-Format vor. Die Transkripte sind größtenteils ebenfalls digitalisiert bzw. digital erstellt worden. Durch die unterschiedliche Bearbeitung der Transkripte liegt hier eine große Heterogenität vor, an deren Auflösung stetig gearbeitet wird. Zusätzlich steht eine umfangreiche Korpusdokumentation zur Verfügung, die zahlreiche Metadaten zu Aufnahmen, Sprechern usw. enthält. Nach einmaliger Online-Registrierung kann die DGD2 genutzt werden. Je nach Erkenntnisinteresse bestehen unterschiedliche Zugriffs- und Bearbeitungsformen. Die DGD2 bietet zwei grundlegende Funktionen an: Browsen und Suchen. „Der Menüpunkt ‚Korpora‘ der DGD2 bietet die Möglichkeit, Metadaten, Transkripte und Zusatzmaterial anzusehen sowie Audioaufnahmen ausschnittsweise anzuhören.“ (Schmidt, Dickgießer & Gasch 2013: 13) Dies wird als Browsen verstanden und ermöglicht

¹⁸ Zusätzlich können die Sprachaufnahmen der Korpora auch im persönlichen Service über die DGD2 bestellt werden bzw. bietet die DGD2 auch einen Download-Service an.

dem Nutzer eine erste Orientierung in den Korpora. Im Vergleich zur gezielten Suche entspricht diese Funktion eher einem „Stöbern“ und wird durch verschiedene Formen der Verknüpfungen (Links) zusätzlich verbessert. Beim Browsen lassen sich jeweils 15-sekündige Ausschnitte anhören und, sofern das Transkript vorliegt, auch die Verschriftlichungen mitlesen (vgl. Schmidt, Dickgießer & Gasch 2013: 14–16). Bei der gezielten Suche lassen sich zwei Funktionen unterscheiden. Bei der Volltextsuche wird innerhalb der Transkripte nach dem jeweiligen Begriff gesucht, wobei über Wildcards usw. Spezifizierungen möglich sind. Die struktursensitive Suche wiederum greift nicht nur auf die Transkripte, sondern auch auf die Annotationen zurück und ermöglicht zudem über Metadatenfilter eine Spezifizierung der Fundstellen (vgl. Schmidt, Dickgießer & Gasch 2013: 17–19). Mit der DGD2 – seit April 2017 steht die Version 2.8 mit einigen Verbesserungen und Neuerungen zur Verfügung – liegt somit eine stabile, stets aktualisierte Online-Datenbank vor, die sich nicht nur durch die umfangreichen Bestände, sondern auch durch Zugänglichkeit und Funktionalität auszeichnet und somit für die variationslinguistische Forschung eine wichtige Datenquelle darstellt.

Wie oben erwähnt besteht eines der Teilziele im Projekt REDE im Aufbau einer interaktiven Forschungsplattform zu den modernen Regionalsprachen des Deutschen. Diese Plattform ist in den letzten Jahren auf Basis des Datenbestandes aus dem *Digitalen Wenker-Atlas* aufgebaut worden und sie wird stetig erweitert.¹⁹ Das Herzstück des REDE-Systems bildet das sprachgeographische Informationssystem REDE SprachGIS, über das kartenbasiert auf den gesamten Datenbestand zugegriffen werden kann. Um dies zu ermöglichen, sind alle Daten des Systems orts- oder raumbezogen gespeichert und mit einer eindeutigen geographischen Identifikationsnummer (GID) versehen, die auf geographische Koordinaten verweist. Teil des Datenbestandes bilden auch die im vorliegenden Beitrag beschriebenen oder genannten Sprachatlas- und Tonkorpora.

Die kartierten Phänomene aus den im System enthaltenen Sprachatlas-korpora sind über die Kartensuche aufrufbar. In dieser Kartensuche kann neben der Freitexteingabe auch eine Filterung über ein morphologisches, ein phonetisch-phonologisches und ein historisches Register vorgenommen werden. Karten können dann einzeln oder übereinander in das REDE SprachGIS geladen und so betrachtet und direkt verglichen werden.

¹⁹ Die Forschungsplattform ist frei unter der URL www.regionalsprache.de (letzter Zugriff: 31. 7. 2017) zugänglich.

Alle anderen Daten, also auch die orthographisch transkribierten Teile der Tonkorpora,²⁰ können über das Recherche-Werkzeug des REDE SprachGIS durchsucht werden. Die Suche kann dabei über alle Korpora erfolgen oder sie kann auf einzelne oder mehrere ausgewählte Korpora eingeschränkt werden. Zusätzlich besteht die Möglichkeit, Suchergebnisse über einen Kartenelementefilter auf eine ausgewählte Region (z. B. ein Dialektverband, ein Bundesland oder auch ein vorher frei gezeichnetes Polygon) zu begrenzen. Diese Recherche- und Filterfunktionen stehen nicht nur für die Tonkorpora, sondern für alle im System enthaltenen Daten zur Verfügung, z. B. auch für das Wenkerbogen-Korpus (vgl. Kehrein i. Dr.). Für dieses Korpus kann bisher allerdings nur angezeigt werden, an welchen Orten ein Bogen verfügbar ist. Eine Suche in den Bogen ist (noch) nicht möglich, sie liegen lediglich als Bilddateien vor. Die Suchergebnisse für die Tonkorpora lassen sich im REDE SprachGIS einzeln oder als Zusammenstellung („Playlist“) anhören und auf diese Weise vergleichend analysieren.

Die beiden beschriebenen Suchwerkzeuge (Kartensuche und Recherche) und ihre je spezifischen Filterfunktionen werden derzeit in einem übergreifenden Suchwerkzeug zusammengeführt. Das bedeutet, dass nach erfolgter Umstellung auch bei der Suche in den entsprechend erschlossenen Tonkorpora nach den Kategorien der vorhandenen Register gefiltert werden kann.²¹

6 Ausblick

Die vorhandenen Korpora zur Sprachvariation bzw. zu sprachlichen Varietäten im Deutschen bilden bereits eine solide und umfangreiche Grundlage für wissenschaftliche Studien. Dennoch könnte und sollte diese Basis in den

20 Außer den im vorliegenden Beitrag beschriebenen Korpora sind im REDE-System noch Aufnahmen aus weiteren Korpora enthalten. Da es sich dabei um einen Bestand handelt, der bereits für den *Digitalen Wenker-Atlas* zusammengestellt wurde, sind im REDE-System meist Wenkersatzaufnahmen aus den Korpora integriert (vgl. für eine Übersicht Kehrein i. Dr. sowie <https://regionalsprache.de/tonkorpora.aspx> [letzter Zugriff: 31. 7. 2017]).

21 Aktuell zu nennen ist auch [moca] (*multimodal oral corpora administration*). Hierbei handelt es sich um ein Online-System der Universität Freiburg (Romanisches Seminar), das Tonaufnahmen mit alignierten Transkripten und Metadaten speichert und zur Verfügung stellt. Es werden umfangreiche und detaillierte Suchmöglichkeiten sowie die Funktion der individuellen Bearbeitung der Korpora angeboten. Das System ist ohne Einschränkungen online zugänglich und soll in der dritten Version ab November 2017 zur Verfügung stehen (vgl. <http://www.hpsl.uni-freiburg.de/> [letzter Zugriff: 31. 7. 2017]).

kommenden Jahren sowohl in quantitativer als auch in qualitativer Hinsicht erweitert werden. Quantitativ sind folgende Maßnahmen denkbar:

- Erhöhung der Ortsnetzdichte der modernen variationslinguistischen Korpora durch (regionale) Neuerhebungen mit vergleichbaren Methoden,
- Ausbau vorhandener Datensammlungen, z. B. von Dialektatlanten, zu analysierbaren Korpora – ähnlich wie beim Mittelrheinischen Sprachatlas oder der Bayerischen Dialektdatenbank – durch Transkriptionen, durch Ton-Text-Synchronisierung, durch Erschließung über Register und evtl. durch Datenbankerfassung der Transkriptionen,
- Volltextdigitalisierung von handschriftlichen (Korpus-)Transkriptionen, die bisher nicht oder nur in Bildform digital vorliegen (z. B. das Wenkerbogen-Korpus im REDE-System oder die in Tübingen bearbeiteten Ruoff-Transkriptionen),
- Integration solcher neuen Korpora in vorhandene Datenbanksysteme oder zumindest Vernetzung mit diesen.

In qualitativer Hinsicht könnten vorhandene und neue Dialekt- und Variationskorpora nicht nur durch orthographische und – wie teilweise bereits durchgeführt – phonetische Transkriptionen erschlossen werden. Es könnten vielmehr kontrollierte Annotationen für alle linguistischen Systemebenen vorgenommen werden, was die (technisch bereits möglichen) Analyseoptionen um ein Vielfaches erhöhen würde. Wichmann (2008) bringt dieses Desiderat folgendermaßen auf den Punkt:

One hopes that future spoken corpora will provide linguistically sophisticated syntactic, pragmatic and discourse annotation together with an equally sophisticated prosodic annotation that can then be complemented by automatic analysis of global trends, such as pitch, pause, loudness and voice quality. At present, the technology outstrips the linguistics. (Wichmann 2008: 205)

Literatur

- Arend, Stefan (2002/2003): Reden, wie der Schnabel gewachsen ist. Tonarchiv osthessischer Mundarten (ToM) will Dialekte der Region umfassend und flächendeckend dokumentieren. *Jahrbuch Landkreis Fulda* 30, 124–138.
- Auer, Peter (1990): *Phonologie der Alltagssprache. Eine Untersuchung zur Standard/Dialekt-Variation am Beispiel der Konstanzer Stadtsprache*. Berlin, New York: de Gruyter.
- Bellmann, Günter (1964): Wege und Möglichkeiten der Schallaufnahme ostdeutscher Mundarten heute. Zur Tonbandaufnahme der Vertriebenenmundarten. *Zeitschrift für Mundartforschung* 31, 62–79.

- Bellmann, Günter (1970): Einleitung. In Günther Bellman & Joachim Göschel (Hrsg.), *Tonbandaufnahmen ostdeutscher Mundarten 1962–1965. Gesamtkatalog mit 10 Karten*, 7–29. Marburg: Elwert.
- Bellmann, Günter (1994): *Einführung in den Mittelrheinischen Sprachatlas*. Tübingen: Niemeyer.
- Bellmann, Günter, Joachim Herrgen & Jürgen Erich Schmidt (Hrsg.) (1994–2002): *Mittelrheinischer Sprachatlas*. 5 Bde. Tübingen: Niemeyer.
- Bethge, Wolfgang (1976): Vom Werden und Wirken des Deutschen Spracharchivs. *Zeitschrift für Dialektologie und Linguistik* 43 (1), 22–43.
- Chambers, J. K. & Peter Trudgill (1998): *Dialectology. Second edition*. Cambridge: Cambridge University Press.
- Christen, Helen, Ingrid Hove & Marina Petkova (2015): Gesprochene Standardsprache im Deutschschweizer Alltag. In Roland Kehrein, Alfred Lameli & Stefan Rabanus (Hrsg.), *Regionale Variation des Deutschen: Projekte und Perspektiven*, 379–396. Berlin, Boston: de Gruyter Mouton.
- Christen, Helen, Manuela Guntern, Ingrid Hove & Marina Petkova (2010): *Hochdeutsch in aller Munde. Eine empirische Untersuchung zur gesprochenen Standardsprache in der Deutschschweiz* (Zeitschrift für Dialektologie und Linguistik. Beiheft 140). Stuttgart: Steiner.
- Coseriu, Eugenio (1992): *Einführung in die Allgemeine Sprachwissenschaft*. 2. Auflage. Tübingen: Francke.
- Digitaler Wenker-Atlas (DiWA)* (2001–2009). Hrsg. von Jürgen Erich Schmidt & Joachim Herrgen. Bearbeitet von Alfred Lameli/Alexandra N. Lenz/Jost Nickel & Roland Kehrein/Karl-Heinz Müller/Stefan Rabanus. Marburg: Forschungszentrum Deutscher Sprachatlas. <http://www.diwa.info> (letzter Zugriff: 31. 7. 2017).
- Elementar Michael, Joachim Gessinger, Jens Lanwer, Peter Rosenberg, Ingrid Schröder & Jan Wirrer (2015): Sprachvariation in Norddeutschland (SiN). In Roland Kehrein, Alfred Lameli & Stefan Rabanus (Hrsg.), *Regionale Variation des Deutschen: Projekte und Perspektiven*, 397–424. Berlin, Boston: de Gruyter Mouton.
- Falck, Oliver, Alfred Lameli & Jens Ruhose (2016): Cultural biases in migration: Estimating non-monetary migration costs. *Papers in Regional Science*, doi: 10.1111/pirs.12243.
- Fleischer, Jürg (2002): *Die Syntax von Pronominaladverbien in den Dialekten des Deutschen. Eine Untersuchung zu Preposition Stranding und verwandten Phänomenen*. Stuttgart: Steiner.
- Fleischer, Jürg (2012): Pronominalsyntax im nordwestlichen Niederdeutsch: eine Auswertung des Wenker-Materials (mit Einbezug der friesischen und dänischen Formulare). *Niederdeutsches Jahrbuch* 135, 59–80.
- Fleischer, Jürg (2017): *Geschichte, Anlage und Durchführung der Fragebogen-Erhebungen von Georg Wenkers 40 Sätzen. Dokumentation, Entdeckungen und Neubewertungen*. Hildesheim u. a.: Olms.
- Fleischer, Jürg & Thomas Gadmer (Hrsg.) (2002): *Schweizer Aufnahmen. Tondokumente aus dem Phonogrammarchiv der Österreichischen Akademie der Wissenschaften. Gesamtausgabe der Historischen Bestände 1899–1950*. Wien: Österreichische Akademie der Wissenschaften.
- Freywald, Ulrike (2010): Obwohl vielleicht war es ganz anders. Vorüberlegungen zum Alter der Verbzweitstellung nach subordinierenden Konjunktionen. In Arne Ziegler (Hrsg.), *Historische Textgrammatik und Historische Syntax des Deutschen*, 55–84. Berlin, Boston: de Gruyter.

- Friginal, Eric & Jack A. Hardy (2014): *Corpus-based sociolinguistics. A guide for students*. New York, London: Routledge.
- Gabriel, Eugen (Hrsg.) (1985): *Vorarlberger Sprachatlas mit Einschluss des Fürstentums Liechtenstein, Westtirols und des Allgäus (VALTS). Einführung in den Vorarlberger Sprachatlas*. Bregenz: Vorarlberger Landesregierung.
- Ganswindt, Brigitte, Roland Kehrein & Alfred Lameli (2015): Regionalsprache.de (REDE). In Roland Kehrein, Alfred Lameli & Stefan Rabanus (Hrsg.), *Regionale Variation des Deutschen: Projekte und Perspektiven*, 425–453. Berlin, Boston: de Gruyter Mouton.
- Girnth, Heiko (2015): Der Mittelrheinische Sprachatlas (MRhSA). In Roland Kehrein, Alfred Lameli & Stefan Rabanus (Hrsg.), *Regionale Variation des Deutschen: Projekte und Perspektiven*, 29–51. Berlin, Boston: de Gruyter Mouton.
- Göschel, Joachim (Hrsg.) (1977): *Die Schallaufnahme deutscher Dialekte im Forschungsinstitut für deutsche Sprache*. Marburg.
- Heblich, Stephan, Alfred Lameli & Gerhard Riener (2015): The impact of regional accents on economic behavior: A lab experiment on linguistic performance, cognitive ratings and economic decisions. *PLoS ONE* 10/2, doi: 10.1371/journal.pone.0113475.
- Herrmann-Winter, Renate (2013): *Sprachatlas für Rügen und die vorpommersche Küste*. Kartographie Martin Hansen. Rostock: Hinstorff.
- Hunston, Susan (2008): Collection strategies and design decisions. In Anke Lüdeling & Merja Kytö (Hrsg.), *Corpus linguistics. An international handbook*. Vol. 1 (Handbücher zur Sprach- und Kommunikationswissenschaft 29.1), 154–168. Berlin, Boston: de Gruyter Mouton.
- Jaberg, Karl & Jakob Jud (1928): *Der Sprachatlas als Forschungsinstrument. Kritische Grundlegung und Einführung in den Sprach- und Sachatlas Italiens und der Südschweiz*. Halle (Saale): Niemeyer.
- Kasper, Simon & Alexander Werth (2015): Fundierung linguistischer Basiskategorien (LingBas). In Roland Kehrein, Alfred Lameli & Stefan Rabanus (Hrsg.), *Regionale Variation des Deutschen: Projekte und Perspektiven*, 349–377. Berlin, Boston: de Gruyter Mouton.
- Kehrein, Roland (2006): Regional accent in the German language area – How dialectally do German police answer emergency calls? In Frans Hinskens (Hrsg.), *Language variation. European perspectives*, 83–96. Amsterdam, Philadelphia: Benjamins.
- Kehrein, Roland (2012): *Regionalsprachliche Spektren im Raum. Zur linguistischen Struktur der Vertikale*. Stuttgart: Steiner.
- Kehrein, Roland (i. Dr.): Das Wenker-Material im REDE SprachGIS. In Jürg Fleischer et al. (Hrsg.), *Minderheitensprachen und Sprachminderheiten: Deutsch und seine Kontaktsprachen in der Dokumentation der Wenker-Materialien*. Hildesheim u. a.: Olms.
- Kleiner, Stefan (2011 ff.): *Atlas zur Aussprache des deutschen Gebrauchsstandards (AADG)*. Unter Mitarbeit von Ralf Knöbl. <http://prowiki.ids-mannheim.de/bin/view/AADG/> (letzter Zugriff: 31. 7. 2017).
- Kleiner, Stefan (2015): „Deutsch heute“ und der Atlas zur Aussprache des deutschen Gebrauchsstandards. In Roland Kehrein, Alfred Lameli & Stefan Rabanus (Hrsg.), *Regionale Variation des Deutschen: Projekte und Perspektiven*, 489–518. Berlin, Boston: de Gruyter Mouton.
- König, Werner (1989): *Atlas zur Aussprache des Schriftdeutschen in der Bundesrepublik Deutschland*. 2 Bde. Ismaning: Hueber.
- König, Werner (Hrsg.) (1996–2009): *Sprachatlas vom Bayerisch-Schwaben* (Bayerischer Sprachatlas: Regionalteil 1). 14 Bde. Heidelberg: C. Winter.

- Kunst, Jan Pieter & Sjeff Barbiers (2010): Generating maps on the internet. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (Hrsg.), *Language and space. An international handbook of linguistic variation. Volume 2: Language mapping* (Handbücher zur Sprach- und Kommunikationswissenschaft 30.2), 401–415. Berlin, Boston: de Gruyter Mouton.
- Lameli, Alfred (2004): *Standard und Substandard. Regionalismen im diachronen Längsschnitt*. Stuttgart: Steiner.
- Lameli, Alfred (2009): Kommentar zum phonologischen Register der standardsprachlichen Vergleichslaute. <http://www.regionalsprache.de/phonologisches-register.aspx> (letzter Zugriff: 31. 7. 2017).
- Lameli, Alfred (2013): *Strukturen im Sprachraum. Analysen zur arealtypologischen Komplexität der Dialekte in Deutschland*. Berlin, Boston: de Gruyter.
- Lauf, Raphaela (1994): *Datenbank regionaler Umgangssprachen des Deutschen. DRUGS. Abschlussbericht*. Manuskript. Universität Marburg.
- Lehmann, Christian (2007): Daten – Korpora – Dokumentation. In Werner Kallmeyer & Gisela Zifonun (Hrsg.), *Sprachkorpora. Datenmengen und Erkenntnisfortschritt*, 9–27. Berlin, Boston: Walter de Gruyter.
- Lenz, Alexandra N. (2003): *Struktur und Dynamik des Substandards. Eine Studie zum Westmitteldeutschen (Wittlich/Eifel)*. Stuttgart: Steiner.
- Lenz, Alexandra N. (2007): Zur variationslinguistischen Analyse regionalsprachlicher Korpora. In Werner Kallmeyer & Gisela Zifonun (Hrsg.), *Sprachkorpora. Datenmengen und Erkenntnisfortschritt*, 187–202. Berlin, Boston: de Gruyter.
- Lenz, Alexandra N. (2013): *Vom kriegen und bekommen. Kognitiv-semantische, variationslinguistische und sprachgeschichtliche Perspektiven*. Berlin, Boston: de Gruyter.
- Munske, Horst Haider (2015): Der Bayerische Sprachatlas (BSA). In Roland Kehrein, Alfred Lameli & Stefan Rabanus (Hrsg.), *Regionale Variation des Deutschen: Projekte und Perspektiven*, 1–27. Berlin, Boston: de Gruyter Mouton.
- Pfeffer, J. Alan (1975): *Grunddeutsch: Erarbeitung und Wertung dreier deutscher Korpora; ein Bericht aus dem Institute for Basic German, Pittsburgh*. Tübingen: Narr.
- Pfeffer, J. Alan & Walter F. W. Lohnes (Hrsg.) (1984): *Grunddeutsch. Texte zur gesprochenen deutschen Gegenwartssprache*. Tübingen: Niemeyer.
- Pröll, Simon, Simon Pickl, Aaron Spetttl, Volker Schmidt, Evgeny Spodarev, Stephan Elspaß & Werner König (2015): Neue Dialektometrie mit Methoden der stochastischen Bildanalyse. In Roland Kehrein, Alfred Lameli & Stefan Rabanus (Hrsg.), *Regionale Variation des Deutschen: Projekte und Perspektiven*, 173–194. Berlin, Boston: de Gruyter Mouton.
- Rabanus, Stefan (2005): Kommentar zum morphologischen Register. <http://www.regionalsprache.de/morphologisches-register.aspx> (31. 7. 2017).
- Regionalsprache.de (REDE)* (2008 ff.). Hrsg. von Jürgen Erich Schmidt, Joachim Herrgen & Roland Kehrein. Forschungsplattform zu den modernen Regionalsprachen des Deutschen. Bearbeitet von Dennis Bock, Brigitte Ganswindt, Heiko Girnth, Simon Kasper, Roland Kehrein, Alfred Lameli, Slawomir Messner, Christoph Purschke, Anna Wolańska. Marburg: Forschungszentrum Deutscher Sprachatlas.
- Rocholl, Josephine (2015): *Ostmitteldeutsch – eine moderne Regionalsprache? Eine Untersuchung zu Konstanz und Wandel im thüringisch-obersächsischen Sprachraum*. Hildesheim u. a.: Olms.
- Rowley, Anthony Robert (1997): *Morphologische Systeme der nordostbayerischen Mundarten in ihrer sprachgeographischen Verflechtung*. Stuttgart: Steiner.

- Schädlich, Hans-Joachim & Heinrich Eras (1964): Deutsche Dialektologie und Tonaufnahmetechnik. *Spektrum. Mitteilungsblatt für die Mitarbeiter der Deutschen Akademie der Wissenschaften zu Berlin* 10, 375–382.
- Schädlich, Hans-Joachim & Heinrich Eras (1965): Bericht über die Tonbandaufnahmen der deutschen Mundarten in der Deutschen Demokratischen Republik. In *Berichte über dialektologische Forschungen in der Deutschen Demokratischen Republik*, 24–27. Berlin.
- Schmidt, Jürgen Erich & Joachim Herrgen (2011): *Sprachdynamik. Eine Einführung in die moderne Regionalsprachenforschung*. Berlin: Erich Schmidt.
- Schmidt, Thomas, Sylvia Dickgießer & Joachim Gasch (2013): *Die Datenbank für Gesprochenes Deutsch – DGD2*. Mannheim: Institut für Deutsche Sprache.
- Schmitt, Ludwig Erich (Hrsg.) (1965–1967): *Schlesischer Sprachatlas*. Band 1: Laut- und Formenatlas. Von Günter Bellmann. Unter Mitarbeit von Wolfgang Putschke und Werner Veith. Band 2: Wortatlas. Von Günter Bellmann. Marburg: Elwert.
- Spiekermann, Helmut (2008): *Sprache in Baden-Württemberg: Merkmale des regionalen Standards*. Tübingen: Niemeyer.
- Teuchert, Hermann (1924/1925): Lautschrift des Teuthonista. In *Teuthonista* 1, 5.
- Viëtor, Wilhelm (1888): Beiträge zur Statistik der Aussprache des Schriftdeutschen. *Phonetische Studien. Zeitschrift für wissenschaftliche und praktische Phonetik* I, 95–114; 209–226.
- Vorberger, Lars (2017): *Regionalsprache in Hessen. Eine Untersuchung zu Sprachvariation und Sprachwandel im mittleren und südlichen Hessen*. Dissertationsschrift Universität Marburg.
- Wecker-Kleiner, Bernadette (2009): *Sprechen nach der Schrift. Die Vorleseaussprache von DialektsprecherInnen in Bayerisch-Schwaben im Spannungsfeld zwischen Dialekt und Orthoepie*. Berlin: dissertation.de.
- Wenker, Georg (1881): *Sprach-Atlas von Nord- und Mitteldeutschland. Auf Grund von systematisch mit Hilfe der Volksschullehrer gesammeltem Material aus circa 30 000 Orten. Bearbeitet, entworfen und gezeichnet von Dr. G. Wenker. Text. Einleitung*. Straßburg, London: Trübner.
- Wenker, Georg (1886): [Über das Sprachatlasunternehmen]. In *Verhandlungen der achtundreißigsten Versammlung deutscher Philologen und Schulmänner in Gießen vom 30. September bis 3. Oktober 1885*, 187–194. Leipzig: Teubner.
- Wenker, Georg (1889–1923): *Sprachatlas des Deutschen Reichs*. Handgezeichnet von Emil Maurmann, Georg Wenker & Ferdinand Wrede. Forschungszentrum Deutscher Sprachatlas. Marburg. [Publiziert als *Digitaler Wenker-Atlas (DiWA)*].
- Wenker, Georg (2013): *Schriften zum Sprachatlas des Deutschen Reichs. Gesamtausgabe. Band 1: Handschriften: Allgemeine Texte, Kartenkommentare 1889–1897*. Hrsg. und bearb. von Alfred Lameli. Unter Mitarbeit von Johanna Heil und Constanze Wellendorf. Hildesheim u. a.: Olms.
- Werth, Alexander (2014): Die Funktionen des Artikels bei Personennamen im norddeutschen Sprachraum. In Friedhelm Debus, Rita Heuser & Damaris Nübling (Hrsg.), *Linguistik der Familiennamen*, 139–174. Hildesheim u. a.: Olms.
- Wichmann, Anne (2008): Speech corpora and spoken corpora. In Anke Lüdeling & Merja Kytö (Hrsg.), *Corpus linguistics. An international handbook*. Vol. 1 (Handbücher zur Sprach- und Kommunikationswissenschaft 29.1), 187–207. Berlin, Boston: de Gruyter Mouton.
- Wiesinger, Peter (1970): *Phonetisch-phonologische Untersuchungen zur Vokalentwicklung in den deutschen Dialekten*. 2 Bände (Studia Linguistica Germanica 2/1+2). Berlin, New York: de Gruyter.

- Wiesinger, Peter (1983): Die Einteilung der deutschen Dialekte. In Werner Besch et al. (Hrsg.), *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. 2. Teilband (Handbücher zur Sprach- und Kommunikationswissenschaft 1.2), 807–900. Berlin, New York: de Gruyter.
- Wrede, Ferdinand (1937): Ferdinand Wredes Einteilungskarte der deutschen Mundarten. In Walther Mitzka & Bernhard Martin (Hrsg.), *Deutscher Sprachatlas. 9. Lieferung. Karte 56*. Marburg: Elwert.
- Zwirner, Eberhard & Wolfgang Bethge (1958): *Erläuterung zu den Texten* (Lautbibliothek der deutschen Mundarten 1). Göttingen: Vandenhoeck & Ruprecht.