

[erscheint als: Carstensen, Kai-Uwe/Spiekermann, Helmut/Tophinke, Doris/Vogel, Petra M./Wich-Reif, Claudia (2020): Zur Methodik des Dialektatlas Mittleres Westdeutschland (DMW). In: *Korrespondenzblatt des Vereins für niederdeutsche Sprachforschung* 127. S. 107-114.]

Zur Methodik des Dialektatlas Mittleres Westdeutschland (DMW)

Kai-Uwe Carstensen, Helmut Spiekermann, Doris Tophinke, Petra M. Vogel, Claudia Wich-Reif

1. Einführung und Kurzüberblick

Der *Dialektatlas Mittleres Westdeutschland* (DMW) ist eine von der Nordrhein-Westfälischen Akademie der Wissenschaften und der Künste bzw. von der Union der deutschen Akademien der Wissenschaften geförderte Atlasdokumentation (Laufzeit: 17 Jahre, 2016-2032).¹ Die Erhebungsorte sind Orte, für die bereits Georg Wenker (1852-1911) Sprachdaten erhoben hat, so dass ein Langzeitvergleich der Sprachdaten möglich ist.² Methodisch schließt der DMW an die bereits abgeschlossenen Atlasprojekte im deutschsprachigen Raum an.³ Die direkten Erhebungen von männlichen und weiblichen Befragten, sog. Gewährspersonen (im Folgenden GPs), basieren auf einem Fragebuch.⁴ Ergänzend zu den fertiggestellten, Großräume erfassenden Sprachatlanten, etwa zum *Mittelrheinischen Sprachatlas* (MRhSA) oder zum *Sprachatlas von Mittelfranken* (SMF), wird beim DMW außer den Wenkersätzen sowie den Systemebenen Phonologie, Morphologie und Lexik auch die grammatische Ebene der Syntax in Anlehnung an den *Atlas zur Syntax hessischer Dialekte* (SyHD) erhoben. Wie beim MRhSA sind für den DMW zwei altersabhängige Aufnahmeserien vorgesehen, so dass für Auswertungen ein *apparent time*-Vergleich möglich ist, um etwa beurteilen zu können, welche sprachlichen Merkmale im Laufe der Zeit abnehmen oder auch, welche sich in einem vertikal über den Dialekten zu verortenden Regiolekt verstärken. Im Vergleich zu anderen Atlanten ist u. a. auch die Nutzung der *draw-a-map-Methode* aus dem Bereich der Wahrnehmungsdialektologie neu; mithilfe der *mental maps* und deren Kommentierung

¹ Beteiligt sind die Universitäten Bonn (Projektleitung: Claudia Wich-Reif), Münster (Projektleitung: Helmut Spiekermann), Paderborn (Projektleitung: Doris Tophinke) und Siegen (Projektleitung: Petra M. Vogel; Sprecheruni/Koordination). Eine Übersicht der Grunddaten findet sich im Anhang sowie auf der Projekt-Homepage (<https://www.dmw-projekt.de/>).

² Einen kurzen Überblick zu Wenker und seinem Werk bietet <http://www.diwa.info/Geschichte/uebersicht.aspx>.

³ Eine aktuelle Übersicht, in die auch der DMW aufgenommen ist, bietet Schmidt, J. E. u. a. 2019: 31.

⁴ Grundlage ist der *Mittelrheinische Sprachatlas*, der an die Bedarfe des DMW (niederdeutsch – hochdeutsch u. a.) angepasst wurde. Erhoben wird mittels Umschreibungen, Lückentexten, Bildern, Übersetzungen und Videos.

können so auch Sprachraumkonzeptualisierungen analysiert werden. Ein weiteres wesentliches Merkmal des DMW ist, dass in einem einzigen Atlas die standardfernsten Sprachvarietäten im niederdeutschen und im hochdeutschen Raum gemeinsam erhoben werden. Aus dieser Kombination resultieren spezifische Herausforderungen, was die Genese (niederdeutsch oder hochdeutsch) der von den GPs geäußerten Formen, aber auch die Abbildung auf Sprachkarten sowie die Analyse und Bewertung betrifft.⁵

Mit der digitalen und audiovisuellen Dokumentation, für die Sprachdaten erhoben, in einer zentralen Datenbank gespeichert, analysiert und auf Karten präsentiert werden, unterscheidet sich der DMW grundlegend von den bereits bestehenden Großraumatlanten. In der folgenden Darstellung liegt der Schwerpunkt auf der Vorstellung des Workflows, also der Arbeitsabläufe, bei dem unterschiedliche Tools/Werkzeuge eingesetzt werden. Vorbild für diese digitalen, „sprechenden“ Karten mit anhörbaren Belegen ist der *Siegerländer Sprachatlas* (SiSAL).⁶

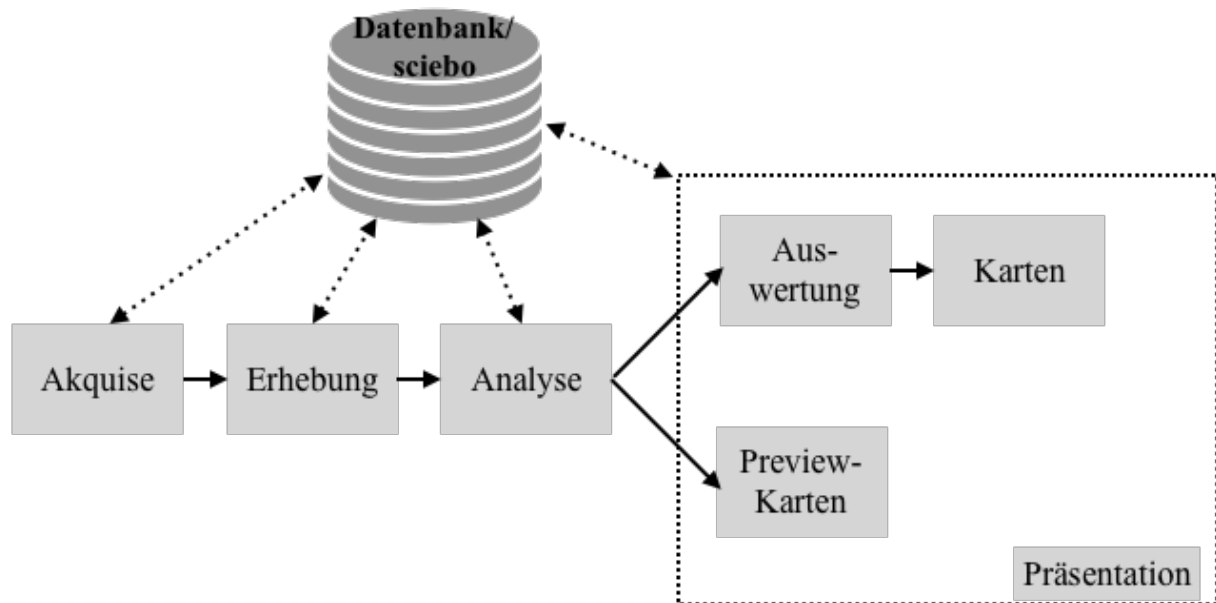
2. Workflow im Projekt

Der Workflow lässt sich mit den vier Domänen *Akquise*, *Erhebung*, *Analyse* und *(Karten-)Präsentation* darstellen, für die jeweils eigene Schnittstellen (sog. *Interfaces*) zur Datenbank existieren.⁷ Die Kartenpräsentation umfasst zunächst dynamisch erzeugte Previewkarten mit linguistisch noch nicht tiefergehend analysierten Sprachdaten. Später (ab 2026) wird zusätzlich ein Teil der Daten ausgewertet und auf statischen Karten präsentiert (vgl. Grafik 1).

⁵ Für eine Kurzbeschreibung vgl. <https://www.dmw-projekt.de/projektbeschreibung/>. – Erhoben werden die folgenden Dialekte: Westfälisch, Niederfränkisch, Ripuarisch, Moselfränkisch (NRW), Westfälisch, Nordniederdeutsch (SW-Niedersachsen), Moselfränkisch (NO-Rheinland-Pfalz).

⁶ <https://www.uni-siegen.de/sisal/>.

⁷ Entwickelt von Kai-Uwe Carstensen, Universität Siegen, dem die computerlinguistische Koordination des Projekts obliegt.



Grafik 1: Workflow des DMW

sciebo ist der Filehosting-Dienst von Hochschulen und Forschungseinrichtungen in NRW, über den die Audiodaten und projektrelevanten Dokumente (sowohl erhebungsbezogen als auch intern) verwaltet werden. Die gesamte Datenverwaltung (samt Datenbank und entsprechender Programmierfähigkeit) wird vom *ZIMT*⁸ der Universität Siegen bereitgestellt. Alle Daten können auf Anfrage für wissenschaftliche Zwecke zur Verfügung gestellt werden.

2.1 Akquise

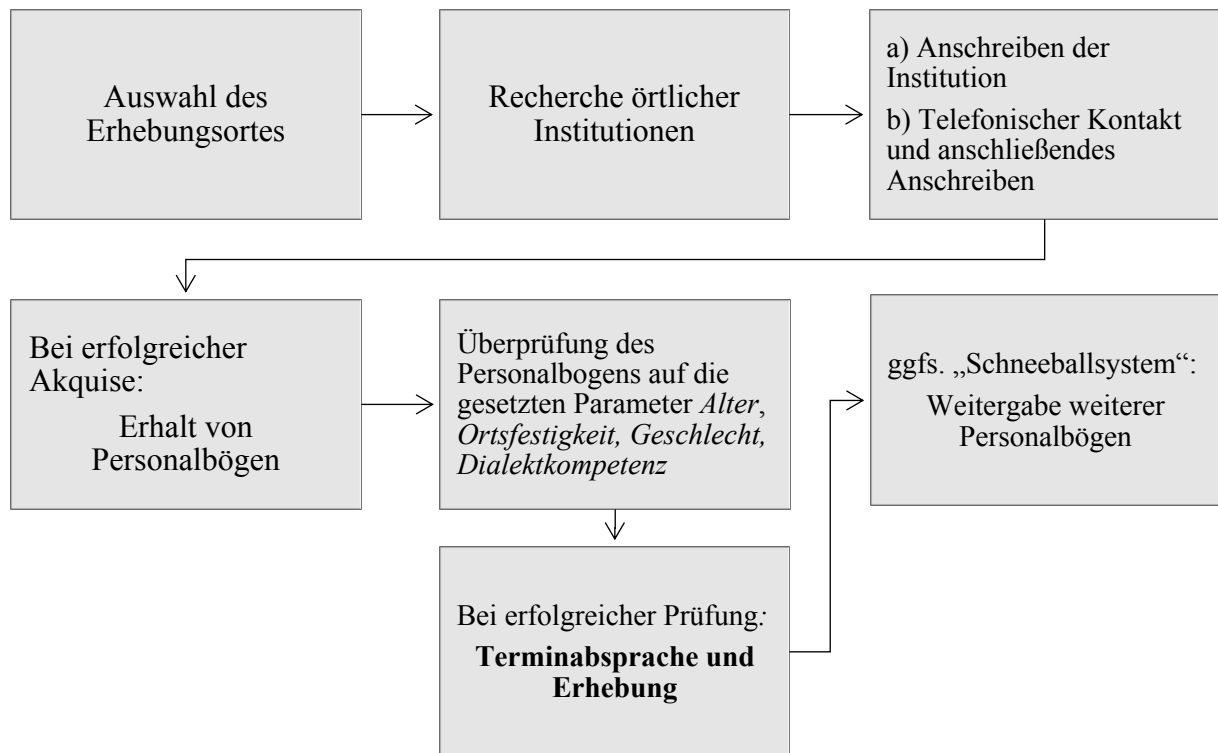
Die Orte, in denen Sprachdaten erhoben werden, sind Wenkerorte oder Orte, in denen im Rahmen von anderen Projekten schon Erhebungen von Sprachdaten durchgeführt wurden.⁹ Die Auswahl der Orte erfolgt nach der jeweiligen (heutigen) Größe des Wenkerortes sowie nach seiner Lage im Gradnetzfeld (vgl. dazu Übersicht im Anhang). Ansprechpartner*innen von örtlichen Institutionen wie Kirchengemeinden, Feuerwehren usw. werden gebeten, für das Projekt geeignete Personen zu identifizieren. Daneben reagieren Personen auf Medienberichte über den DMW (bzw. Aufrufe zur Teilnahme in lokalen Zeitungen) und melden sich für die Erhebungen. Potentielle GPs erhalten einen Personalbogen zum Ausfüllen. Auswahlkriterien sind Alter, Ortsfestigkeit und damit Sprachkompetenz (erhoben werden soll das Standardfernste) und Geschlecht.¹⁰ Erweist sich eine Person nach Prüfung des Personalbogens

⁸ Zentrum für Informations- und Medientechnologie.

⁹ Zum Beispiel *Sprachvariation in Norddeutschland* (SiN) (<https://corpora.uni-hamburg.de//sin/startseite.html>).

¹⁰ Vgl. Übersicht im Anhang.

als geeignet, wird mit der jeweiligen GP ein Ortstermin vereinbart. Die folgende Grafik bietet eine Übersicht über die Schritte, die sich für die Akquise bewährt haben:¹¹



Grafik 2: Schritte zur Gewinnung von Gewährspersonen (GPs)

Ist eine Terminabsprache erfolgt, fährt ein*e Explorator*in an den Wohnort der GP. Die Erhebung findet im Allgemeinen in deren häuslicher Umgebung statt.

Über das *Akquiseinterface* ist ein einfacher Überblick über bisherige und potentielle Erhebungsorte sowie eine ökonomische Verwaltung der Akquisevorgänge (Kontakterstellung, Kommunikation, Rückmeldungen etc.) möglich.¹²

2.2 Erhebung (Exploration)

Eine Erhebung dauert zweieinhalb bis fünf Stunden. Sie enthält die folgenden Komponenten:

1. Ausfüllen einer Einverständniserklärung, dass die (anonymisierten) Sprachdaten für wissenschaftliche Zwecke weiterverwendet werden dürfen.
2. Überprüfung der Personaldaten.

¹¹ Die Grafik wurde von Gero Gehrke und Stephanie Sauer Milch, beide Universität Münster, erstellt, bei denen wir uns herzlich dafür bedanken, dass sie sie uns für die vorliegende Publikation zur Verfügung stellen.

¹² Die bereits erhobenen Orte können unter https://dmw.zimt.uni-siegen.de/?page_id=8762 eingesehen werden.

3. Ausfüllen einer *mental map* (vgl. Abschnitt 1).
 4. Gewinnung der Sprachdaten mittels eines digitalisierten Fragebuchs, das sich aus den drei Teilen Kernfragebuch, Syntaxfragebuch und Wenkersätze zusammensetzt.
 5. Aufnahme der Sprachdaten mit dem *SpeechRecorder* (Draxler/Jänsch 2019), der es auf der Grundlage des Fragebuchs ermöglicht, die Antworten der GP vorzuschneiden und für die spätere Bearbeitung systematisch zu speichern.
 6. Sicherheitsaufnahme mit einem zweiten Aufnahmegerät.
 7. Protokoll (Hinweise zur Erhebungssituation, zur Dialektfestigkeit usw.). Zusätzliche Hinweise, die die Gesamtaufnahme oder Teile davon betreffen, können in vielen Fällen Abweichungen von den für den Ort erwartbaren Antworten bzw. von der zweiten GP am Ort erklären (plötzliche Unsicherheit, Korrekturen, Gebrauch der Standardsprache usw.).
- Das unter Punkt 4 bis 6 Angeführte liegt den Explorator*innen auf einem Laptop vor. Hier sehen sie die ID, die Nummer im Fragebuch, die betreffenden Wortform(en), die Systemebenen, für die das/die Element(e) analysiert und ausgewertet werden, sowie direkt vorausgehende und folgende Fragen. Zudem wird die Tonspur angezeigt, die schon bei der Erhebung Auskunft über die Qualität der Aufnahme gibt, sodass die GP gegebenenfalls gebeten werden kann, die Antwort zu wiederholen. Schwierigkeiten im niederdeutschen wie auch im hochdeutschen Raum ergeben sich dadurch, dass die GPs bestimmte Wörter im Dialekt nicht (mehr) kennen, die jüngeren zum Teil auch nicht mehr in der Standardsprache. Dies betrifft u. a. Bezeichnungen aus dem landwirtschaftlichen Bereich.
- Die Erhebungsdaten werden nach der Exploration über das *Explorationsinterface* auf *sciebo* hochgeladen bzw. in die Datenbank eingetragen. Dazu gehören: die Einverständniserklärung, die Gesamtaufnahme, die vom *SpeechRecorder* erzeugten Audio-Dateien (im wav-Format¹³), die *mental map* und der Personalfragebogen sowie die im handschriftlichen Protokoll notierten Gesprächs(meta)daten (z. B. über Aufteilung einer Erhebung).

2.3 Analyse

Die Analyse im *Analyseinterface* umfasst die Weiterbearbeitung der vorgeschrittenen Antworten. Studentische oder wissenschaftliche Hilfskräfte schneiden die interessierende sprachliche Form, die später auf der Karte zu hören sein wird, aus der Antwort und transkribieren sie nach IPA (International Phonetic Alphabet). Schließlich erfolgt die

¹³ Containerformat zur digitalen Speicherung von Audiodaten. – Auch die Gesamtaufnahmen liegen als wav-Dateien vor.

Phänomenbearbeitung, in der spezifische, phänomenbezogene Aspekte der betreffenden Form in die Datenbank eingetragen werden.

Das *Analyseinterface* unterstützt die Bearbeiter*innen bei ihrer Tätigkeit. Hierzu gehören u. a. die Auswahl der jeweiligen Aufgaben¹⁴, die Bearbeitung der wav-Dateien und eine im Hinblick auf Effizienz und Qualität automatische Unterstützung der Transkription (SAMPA-Eingabe, Fehlerchecks usw.). Zudem erfolgt eine automatische Umwandlung eines Transkripts in eine leichter lesbare populärwissenschaftliche Transkription (POP), die von phonetischen Details abstrahiert.¹⁵

Insgesamt ermöglicht das Analyseinterface die Anzeige relevanter Informationen, die für die Einschätzung der GP (z. B. relevante Metadaten wie das jeweilige Dialektgebiet der GP), der Antwort (im Protokoll notierte Beeinflussung durch anwesende Personen), der Aufnahme (dort notierte Störgeräusche) oder für die Qualität der Transkription (Anhören der anderen GP am Ort) gegebenenfalls wichtig sind.

2.4 Kartenpräsentation

Sobald Analysedaten in die Datenbank eingespeist worden sind, können aus ihr über das *Karteninterface* dynamische Previewkarten generiert werden, auf denen nicht nur in unterschiedlicher Granularität angezeigt wird, wo wie gesprochen wird, sondern auf denen sich die vorhandenen standardfernen Äußerungen (die geschnittenen wav-Dateien) per Mausklick auf einen Ort direkt anhören lassen (dies gilt auch für die Wenkersätze, die separat auswählbar sind). Der einfache, explorative Zugang zu den Sprachdaten macht den DMW auch für Laien attraktiv (Wie spricht man denn in der Nähe von mir/weiter entfernt von mir/wo meine Tochter wohnt/...?).

Wie der Terminus *Previewkarte* anzeigt, handelt es sich dabei um Vorschau-/Voransichtskarten, die nicht linguistisch ausgewertet sind. Die Variantentypen, die die Previewkarten zeigen, werden in variabler Granularität rein automatisch erzeugt. Hiermit lässt sich das Problem der enormen phonetischen Variation handhaben, mit dem eine Präsentation nicht ausgewerteter Sprachdaten konfrontiert ist.

3. Schlussbemerkungen

¹⁴ Zur Vermeidung von Bearbeiter*innenisoglossen wird die Bearbeitung einer (Teil-)Aufgabe/Frage möglichst einer einzigen Person zugewiesen.

¹⁵ Die POP-Notation folgt den sog. Hamburger Konventionen (Bieberstedt/Ruge/Schröder 2016).

Der DMW ist ein moderner Sprachatlas, für dessen Erstellung traditionelle und innovative Methoden zusammengeführt werden. Für die Datengewinnung wird mit dem Fragebuch auf die bewährte Erhebungsgrundlage und mit der direkten Erhebung auf die bewährte Form zurückgegriffen. Bei der Erhebung, der Analyse wie der Präsentation der Sprachdaten wird modern(st)e Technik verwendet. Nicht nur durch die Abbildung von niederdeutschem und hochdeutschem Sprachmaterial in einem Atlas wird methodisch Neuland betreten, sondern auch bei der Anwendung von Tools, die für die deutsche Standardsprache entwickelt wurden. Der DMW ist (insbesondere auch in technischer Hinsicht) in engem Austausch mit verwandten Projekten wie *regionalsprache.de* (REDE) und *Deutsch in Österreich. Variation – Kontakt – Perzeption* (DiÖ).

Literaturverzeichnis

Bieberstedt, Andreas/Ruge, Jürgen/Schröder, Ingrid (2016): Hamburger

Transkriptionskonventionen. In: Bieberstedt, Andreas/Ruge, Jürgen/Schröder, Ingrid (Hg.): Hamburgisch. Struktur, Gebrauch, Wahrnehmung der Regionalsprache im urbanen Raum (Sprache in der Gesellschaft 34). Frankfurt am Main u. a., S. 421-428.

Deutsch in Österreich. Variation – Kontakt – Perzeption (DiÖ) (<https://dioe.at/>, letzter Zugriff 15.01.2020).

Digitaler Wenker-Atlas (DiWA) (<http://www.diwa.info/Geschichte/Uebersicht.aspx>, letzter Zugriff 15.01.2020).

Dialektatlas Mittleres Westdeutschland (DMW) (<https://www.dmw-projekt.de/>, letzter Zugriff 15.01.2020).

Draxler, Christoph/Jänsch, Klaus (2019): SpeechRecorder (<https://www.bas.uni-muenchen.de/Bas/software/speechrecorder/>, letzter Zugriff 15.01.2020).

regionalsprache.de (REDE) (<https://www.regionalsprache.de/>, letzter Zugriff 15.01.2020).

Schmidt, Jürgen Erich/Dammel, Antje/Girnth, Heiko/ Lenz, Alexandra N. (2019): Sprache und Raum im Deutschen: Aktuelle Entwicklungen und Forschungsdesiderate. In: Herrgen, Joachim/Schmidt, Jürgen Erich (Hg.): Sprache und Raum. Ein internationales Handbuch der Sprachvariation. Bd. 4: Deutsch. Unter Mitarbeit von Hanna Fischer und Brigitte Ganswindt (Handbücher zur Sprach- und Kommunikationswissenschaft. 30.4). Berlin/Boston, S. 28-60.

sciebo. www.hochschulcloud.nrw (<https://www.sciebo.de/>, letzter Zugriff 15.01.2020).

Spiekermann, Helmut/Tophinke, Doris/Vogel, Petra M./Wich-Reif, Claudia (2017):
Dialektatlas Mittleres Westdeutschland (DMW). Dialect Atlas of Middle Western
Germany. In: Zeitschrift für Germanistische Linguistik 45, S. 386-388.
Sprachvariation in Norddeutschland (<https://corpora.uni-hamburg.de//sin/startseite.html>,
letzter Zugriff 15.01.2020).

Anhang

Grunddaten DMW

Gradnetzfelder: 9 x 12 km (in Anlehnung an den MRhSA) = 449 Gradnetzfelder,
davon 320 ganz, 129 partiell; möglichst gleichmäßige Verteilung der
Orte

Aufnahmeserien: 2 altersabhängige Serien (in Anlehnung an den MRhSA)
Serie I: 70 oder älter (ortsfest)
Serie II: ca. 30-45 (möglichst ortsfest)

Orte: Einwohnerzahl 500-ca. 8.000
Serie I: 1.089 Erhebungsorte
Serie II: 449 Erhebungsorte

Erhebungen: insg. 3.076; pro Ort 2 Erhebungen (möglichst ein Mann und eine Frau)
Serie I: 2.178
Serie II: 898

Fragebuch: insg. rd. 800 Fragen
Phonologie (42,5 %), Morphologie (26,5 %), Syntax (16,5 %), Lexik
(14,5 %)