Zur dialektometrischen Analyse des „Atlas zur deutschen Alltagssprache“

- Zulassungsarbeit -

eingereicht bei

Prof. Dr. Stephan Elspaß
Lehrstuhl für Deutsche Sprachwissenschaft

Philologisch-Historische Fakultät
Universität Augsburg

Lang, Benjamin

Abgabetermin: 02.10.2008

Inhaltsverzeichnis

1. Einführung 1

2. Der „Atlas zur deutschen Alltagssprache“ (AdA) 4

2.1 Erhebung.....6

2.2 Auswertung..... 7

2.3 Zusammenfassung.....9

3. Die dialektometrische Verfahrenskette 11

3.1 Datenmatrix.....12

3.2 Ähnlichkeitsmatrix..... 17

3.3 Clusteranalyse.....20

4. Auswertung 26

4.1 Wortgebrauchsräume.....26

4.2 Merkmalsdifferenzierte Wortgebrauchsräume…36

4.2.1 Geschlecht..... 38

4.2.2 Alter.....38

4.3 Gesamtgliederung……..46

5. Ausblick 57

Anhang 59

Literaturverzeichnis 67

Kapitel 1

Einführung

Der Titel der vorliegenden Arbeit besteht aus zwei Komponenten, die miteinander verbunden werden: „Dialektometrie“ sowie „Atlas zur deutschen Alltagssprache“. Auf der einen Seite also eine spezielle Methodik innerhalb der Dialektgeographie, auf der anderen Seite ihr wichtigstes Untersuchungsmedium, der Sprachatlas als Sammlung von Sprachkarten eines bestimmten Gebietes.

Der Begriff „Dialektometrie“ existiert dem Namen nach seit 1973[1] und ist im Zusammenhang mit dem zunehmenden Einsatz des Computers in der Sprachgeographie zu sehen. GOEBL, ein wichtiger Vertreter dieser Forschungsrichtung, definiert „Dialektometrie“ prägnant als „eine auf geolinguistische Daten angewendete numerisch-klassifikatorische Methode“[2]. Was verbirgt sich hinter dieser abstrakten Definition? Die *geolinguistischen Daten* eines Sprachatlasses bauen sich aus mehreren Einzelkarten, d. h. mehreren Variablen, auf. Jede Einzelkarte behandelt eine spezifische Fragestellung und weist eine bestimmte diatopische Verteilung auf, die sich von Karte zu Karte mehr oder weniger unterscheiden kann. Möchte man nun in dem in seiner Gesamtheit unüberschaubaren Material eines Sprachatlasses versuchen, verborgene Ordnungsstrukturen sichtbar zu machen, bietet sich der Einsatz von Computern an. Mit ihm lassen sich *numerisch-klassifikatorische Methoden* (s. Kapitel 3.3),

[1] Goebl (1993, 37)

[2] Goebl (1993, 38)

die der „Erkennung von Ordnungsstrukturen bzw. Ordnungsmustern“[3] dienen, durchführen. Die Forschungsrichtung der Dialektometrie steht damit ganz in der Tradition der Dialektgeographie, die versucht, sprachlich einheitliche Gebiete mittels Isoglossen voneinander abzugrenzen, unterscheidet sich jedoch in ihrer methodischen Vorgehensweise.

Die vorliegende Arbeit untersucht mit den methodischen Mitteln der Dialektometrie das sprachliche Material des „Atlas zur deutschen Alltagssprache“. Das Besondere des untersuchten Sprachatlasses liegt in seiner Erhebungsmethode, die entgegen der systematischen Vorgehensweise konventioneller Atlanten unsystematisch über eine Online-Erhebung erfolgt. Für nähere Erläuterungen wird deshalb zunächst der „Atlas zur deutschen Alltagsprache“ in Kapitel 2 näher vorgestellt, wobei neben der eben erläuterte Erhebungsmethode (Kapitel 2.1) auf weitere für die dialektometrische Analyse wichtige Eigenschaften eingegangen wird (Kapitel 2.2 und 2.3).

Eine dialektometrische Analyse des „Atlas zur deutschen Alltagssprache“ stellt insofern ein Experiment dar, als dass erstmals überprüft wird, ob sich das über eine unsystematische Online-Erhebung gewonnene sprachliche Material für ein derartiges methodisches Vorgehen überhaupt eignet. Experimente können gelingen, sie können aber auch schief gehen. Deshalb liegt der Schwerpunkt der Arbeit zunächst einmal auf der Frage, ob die Anwendung einer in der Dialektometrie etablierten *dialektometrischen Verfahrenskette* (Kapitel 3) auf das sprachliche Material des „Atlas zur deutschen Alltagssprache“ zu plausiblen Ergebnissen führt. In Kapitel 3 wird dazu detailliert auf das methodische Vorgehen eingegangen, indem die Erzeugung einer „Datenmatrix“ (Kapitel 3.1), einer „Ähnlichkeitsmatrix“ (Kapitel 3.2) und das numerisch-klassifikatorische Verfahren der „Clusteranalyse“ (Kapitel 3.3) geschildert werden.

Auch wenn ein Hauptaugenmerk auf den dialektometrischen Analyseschritt gerichtet ist, darf Dialektometrie dennoch nicht lediglich um seiner selbst Willen betrieben werden. Die Methoden der Dialektometrie müssen angewendet werden, um bestehende Hypothesen und Theorien zu überprüfen oder neue zu

[3] Goebl (1982, 11)

gewinnen. In Kapitel 4 wird dieser Forderung Rechnung getragen und einige exemplarische Anwendungsbeispiele vorgestellt. Während in Kapitel 4.1 anhand so genannter *Ähnlichkeitskarten* versucht wird, bestimmte Wortgebrauchsräume voneinander abzugrenzen, geht es in Kapitel 4.2 vor allem um die Frage, inwiefern sich der regionale Wortgebrauch innerhalb einer Generation verändert hat, was durch eine altersspezifische Differenzierung des Datensatzes beantwortet werden kann. In Kapitel 4.3 wird schließlich versucht, anhand der Ergebnisse einer Clusteranalyse eine Gesamtgliederung des deutschen Sprachgebiets vorzunehmen.

Abschließend wird in Kapitel 5 ein kurzer Ausblick über zukünftige Anwendungsmöglichkeiten gegeben und methodische Verbesserungsvorschläge erörtert.

Kapitel 2

Der „Atlas zur deutschen Alltagssprache“ (AdA)

Der „Atlas zur deutschen Alltagssprache“ (AdA) ist ein von STEPHAN ELSPAß und ROBERT MÖLLER im Jahr 2003 initiiertes Projekt[4], das sich konzeptionell auf den „Wortatlas der deutschen Umgangssprachen“ (WDU) von JÜRGEN EICHHOFF[5] stützt.

Beiden Atlanten gemein ist die Erforschung und Dokumentation der „Alltagssprache“ (ELSPAß / MÖLLER) bzw. der „Umgangssprache“ (EICHHOFF)[6], also der nicht standardisierten Sprache, welche im täglichen Umgang (z. B. im Gespräch im Supermarkt) benutzt wird. Die sich im ständigen Wandel befindliche Umgangssprache folgt weder ganz einer dialektalen Norm noch einer standardsprachlichen. Vielmehr sind die „in der Alltagssprache dominierenden Sprachformen [...] durch eine häufig kaum berechenbare Variation“[7] gekennzeichnet, für die man ein Modell von Kontinua zwischen standardnäheren und dialektnäheren Erscheinungsformen als jeweilige Pole annimmt.

---

[4] Vgl. zur Entstehungsgeschichte des Atlasprojekts Elspaß (2005, 1f.).

[5] Eichhoff (1977ff.)

[6] Im Folgenden soll terminologisch zwischen Verwendungsweise („Alltagssprache“) und Existenzform („Umgangssprache“) differenziert werden (vgl. Möller (2003, 259)).

[7] Möller (2003, 260)

Diese Tatsache ist für Untersuchungen der räumlichen Variation in doppelter Hinsicht bedeutsam: Zum einen besitzt die Umgangssprache eine *nicht-diatopische* Variation, was bedeutet, dass für eine Region keine völlige Homogenität postuliert werden kann. Diese stellt für diatopische Erhebungen einen in der Natur des Untersuchungsgegenstandes liegenden Störfaktor, statistisch gesprochen ein so genanntes *Hintergrundrauschen*, dar. Zum anderen zeichnet sich die Umgangssprache „in der Regel fast immer noch durch regionalspezifische, auf die Dialekte zurückweisende Merkmale aus“[8]. Die *vertikale* Variation zwischen Standard- und Dialektelementen besitzt also auch eine diatopische Dimension, die ohne Zweifel auf die Unterschiede in den Basisdialekten zurückgeht, jedoch nicht 1:1 mit dieser übereinstimmt.[9] Auch wenn ein fortschreitender Rückgang des Dialektgebrauchs in der Bevölkerung allgemein festzustellen ist, können Sprecherinnen und Sprecher aufgrund bestimmter Merkmale (Lexik, Aussprache, Grammatik) weiterhin einer bestimmten Region zugeordnet werden.

Die Erfassung und Dokumentation der räumlichen Verteilung dieser „regional gefärbten Sprache oberhalb der Dialekte“[10] war und ist (auch in Zukunft) ein wichtiges Forschungsgebiet der Dialektgeographie. Als Pionier kann in dieser Hinsicht EICHHOFF bezeichnet werden, der im Zeitraum zwischen 1977 und 2000 vier Bände seines „Wortatlas der deutschen Umgangssprachen“ (WDU) veröffentlichte. Dieselben Ziele verfolgt das Forschungsprojekt „Atlas zur deutschen Alltagssprache“ (AdA), das in Anlehnung an EICHHOFFs WDU die diatopische Variation der Umgangssprache untersucht. Intention ist zum einen, die Vielfalt des Deutschen zu erfassen, zum anderen durch den Vergleich der alten Karten des WDU mit den entsprechenden neuen Sprachkarten des AdA den Wandel der deutschen Sprache zu belegen.

---

8 Elspaß / Möller (2006, 144)

9 Gründe hierfür s. Möller (2003, 260).

10 Möller / Elspaß (2008, 115)

2.1 Erhebung[11]

Die Erhebung erfolgt beim „Atlas zur deutschen Alltagssprache“ wie bei EICHHOFFs WDU über die indirekte Befragungsmethode, was bedeutet, dass die Informanten als Experten für die jeweilige regionale Umgangssprache dienen.[12] Das Neue im Vergleich zum Verfahren der Datenerhebung des WDU liegt darin, dass beim AdA die Möglichkeiten der neuen Medien genutzt werden. Der Fragebogen, der vor allem lexikalische aber auch phonologische, morphologische und syntaktische Phänomene erfasst, befindet sich für alle frei zugänglich im Internet[13]. Neben der Ortschaft, für die die Angaben gemacht werden, und den Antworten zu den einzelnen Fragen können die Informanten freiwillig Auskunft über ihr Geschlecht, Alter, Wohndauer in der Ortschaft, Beruf, etc. geben, was weitere Analysemöglichkeiten (s. Kapitel 4.2) eröffnet. Mittels eines *php-Codes* werden die Antworten, die die einzelnen Personen in das Formular eingegeben haben, in eine Textdatei umgewandelt und stehen der weiteren Auswertung zur Verfügung.

Der „Atlas zur deutschen Alltagssprache“ befindet sich mittlerweile (Stand: September 2008) in der fünften Umfragerunde, wobei eine Erhebungsphase im Durchschnitt einen einjährigen Zeitraum einnimmt. Nachdem für das Pilotprojekt im Jahr 2003 gezielt Personen per E-Mail angeschrieben wurden, hat sich die Zahl der Informanten seitdem durch das *Schneeballsystem* deutlich erhöht. Während für Runde 1 lediglich 1.763, für Runde 2 2.478 und für Runde 3 1.980 Meldungen registriert werden konnten, erlebte der Atlas mit 6.863 Informanten in Runde 4 einen deutlich gesteigerten Zuspruch. Für Runde 5 ge-

---

[11] Für genauere Informationen zur Erhebungsmethode, die an dieser Stelle nicht gegeben werden können, s. Elspaß (2005, 2-5), Elspaß / Möller (2006, 144-149) und Möller / Elspaß (2008, 116-119).

[12] Verwirklicht wird die indirekte Befragungsmethode durch die dem Fragebogen vorangestellte Aufforderung an die Informanten: „Bitte geben Sie bei den folgenden Fragen jeweils an, welchen Ausdruck man in Ihrer Stadt normalerweise hören würde - egal, ob es mehr Mundart oder Hochdeutsch ist.“

[13] http://www.philhist.uni-augsburg.de/ada

hen nach dem aktuellem Stand der Erhebung 4.360 Antworten in die weiteren Analyseschritte ein.

Für eine dialektometrische Auswertung der Sprachdaten ist die Zunahme der Rückmeldungen eine erfreuliche Nachricht. Zum einen können umso detailliertere (z. B. Selektion der Daten nach Geschlecht, Alter oder sozialem Milieu) und statistisch sicherere Aussagen getroffen werden, je mehr Personen für einen Ort Angaben machen. Zum anderen wird das *Hintergrundrauschen* durch eine möglichst große Anzahl an Ortsmeldungen reduziert, sodass z. B. auch gezielt falsche Rückmeldungen, die bei der Online-Erhebung nicht ausgeschlossen werden können, umso geringer ins Gewicht fallen, je mehr Personen für denselben Ort Angaben machen. Der Ausgleich erfolgt hier also über die Masse, was als großer Vorteil der neuen Erhebungsmethode des AdA gewertet werden kann.

2.2 Auswertung[14]

Bei der dialektographischen Auswertung des Rückmeldedatensatzes orientiert sich der AdA an der Grundkarte des WDU von EICHHOFF. Die jeweiligen Antworten der Gewährspersonen werden mittels eines *Spitbol-Skripts* bei Übereinstimmung der Ortsangabe zu einem der (zurzeit) 484 Ortspunkte zugeordnet.[15] In diesen sind die 402 Ortspunkte von EICHHOFFs WDU enthalten, die im Verlaufe der einzelnen Runden um 82 zusätzliche ergänzt wurden. Für jeden einzelnen Ortspunkt der jeweiligen Frage ermittelt das Skript in einem nächsten Schritt die häufigste Variante aus den Antworten, die zu einem Ort zugewiesen wurden, und schreibt diese in eine Datenbank. Zudem wird nach dem gleichen Prinzip eine weitere Datenbank generiert, welche die zweithäufigsten Varianten

[14] Für genauere Informationen zum Thema *Auswertung*, die an dieser Stelle nicht gegeben werden können, s. Elspaß / Möller (2006, 150-152).

[15] Bei dieser Methode können Meldungen, die aus im Spitbol-Skript nicht vorhandenen Ortschaften kommen, nicht berücksichtigt werden. Für Verbesserungsvorschläge s. Kapitel 5.

enthält, insofern deren prozentuale Nennung einen Schwellenwert von 33 % überschreitet.

Die Kartierung erfolgt mit Hilfe eines **G**eographischen **I**nformations**S**ystems (*ArcView*), das die während der Auswertung entstandene Datenbank in Karten umsetzt. Jedem der 484 Ortspunkte der Grundkarte wird, wenn sich für ihn mindestens eine Gewährsperson in der Erhebungsphase gemeldet hat, aus der generierten Datenbank für jede Frage die Haupt- und – wenn vorhanden – die Nebenvariante zugewiesen. Das Programm bietet verschiedene Möglichkeiten, Kartenelemente (d. h. die Ortspunkte der Grundkarte) in Abhängigkeit von den zugehörigen Einträgen eines ausgewählten Feldes (d. h. die aus den Orten gemeldeten Varianten) differenziert darzustellen (s. Abb. 1).[16]

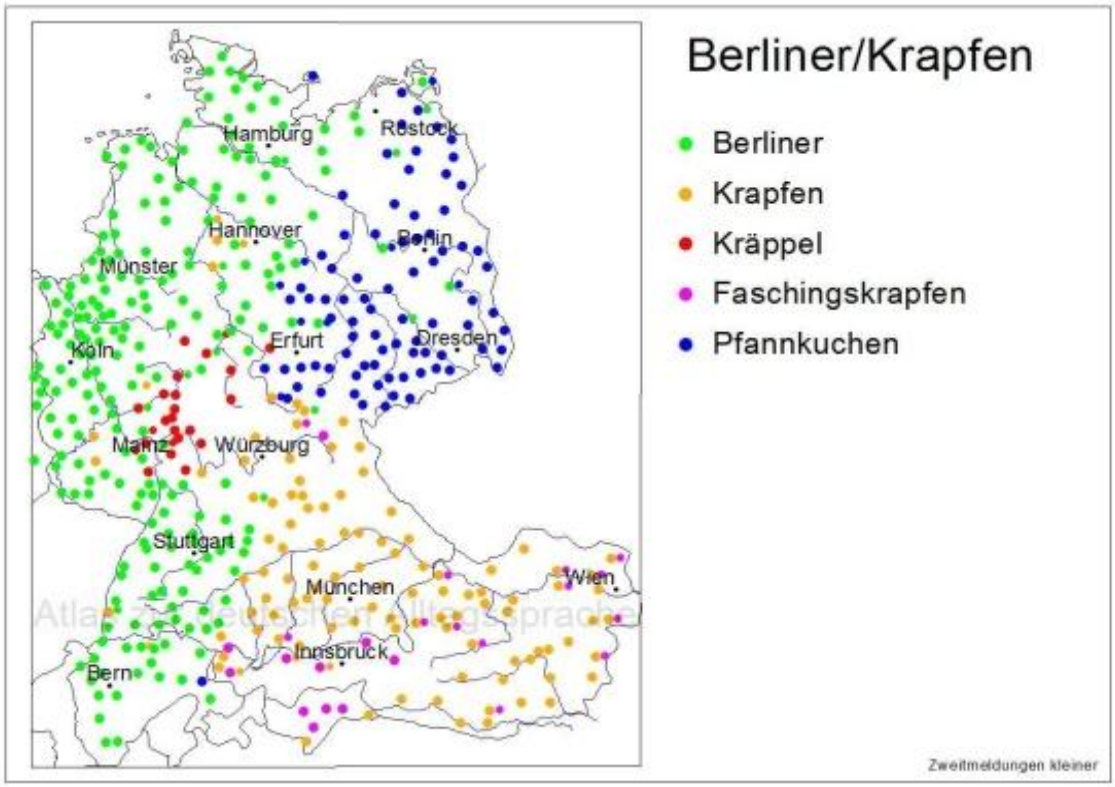


Abb. 1: Beispiel einer online veröffentlichten Karte des „Atlas zur deutschen Alltagssprache“ (Quelle: http://www.philhist.uni-augsburg.de/lehrstuehle/germanistik/sprachwissenschaft/ada/runde_4/f03/)

[16] Vgl. Elspaß / Möller (2006, S. 150).

2.3 Zusammenfassung

Als Endprodukt der Erhebungs- und Auswertungsphase des AdA entsteht für jede Frage eine anschauliche Karte, bei der jeweils ein Punktobjekt einen Ortspunkt repräsentiert. Aufgrund der großen Anzahl der Punkte entsteht insgesamt, vor allem bei ausgeprägter diatopischer Variation, ein flächenhafter Eindruck, der unterschiedliche Sprachgebiete bzw. Wortgebrauchsräume erkennen lässt.

Werden mehrere Ortspunkte nicht kartiert, geht dieser Eindruck allerdings verloren. Das geschieht dann, wenn sich während der Erhebungsphase für einen Ort kein Informant gemeldet hat. Damit ist bereits ein Nachteil der Erhebungsmethode des AdA angesprochen, der sich, wie sich zeigen wird, auf die dialektometrische Auswertung niederschlägt. Im Vergleich zu der herkömmlichen Methode, wie sie EICHHOFF bei der Erfassung der Daten für den WDU anwendet, d. h. nach traditionell dialektologischem Muster mit Befragung von ein bis zwei Gewährspersonen pro Ort, besitzt die Online-Erhebung zwar einen geringeren Aufwand und ist weitaus kostengünstiger, allerdings erfolgt keine systematische Aufnahme der Daten. Während aus Ortspunkten mit hoher Bevölkerung (z. B. Berlin, München, Wien, Zürich, ...) sehr viele Gewährspersonen Antworten senden, treffen aus anderen nur sehr wenige oder gar keine Meldungen ein, sodass für diese Orte keine Kartierung erfolgen kann. Somit bleibt jede Runde darauf zu „hoffen“, dass sich für alle Ortspunkte Gewährspersonen melden, was in den bisherigen Runden noch nicht gelungen ist.

Allerdings besitzt die Online-Erhebung des AdA einen entscheidenden Vorteil gegenüber der herkömmlichen Methode: die enorme Masse der Antworten. Auch wenn diese sich nicht gleichmäßig auf die einzelnen Ortspunkte verteilen, so lassen sich dennoch genauere und differenziertere Analysen durchführen sowie statistische sicherere Aussagen treffen. Zum einen erfolgt bei mehreren Informanten pro Ort eine bessere Repräsentation der lokalen Gegebenheiten, zum anderen wirkt die Masse, wie bereits erwähnt wurde, als Korrektiv bei versehentlichen oder bewusst falschen „Ausreißerdaten“.

Der größte Vorteil liegt allerdings darin, dass differenzierte dialektometrische Analysen durchgeführt werden können. Da neben den Antworten zu den einzelnen Fragen zusätzlich Sozialdaten zu Alter, Geschlecht, Beruf, ... erfasst werden, können gezielt Untergruppen aus dem gesamten Datensatz selektiert werden. Beispielsweise können die diatopischen Verhältnisse verschiedener Berufs- oder Altersgruppen miteinander verglichen bzw. Analysen nach dem Geschlecht getrennt durchgeführt werden. Dadurch, dass pro Ortspunkt mehrere Gewährspersonen Angaben machen, sind derartige Analysen überhaupt erst möglich.[17]

[17] Die Informanten sind laut Fragebogen zwar dazu aufgefordert, als Repräsentanten ihres Ortes anzugeben, „welchen Ausdruck man in Ihrer Stadt normalerweise hören würde“. Dennoch ist davon auszugehen, dass die Gewährspersonen jeweils geprägt von ihrer persönlichen Situation und ihrem Lebensumfeld die Umwelt wahrnehmen und deshalb auf Unterschiede bezüglich Alter, Geschlecht oder Beruf geschlossen werden kann.

Kapitel 3

Die dialektometrische Verfahrenskette

Insgesamt umfasst das Korpus des „Atlas zur deutschen Alltagssprache“ einschließlich der fünften Runde *N* = 224 Einzelkarten (= Variablen), von der jede Einzelinformation zu den diatopischen Verhältnissen eines spezifischen Phänomens enthält. Um in dem in seiner Gesamtheit unüberschaubaren sprachlichen Material höherrangige Muster und Ordnungsstrukturen zu erkennen, was, wie bereits erwähnt wurde, das Ziel einer dialektometrischen Auswertung darstellt, bedarf es statistischer Methoden, die die Informationen der Einzelkarten bündeln. Obwohl von Seiten der Statistik ein breites Spektrum an unterschiedlichen Verfahrensweisen dafür angeboten wird, hat sich in der Dialektometrie eine Vorgehensweise durchgesetzt, die erstmals GOEBL beschrieben hat.[18]

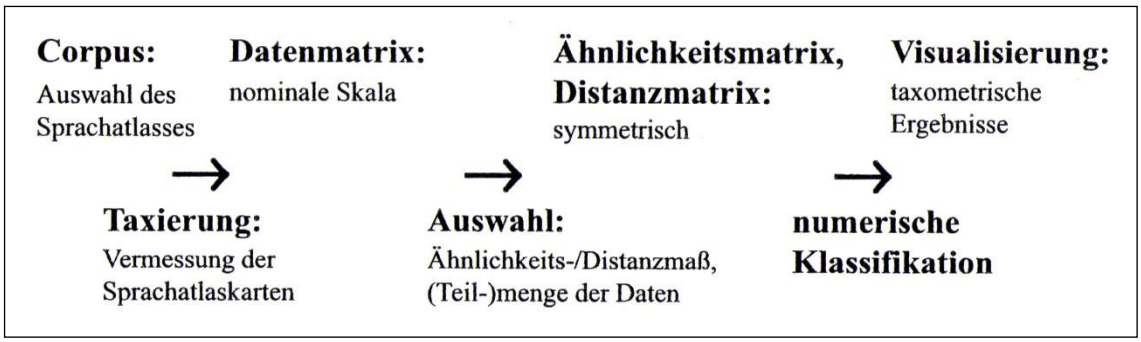


Abb. 2: Die dialektometrische Verfahrenskette (Quelle: Bauer (2002/2003, 216))

[18] Für weitere Informationen siehe Goebl (1982).

Die *dialektometrische Verfahrenskette* (s. Abb. 2) umfasst drei voneinander abgrenzbare methodische Schritte. Zunächst wird durch den hier mit *Taxierung* bezeichneten Vorgang auf Basis des Korpus eines Sprachatlasses eine Datenmatrix mit nominaler Skalierung erzeugt. Diese dient als Ausgangsmaterial für das eigentliche dialektometrische Arbeiten, nämlich der *Umwandlung* der auf nominalem Skalenniveau befindlichen Datenmatrix in eine symmetrische Ähnlichkeits- oder Distanzmatrix. Zu guter Letzt schließt der wichtigste Schritt in der dialektometrischen Verfahrenskette an: die *numerische Klassifikation*, die auf die in der Ähnlichkeitsmatrix enthaltenen Werte angewendet wird.

Die einzelnen Schritte der dialektometrischen Verfahrenskette wurden auch bei der Analyse der sprachlichen Daten des AdA durchgeführt. In den folgenden Kapiteln wird das Vorgehen näher erläutert.

3.1 Datenmatrix

Bevor eine dialektometrische Analyse zur Raumbildung überhaupt durchgeführt werden kann, ist eine geeignete Arbeitsgrundlage notwendig, d. h. eine nominale Datenmatrix. Diese baut sich, wie Abb. 3 verdeutlicht, aus *p* Variablen[19] und *N* Ortspunkten auf. Die Bezeichnung „nominal“ meint, dass die Werte auf nominalem Skalenniveau vorliegen:

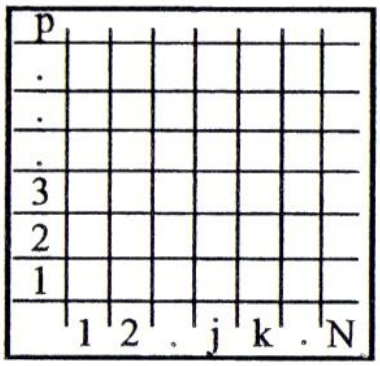


Abb. 3: Aufbau einer Datenmatrix (Quelle: Bauer (2002/2003, 216))

> Stellen die Ausprägungen (Werte) einer Variablen nur „Namen“ im Sinne einer Bezeichnung durch ein Wort, einen Buchstaben oder eine Zahl dar und sind die Ausprägungen nicht im Sinne einer Größer- / Kleiner-Relation vergleichbar, handelt es sich um eine nominalskalierte Variable.[20]

---

[19] Zum Verständnis sei an dieser Stelle nochmals gesagt, dass mit „Variablen“ die Einzelfragen / Einzelkarten eines Sprachatlasses (hier: des AdA) gemeint sind.

[20] Bahrenberg / Giese / Nipper (1999, 15)

Nominal skalierte Daten können demnach bezüglich ihrer Ausprägung nur dahingehend untersucht werden, ob ihre Elemente gleich ($x_i = x_j$) oder ungleich ($x_i \neq x_j$) sind. Angewendet auf die Analyse des AdA bedeutet es, dass die Ortspunkte hinsichtlich der Ausprägung einer Variable, d. h. einer Einzelfrage, (z. B. „Bezeichnung eines Menschen der Fußball spielt“) entweder gleich sind (z. B. *Fußballer* = *Fußballer*) oder sich voneinander unterscheiden (z. B. *Fußballer* ≠ *Tschütteler*). Bei der numerischen Auswertung der Daten des AdA bleibt also, wie es für Analysen von Sprachatlanten üblich ist, unberücksichtigt, ob und inwiefern eine Variante einer anderen möglicherweise ähnlicher ist als einer weiteren.[21] Um derartige Aussagen treffen zu können, bedürfe es eines subjektiven und sehr zeitaufwändigen Eingriffs.

Die Grundlage für die dialektometrische Analyse bilden die bisher in den fünf Runden gewonnenen Daten des „Atlas zur deutschen Alltagssprache“. Sie umfassen einschließlich der fünften Befragungsrunde p = 224 Variablen, von denen die meisten eine ausgeprägte diatopische Variation zeigen und als Karten visualisiert wurden.[22] Problematisch für die statistische Auswertung ist die unsystematische Datengewinnung via Internet, durch die die Anzahl N der Ortspunkte von Runde zu Runde variiert[23] und einige Ortspunkte in manchen Runden nicht belegt sind. Bei der Erzeugung der im Folgenden vorgestellten einzelnen Datenmatrizen wird diese Tatsache insofern berücksichtigt, als dass nur diese Ortspunkte in die Analyse eingegangen sind, von denen in allen Runden Meldungen von Informanten vorliegen. Dies ermöglicht eine einheitliche statistische Auswertung aller Ortspunkte, hat aber bedauerlicherweise eine (teilweise) deutliche Reduktion der Datengrundlage zur Folge.

---

[21] Die Tatsache, dass beispielsweise die Variante *Schütteler* der Variante *Tschütteler* viel näher steht als *Fußballer* bleibt also bei der numerischen Auswertung der Daten unberücksichtigt.

[22] Siehe http://www.philhist.uni-augsburg.de/ada

[23] N_{Runde1} = 442, N_{Runde2} = 417, N_{Runde3} = 418, N_{Runde4} = 449, N_{Runde5} = 408

Insgesamt wurden in einem ersten Schritt aus dem erhobenen und durch die Auswertung bereits als Datenmatrix für jede Runde vorliegenden Material des AdA neun Datensätze für weitere Analysezwecke generiert (s. Tabelle 1):

Datensatz	N	P
D_{gesamt}	322	222
$D_{lexikalisch}$	322	119
$D_{morphologisch}$	322	24
$D_{phonologisch}$	322	29
$D_{syntaktisch}$	322	31
$D_{weiblich}$	244	93
$D_{männlich}$	244	93
D_{alt}	239	93
D_{jung}	239	93

Tabelle 1: Überblick über die erstellten Datensätze

D_{gesamt} enthält alle Variablen und stellt somit den größten Datensatz dar. Von den maximal möglichen 484 Ortspunkten sind 322 in allen Runden vertreten, die bei der Generierung des Datensatzes berücksichtigt werden konnten. Von den insgesamt 224 Variablen wurden bis auf zwei, bei denen es sich um doppelte Fragen handelt, alle ausgewählt, d. h. auch diejenigen mit einer gering ausgeprägten diatopischen Verteilung.[24]

Der Atlas zur deutschen Alltagssprache erkundet vornehmlich lexikalische Phänomene, allerdings auch solche auf morphologischer (z. B. Fugenelemente), phonologischer (z. B. Aussprache von „König“ oder „ich weiß“) und syntaktischer Ebene (z. B. Verlaufsformen, Satzstellung). Für spezielle Fragestellungen wurden deshalb aus dem großen Datensatz D_{gesamt} vier weitere Sub-

[24] Ein Ausschluss von Karten mit einer unklaren Arealbildung wie sie manche Autoren vornehmen (vgl. Möller (2003, 263)) ist nicht notwendig, da sie letztendlich zu keiner für die Clusteranalyse relevanten statistischen Verfälschung führen.

gruppen ($D_{lexikalisch}$, $D_{morphologisch}$, $D_{phonologisch}$, $D_{syntaktisch}$) extrahiert, insofern die Variablen eindeutig einer Ebene zugewiesen werden konnten. Die damit gewonnen Datenmatrizen bestehen alle aus 322 Ortspunkten, weisen jedoch aufgrund der Konzentration des AdA auf lexikalische Fragestellungen eine starke Variation bei der Anzahl der zugeteilten Variablen auf (s. Tabelle 1)[25], was sich, wie sich später zeigt, auf die statistische Aussagekraft auswirkt.

Wie bereits erwähnt wurde, nimmt der AdA neben den Antworten zu den Einzelfragen auch diverse Sozialdaten auf. Dies ermöglicht eine Differenzierung der von den Informanten gegebenen Antworten nach Parametern wie Geschlecht und Alter. Dabei ist es jedoch zwingend notwendig, dass aus einem Ort mindestens zwei Repräsentanten gemeldet haben, die sich in diesem Parameter unterscheiden.[26] Es wird also ein großer Datensatz mit vielen Informanten benötigt, was beim AdA vor allem in der vierten und fünften Runde verwirklicht ist.[27] Deshalb konnten bei der Erzeugung der nach den jeweiligen Parametern differenzierten Datensätzen nur diese beiden Runden berücksichtigt werden, was die geringe Anzahl an Variablen (p = 93 erklärt). Um einen besseren Vergleich bei der Auswertung zu gewährleisten, gingen zudem nur die Ortspunkte ein, die beim jeweiligen Parameter, d. h. bei den z. B. nach dem Geschlecht differenzierten Datensätzen sowohl in dem *weiblichen* als auch in dem *männlichen*, in beiden Runden vertreten sind, was wiederum zu einer starken Reduktion der Datengrundlage führt. Für die nach dem Geschlecht differenzierten Datensätze ($D_{weiblich}$, $D_{männlich}$) konnten auf diese Weise 244 Ortspunkte selektiert werden, für die nach dem Alter differenzierten (D_{alt}, D_{jung}) 239. D_{alt} umfasst die Meldungen von Informanten ab 30 Jahren, D_{jung} diejenigen der unter 30-jährigen.

---

[25] Eine eindeutige Zuordnung zu den vier Ebenen konnte bei 19 der insgesamt 222 Variablen aus D_{gesamt} nicht durchgeführt werden. Diese blieben unberücksichtigt.

[26] Die neue Erhebungsmethode des AdA über das Internet macht gegenüber der herkömmlichen systematischen Aufnahme mit 1-2 Repräsentanten pro Ort überhaupt erst eine Differenzierung des Datensatzes möglich, da hier mitunter sehr viele Antworten für einen Ort zustande kommen können.

[27] Zu den genauen Zahlen siehe Kapitel 2.1.

Die Online-Erhebung des AdA macht im ersten Schritt der dialektometrischen Verfahrenskette einen zusätzlichen zeitaufwändigen Eingriff erforderlich. Dadurch, dass die Informanten bei den meisten Fragen nicht nur von den Autoren vorgegebene Varianten auswählen, sondern auch selbstständig Antworten eingeben können, ist es möglich, dass eigentlich gleiche Antworten durch das Auswertungsprogramm nicht als solche anerkannt werden, was einen manuellen Eingriff erforderlich macht. Aus diesem Grund wurden die eben vorgestellten Datenmatrizen einer Vereinheitlichung unterzogen. Dabei wurde versucht, den Eingriff möglichst gering zu halten, um keine Manipulation der Daten zu verursachen. Die Beispiele in Tabelle 2 sollen die vorgenommenen Eingriffe veranschaulichen. Die zwei Varianten, die bei der computativen Auswertung als nicht identisch anerkannt worden wären, wurden jeweils zu einer vereinheitlicht:

Variante 1	**Variante 2**	**Vereinheitlichungsform**
Huder	*huder*	*Huder*
Nüsslisalat	*Nüssli-Salat*	*Nüssli-Salat*
Pöstler	*Poestler*	*Pöstler*
Peperoni	*Pepperoni*	*Peperoni*
Mäppchen (zumindest zu meiner Schulzeit)	*Mäppchen*	*Mäppchen*

Tabelle 2: Beispiele für manuelle Eingriffe in die Datensätze

Wie Tabelle 2 verdeutlich, handelt es sich bei den Eingriffen vor allem um Vereinheitlichungen bezüglich Rechtschreibfehler, Groß- / Kleinschreibung und des Umlauts. Welche Variante letztendlich durch welche ersetzt wird, ist prinzipiell egal, hauptsache sie sind exakt identisch und werden von dem auswertenden Programm als solche erkannt.

Zusätzlich größere Eingriffe erfolgten neben denjenigen, die der Vereinheitlichungen dienten, bei den Datensätzen $D_{lexikalisch}$, $D_{morphologisch}$, $D_{phonologisch}$, $D_{syntaktisch}$, da hier nur die Variationen auf der spezifischen Ebene interessieren.

Dementsprechend wurden beispielsweise im Datensatz $D_{lexikalisch}$ etymologisch zusammengehörige Varianten zusammengefasst[28] und im Datensatz $D_{syntaktisch}$ dialektale Varianten vereinheitlicht.[29]

3.2 Ähnlichkeitsmatrix

Das dialektometrische Arbeiten beginnt eigentlich erst mit dem zweiten Schritt der dialektometrischen Verfahrenskette (s. Abb. 2), nämlich der *Umwandlung* der im ersten Schritt erzeugten und vereinheitlichten Datenmatrix auf nominalem Skalenniveau in eine symmetrische Ähnlichkeits- / Distanzmatrix. Die Umwandlung beruht dabei auf dem paarweisen Vergleich der Elemente aller Ortspunkte im Hinblick auf Übereinstimmungen und Unterschiede. Ein dafür nützliches und in der Dialektometrie weit verbreitetes Instrumentarium stellt der von GOEBL definierte *Relative Identitätswert* (RIW_{jk}) dar.[30] Die Formel hierfür lautet:

$$RIW_{jk} = \frac{\sum_{i=1}^{p} (KOI_{jk})_i}{\sum_{i=1}^{p} (KOI_{jk})_i + \sum_{i=1}^{p} (KOD_{jk})_i}$$

RIW_{jk}	Maß für die Ähnlichkeit zwischen den Ortspunktvektoren j und k ($0 \leq RIW_{jk} \leq 1$)	
	j, k	Index für die verglichenen Ortspunkte
	p	maximale Anzahl der Einzelkarten
KOI	Koidentitäten	
KOD	Kodifferenzen	

[28] Z. B. Vereinheitlichung der Varianten *Tschütteler* und *Schütteler* zu *Tschütteler* (im Gegensatz zu *Fußballer*).

[29] Z. B. Vereinheitlichung der Varianten *ein ganz lieber Mensch* und *a ganz liabe Mensch* zu *ein ganz lieber Mensch* (im Gegensatz zu *ein ganz ein lieber Mensch*).

[30] Siehe z. B. Goebl (1993, 43-45), Bauer (2002/2003, 217), Hummel (1993, 7). Für ausführliche Informationen zur Methode siehe Goebl (1982, 74-86).

Hinter der auf den ersten Blick kompliziert anmutenden Formel nach GOEBL mit Koidentitäten (übereinstimmende Varianten) und Kodifferenzen (unterschiedliche Varianten) befindet sich ein relativ einfaches methodisches Prinzip, um die Ähnlichkeit zweier Ortspunkte zu berechnen.[31] Nach der Formel wird für zwei Ortspunkt geprüft, inwiefern die Antworten für die einzelnen Variablen (= Einzelkarten / Einzelfragen) übereinstimmen oder sich unterscheiden. Die Zahl der Übereinstimmungen wird für jedes Paar von Ortspunkten addiert, es wird also für jede mögliche Kombination von Ortspunkten ermittelt, in wie vielen Karten insgesamt Übereinstimmung zwischen den beiden verglichenen Ortspunkten vorliegt. Setzt man die Summe der Übereinstimmungen zwischen zwei Ortspunkten in Relation zur Gesamtzahl der Variablen, ergibt sich ein Ähnlichkeitswert zwischen 0 und 1[32], aus dem unmittelbar abzulesen ist, wie viel Gemeinsamkeit zwischen zwei Ortspunkten besteht. Umso näher der Wert an 0 liegt, desto unähnlicher sind sich die beiden verglichenen Ortspunkte (0 = maximale Unähnlichkeit), je näher er an 1 liegt, desto ähnlicher sind sie sich (1 = maximale Ähnlichkeit).

Aus den berechneten Werten lässt sich eine *symmetrische Ähnlichkeitsmatrix*[33] (wie sie Abb. 4 mit Prozentangaben verdeutlicht) erstellen, aus der die Ähnlichkeit zwischen zwei beliebigen Ortspunkten (O1, O2, …) unmittelbar abgelesen werden kann. Die Anzahl der Spalten und Zeilen ist aufgrund der Symmetrie gleich und ergibt sich aus der Anzahl der in die Analyse eingegangenen

	O1	O2	O3	O4
O1	(100)	66	33	0
O2	66	(100)	75	25
O3	33	75	(100)	50
O4	0	25	50	(100)

Abb. 4: Ähnlichkeitsmatrix (Quelle: Möller (2003, 265))

[31] Vgl. für Folgendes Möller (2003, 264).

[32] Bzw. zwischen 0 und 100 % (nach Goebl), wenn man den Wert mit 100 multipliziert.

[33] Zusätzlich lässt sich auch eine so genannte „Distanzmatrix“ berechnen, die vor allem zur Herstellung von Schottenkarten (vgl. z. B. Goebl (1981, 25-28)) dient. Dabei werden die einzelnen Werte der Ähnlichkeitsmatrix von 1 subtrahiert. Diese Methode wird allerdings bei der Untersuchung der Sprachdaten des AdA nicht verfolgt und deshalb nicht weiter ausgeführt.

Ortspunkte. Eine weit verbreitete Anwendung von Ähnlichkeitsmatrizen ist deren Visualisierung in *Ähnlichkeitskarten*. Formal gesehen wird dabei aus einer Ähnlichkeitsmatrix einer der Vektoren (d. h. einer der Ortspunkte) herausgegriffen, der als Bezugspunkt fungiert und dem die Ähnlichkeitswerte der anderen Ortspunkte zugewiesen sind. Ausgehend von diesem konkreten Bezugspunkt lassen sich auf Ähnlichkeitskarten einheitliche Gebiete voneinander abgrenzen. Sie können in dieser Hinsicht bei bestimmten Fragestellungen von Vorteil sein, allerdings besitzen Ähnlichkeitskarten den Nachteil der *Bezugspunktabhängigkeit*, was bedeutet, dass das gesamte Kartenbild vom jeweiligen Bezugsort abhängt. D. h. Ähnlichkeitskarten bieten in erster Linie Aufschluss über die Verhältnisse in nächster Nachbarschaft, nicht aber über weiter entferntere Gebiete:

> Hohe Übereinstimmung verschiedener Orte mit dem Bezugsort weist auch auf hohe Übereinstimmung der betreffenden Orte untereinander hin, während geringe Übereinstimmung mit dem Bezugsort über das Verhältnis dieser Orte zueinander nichts aussagt.[34]

Der soeben erläuterte zweite Schritt der dialektometrischen Verfahrenskette wurde auf die im ersten Schritt erzeugten Datenmatrizen angewendet. Mittels eines *FORTRAN-Skripts* wurden aus den neun nominalskalierten Datenmatrizen (s. Tabelle 1) neun symmetrische Ähnlichkeitsmatrizen mit metrischem Skalenniveau erzeugt, wobei sich die Anzahl der Spalten bzw. der Zeilen jeweils aus der Anzahl *N* der eingegangenen Ortspunkte ergibt.[35] Für weitere Analysezwecke (s. Kapitel 4.1) wurden zudem mit Hilfe eines GIS-Programms (*ArcView*) mehrere Ähnlichkeitskarten nach bereits erläutertem Prinzip erstellt.

---

[34] Möller (2003, 285)

[35] D. h. $Ä_{gesamt}$, $Ä_{lexikalisch}$, $Ä_{morphologisch}$, $Ä_{phonologisch}$, $Ä_{syntaktisch}$ mit jeweils 322; $Ä_{männlich}$, $Ä_{weiblich}$ mit je 244 und $Ä_{alt}$, $Ä_{jung}$ mit 239 Zeilen und Spalten

3.3 Clusteranalyse

Mit der im zweiten Schritt der dialektometrischen Verfahrenskette erfolgten Metrisierung der Informationen durch Umwandlung der nominalen Daten- in eine metrische Ähnlichkeitsmatrix, ist es nun möglich, komplexe numerisch-klassifikatorische Verfahren anzuwenden. Unter den vielen multivariaten Klassifikationsverfahren hat sich in der Dialektometrie die *Clusteranalyse* bewährt und durchgesetzt. Im Folgenden sollen zum besseren Verständnis und Nachvollziehbarkeit der Auswertungsergebnisse einige Grundzüge der *hierarchischen Clusteranalyse* ausführlicher dargestellt werden, ohne dabei allzu tief in theoretische Überlegungen vorzudringen.

Per definitionem ist unter dem Begriff „Clusteranalyse“ ein numerisches Sortierverfahren zu verstehen, das „Untersuchungseinheiten [hier: Ortspunkte], die sich voneinander in Bezug auf bestimmte Ähnlichkeits- und Distanzmaße unterscheiden, dergestalt in Klassen [...] zusammenfasst, dass die einer Klasse zugeteilten Elemente voneinander weniger differieren (also in sich homogener sind) als gegenüber Elementen [bzw. Klassen], die außerhalb der Klasse liegen“[36].

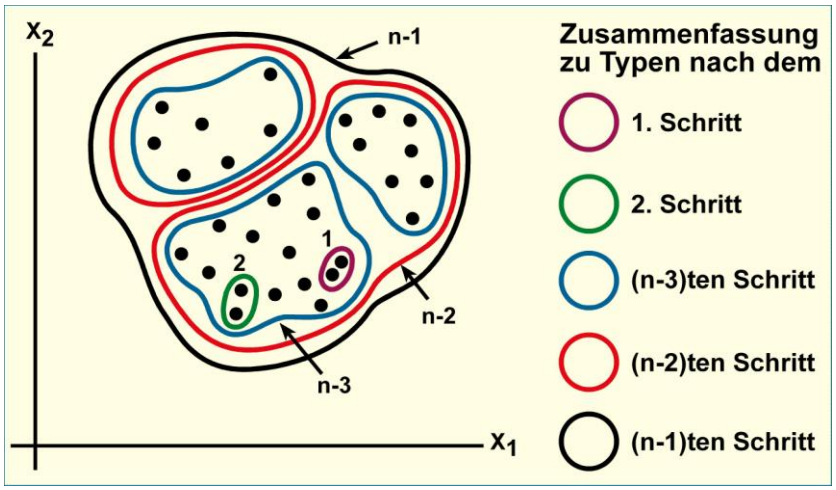


Abb. 5: Prinzip der hierarchischen Clusteranalyse (Quelle: verändert nach Bahrenberg / Giese / Nipper (2003, 284))

Die Zusammenfassung einzelner Untersuchungseinheiten zu größeren Gruppen erfolgt nach dem Prinzip der *schrittweisen Zusammenfassung* (vgl.

[36] Goebl (1982, 172)

Abb. 5). Der Prozess der Klassenbildung beginnt mit den zunächst unverbunden *N* Elementen (hier: Ortspunkte). Im ersten Klassifizierungsschritt werden nun diejenigen Elemente zusammengefasst, die gemäß dem gewählten Ähnlichkeitsmaß (hier: RIW_{jk})[37] am ähnlichsten sind, d. h. bei ihrem Zusammenschluss den geringsten Inhomogenitätswert aufweisen. Die Anzahl der für den nächsten Klassifizierungsschritt verbleibenden Elemente ist also durch die Fusionierung um eines verringert. Im nächsten Schritt werden wiederum die (*N* -1) Elemente zu einer Klasse zusammengefasst, die sich am ähnlichsten sind. Das Klassifizierungsverfahren kann so oft wiederholt werden, bis nach dem (*N* -1)ten Schritt alle Elemente in einem Cluster vereinigt sind (vgl. Abb. 5).

Sind einzelne Elemente zu einem Cluster zusammengefasst, was, wie eben erläutert wurde, bereits nach dem ersten Klassifizierungsschritt der Fall ist, besteht das Problem, welches der Elemente des Clusters herangezogen wird, um die Ähnlichkeit zu anderen Elementen zu messen. Dafür gibt es verschiedene Methoden. Zu den wichtigsten[38] gehört neben dem *Single Linkage*-Verfahren, bei dem die Distanz durch die „nächsten Nachbarn" bestimmt wird, und dem *Complete Linkage*-Verfahren („entfernteste Nachbarn") das *Average Linkage*-Verfahren. Bei diesem Verfahren dient die durchschnittliche Distanz aller Paare von Elementen als Ähnlichkeitsmaß zwischen zwei Clustern. Daneben sind das *Zentroid*-Verfahren mit Bestimmung eines

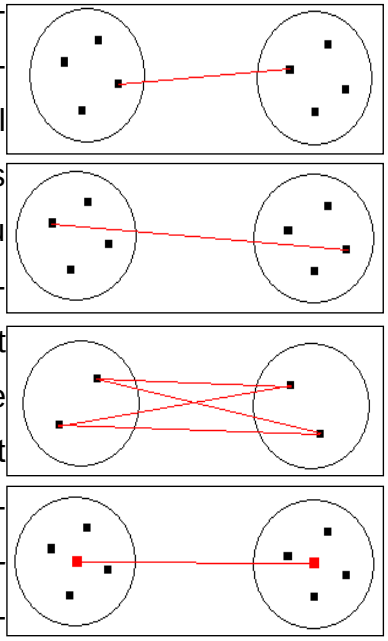


Abb. 6: *Single Linkage-*, *Complete Linkage-*, *Average Linkage-* und *Zentroid-* Verfahren

---

[37] Neben dem *Relativen Identitätswert* (RIW_{jk}) existieren jede Menge weiterer Ähnlichkeitsmaße, auf deren Basis eine Klassifizierung durchgeführt werden kann. Mit *(quadrierter) euklidischer Distanz*, *Manhatten-Distanz* und *Korrelationskoeffizient* seien an dieser Stelle einige genannt und für weitere Informationen auf Bahrenberg / Giese / Nipper (2003, 281-283) verwiesen.

[38] Vgl. Bahrenberg / Giese / Nipper (2003, 284-287).

arithmetischen Mittelzentrums (Zentroid) und das komplexere *Ward*-Verfahren wichtige Methoden. In der Dialektometrie findet vor allem das *Zentroid*-Verfahren Anwendung, das zusammen mit dem *Average Linkage*-Verfahren den Vorteil besitzt, durch den Einbezug aller Elemente eines Cluster (s. Abb. 6) zu relativ homogenen Clustergrößen zu gelangen. Dagegen neigt z. B. das *Single Linkage*-Verfahren zur *Verkettung*, d. h. zur Bildung einiger weniger sehr umfangreichen Cluster, während andere Cluster nur sehr wenige Elemente enthalten.

Nach jedem Schritt des Klassifizierungsverfahrens erhält man jeweils eine neue Lösung für das Klassifizierungsproblem. Im ersten Schritt gibt es genau so viele Klassen wie Elemente, d. h. jedes Element stellt eine eigene Klasse dar. Im letzten Schritt der Klassifizierung gehören alle Elemente einer einzigen Klasse an. Beide Ergebnisse sind für eine sinnvolle Klassifizierung unbefriedigend, sodass die „Lösung“ zwischen diesen beiden Polen gesucht werden muss. Mit jedem Schritt wächst zwar durch die Abnahme der Elemente die Übersichtlichkeit, allerdings steigt auch der Grad der Generalisierung, d. h. der Informationsverlust bedingt durch die Zusammenfassung immer unterschiedlicherer Klassen. Bei der Festlegung der Anzahl der Klassen muss also ein Kompromiss zwischen einem genügend großen Grad an Generalisierung zur Gewährung einer Übersichtlichkeit und einem nicht zu hohen Grad an Generalisierung gefunden werden, um den Informationsverlust zu vermindern.[39]

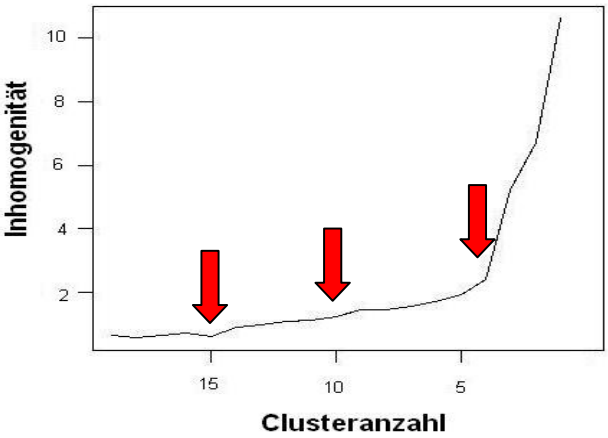


Abb. 7: Das „Ellbogen“-Kriterium

[39] Vgl. Bahrenberg / Giese / Nipper (2003, 287).

Doch wie legt man in der Praxis tatsächlich die Anzahl der Klassen für ein „optimales“ Klassifizierungsergebnis fest? Für dieses schwierige Problem gibt es einerseits ein statistisches Hilfsinstrument, das „Ellbogen“-Kriterium (s. Abb. 7): Trägt man die Inhomogenität (= Fehlerquadratsumme) innerhalb der Cluster in ein Diagramm ein, so erhöht sich diese logischerweise mit jedem Schritt des Fusionierungsprozesses. In bestimmten Schritten steigt sie besonders stark, wenn sehr unähnliche Cluster zusammengefasst werden, was vor allem gegen Ende der Fusionierung auftritt. Das „Ellbogen“-Kriterium besagt, dass vor genau solchen Schritten das weitere Clustern abgebrochen werden soll. Da dieses Kriterium nicht eindeutig ist und theoretisch mehrere Lösungen möglich sind (s. rote Pfeile in Abb. 7), stellt es immer nur ein Hilfsinstrument zur groben Orientierung dar, das letztendlich nur in Kombination mit dem logischen Sachverstand des Bearbeiters einzusetzen ist. Die Clusteranzahl kann also nicht nur auf statistischem Weg bestimmt werden, sondern ist immer auch abhängig vom Untersuchungsgegenstand und der jeweiligen Fragestellung. Deshalb kommt der Bearbeiter nicht umhin, sich das Ergebnis mehrerer Lösungen anzuschauen und auf deren Grundlage eine (auch mehr oder weniger subjektive) Entscheidung über die Clusteranzahl zu treffen.

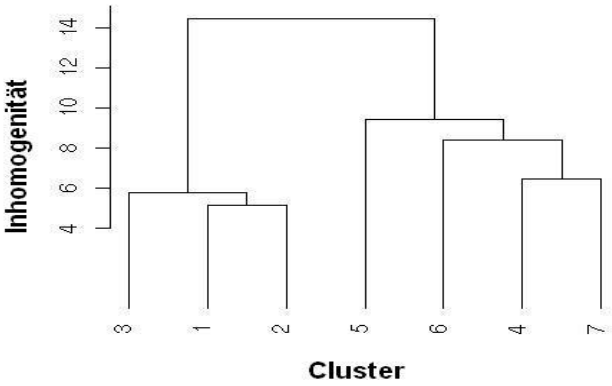


Abb. 8: Das Dendrogramm

Ein Ausweg aus dieser Problemlage bietet das so genannte *Dendrogramm* (vgl. Abb. 8), das eine große Verbreitung in der Dialektometrie

besitzt[40]. Es veranschaulicht den Prozess der Klassenbildung, indem es angibt, welche Cluster bzw. Elemente beim Fusionierungsprozess miteinander zu einem verschmolzen werden. Aus ihm lässt sich für jeden Schritt des Klassifizierungsverfahrens der weitere Fortgang bis zum letzten Schritt ablesen.

Diese Tatsachen wurden bei der Analyse der Daten des „Atlas zur deutschen Alltagssprache“ berücksichtigt. Auf fünf ($Ä_{gesamt}$, $Ä_{lexikalisch}$, $Ä_{morphologisch}$, $Ä_{phonologisch}$, $Ä_{syntaktisch}$) der im zweiten Schritt der dialektometrischen Verfahrenkette erzeugten Ähnlichkeitsmatrizen wurde die hierarchische Clusteranalyse angewendet.[41] Nach der Bestimmung der Clusteranzahl gemäß der bereits genannten Kriterien wurden im weiteren Vorgehen für die jeweiligen ermittelten Clusteranzahlen Dendrogramme erstellt, aus denen der weitere Klassifizierungsprozess abzulesen ist (s. Kapitel 4.3). Die Darstellung der Ergebnisse erfolgte in Anlehnung an den „Dialektometrischen Atlas von Südwest-Baden“ (DASB) von SCHILTZ[42], der die beiden Elemente, Karte mit „optimaler“ Clusteranzahl und Dendrogramm, vereinigt.

Zu Beginn des Kapitels wurde erwähnt, dass das Thema *Clusteranalyse* deswegen ausführlicher behandelt wird, um eine Nachvollziehbarkeit der Ergebnisse auch für einen statistischen Laien zu ermöglichen. Allerdings erfüllt die intensivere Beschäftigung mit der Thematik einen weiteren Zweck, der erwähnt werden muss und sich aus dem bereits Genannten ergibt. Sowohl bei der Wahl des Distanzmaßes als auch bei der Wahl der Methode der Clusterbildung bietet sich ein Potpourri an Vorgehensweisen. Die große Vielfalt klassifikatorischer Methoden und Resultate macht die Bestimmung einer einzigen „richtigen“ Lösung unmöglich:

> „Leider wird sehr oft übersehen, dass Klassifikationen zielgerichtete Heuristika sind und daher stets im Dienst eines bestimmten Forschungsziels bzw. For-

---

[40] Siehe z. B. Schiltz (1996, „Der Dialektometrische Atlas von Südwest-Baden (DASB)“), Videsott (2003) oder Goebl (1982).

[41] Auf Grundlage eines *R-Skripts* (vgl. http://cran.r-project.org/).

[42] Vgl. Schiltz (1996, Teil II-IV: Kartenband).

schungstheorie stehen. Daraus ergibt sich zwangsläufig, dass ein gegebener Datensatz in der verschiedensten Weise klassifiziert werden kann. Der oft leichtfertig ausgesprochene Wunsch nach der „allein richtigen“ Klassifikation muss Schimäre bleiben.“[43]

[43] Goebl (1982, 55)

Kapitel 4

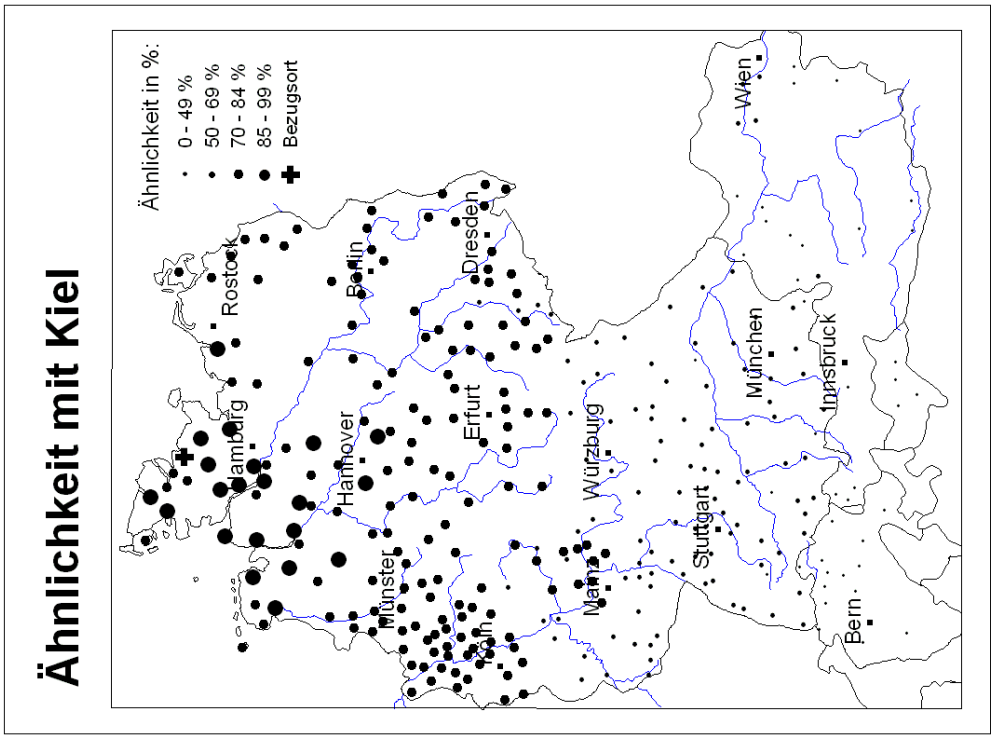
Auswertung

Ein Hauptaugenmerk der vorliegenden Arbeit liegt auf der dialektometrischen Analyse der Daten des „Atlas zur deutschen Alltagssprache“, dennoch darf die Auswertung der erzeugten Ergebnisse nicht außer Acht gelassen werden. Die Dialektometrie hat ein breit gefächertes Anwendungspotential in der sprachwissenschaftlichen Forschung, aus dem exemplarisch einige Problemstellungen herausgegriffen wurden. Anhand dieser Beispiele soll letztendlich die zentrale Frage der dialektometrischen Analyse des AdA geklärt werden: Liefert eine statistische Auswertung unsystematisch erhobener Sprachdaten plausible Ergebnisse?

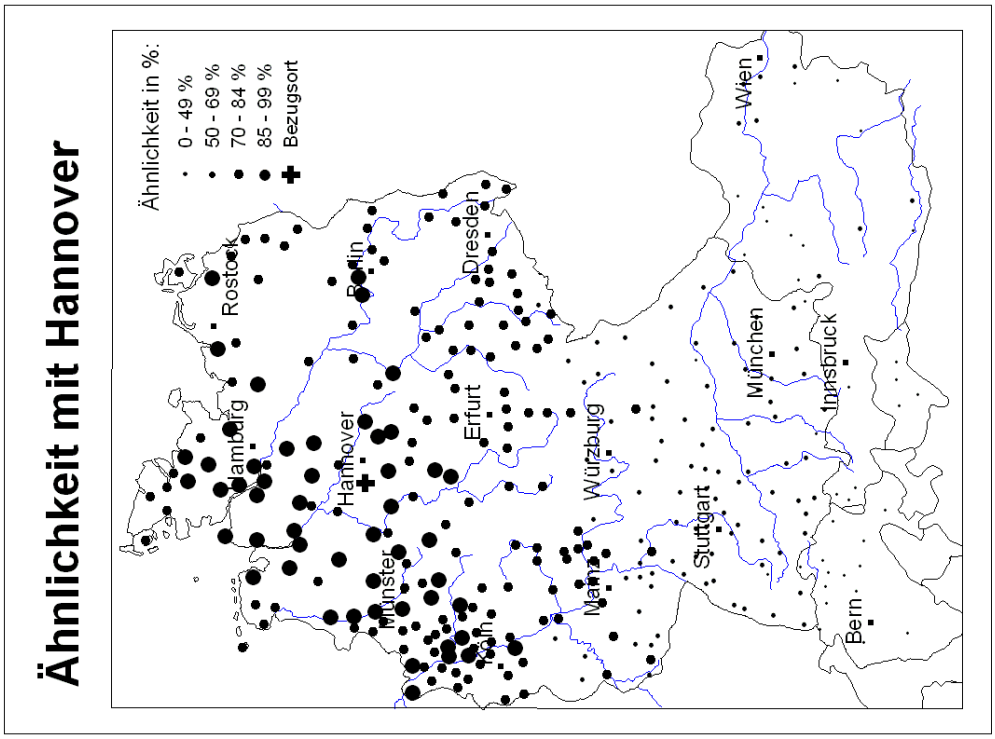
4.1 Wortgebrauchsräume

In Anlehnung an Möller[44] lassen sich anhand der erstellten Ähnlichkeitskarten (auf Grundlage von $Ä_{gesamt}$) unterschiedliche *Wortgebrauchsräume* differenzieren, die Hinweise zur diatopischen Gliederung des Deutschen liefern. Dabei muss allerdings die bereits erwähnte Einschränkung genannt werden, dass die Karten stark vom jeweiligen Bezugspunkt abhängen (s. Kapitel 3.2).

[44] Vgl. Möller (2003, 273-285).



Karte 1: Ähnlichkeitskarte „Kiel“



Karte 2: Ähnlichkeitskarte „Hannover“

Karte 1 und 2 zeigen einen großen homogenen Raum alltagssprachlichen Wortgebrauchs in Norddeutschland. Für Kiel (s. Karte 1) umfasst er das gesamte Bundesland Schleswig-Holstein und reicht in einem Keil bis nach Hannover mit Ähnlichkeitswerten von über 85 %. Die zweite Karte mit Hannover als Referenzpunkt bestätigt die erste und weist sogar eine noch größere Fläche mit über 85 % Ähnlichkeit auf. Sie beinhaltet neben Schleswig-Holstein (mit Ausnahme der nördlichsten Ortspunkte) fast komplett Niedersachsen und teilweise Nordrhein-Westfalen. Zudem zeigen die ostdeutschen Gebiete in beiden Karten einen hohen Bezug zu den norddeutschen Bezugspunkten auf, sodass von einem wort(gebrauchs)geographisch zusammengehörigen Gebiet gesprochen werden kann.

Beide Karten weisen eine deutliche Nord-Süd-Teilung des Gesamtgebiets auf. Die Grenze, die die beiden Gebiete voneinander trennt, verläuft bei beiden Karten auf Höhe des Mains – auf der Karte mit Bezugspunkt Kiel nur leicht nach Norden verschoben. Eine markante Ausbuchtung nach Norden erfährt sie in Rheinland-Pfalz, wo die Orte mit einem Ähnlichkeitswert von unter 70 % bis fast nach Köln reichen. Denkt man sich die Topographie hinter das Kartenbild, kann man erkennen, dass die Eifel hier eine natürliche Grenze bildet (s. andeutungsweise Abb. 9). Das gesamte Gebiet südlich des Mains ist durch Orte gekennzeichnet, die einen Ähnlichkeitswert zwischen 50 und 69 % zu Kiel und Hannover aufweisen. Ob es sich um ein homogenes Gebiet ähnlich dem in Norddeutschland handelt, kann aus den beiden Karten aufgrund bereits genannter Einschränkungen allerdings nicht geschlossen werden. In beiden Karten ist zudem eine weitere Nord-Süd-Teilung erkennbar, die entlang der südlichen deutschen Staatsgrenze verläuft. So weisen fast alle Orte in Österreich und der Schweiz einen Ähnlichkeitswert von unter 50 % zu Kiel bzw. Hannover auf.

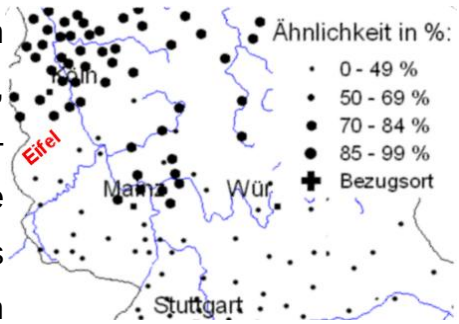
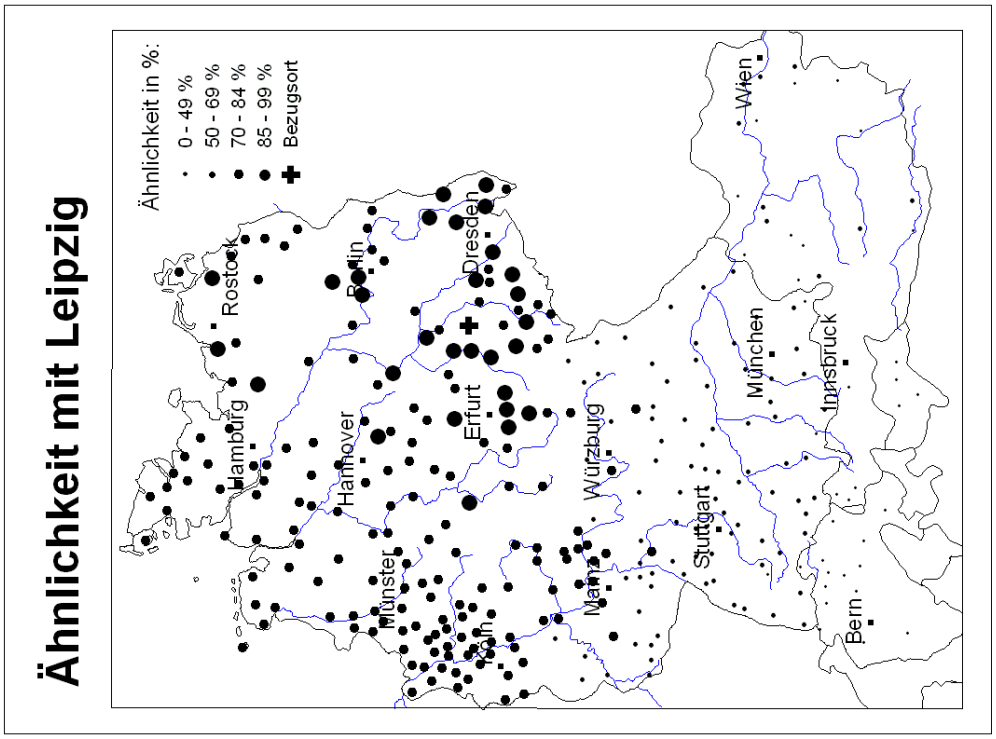
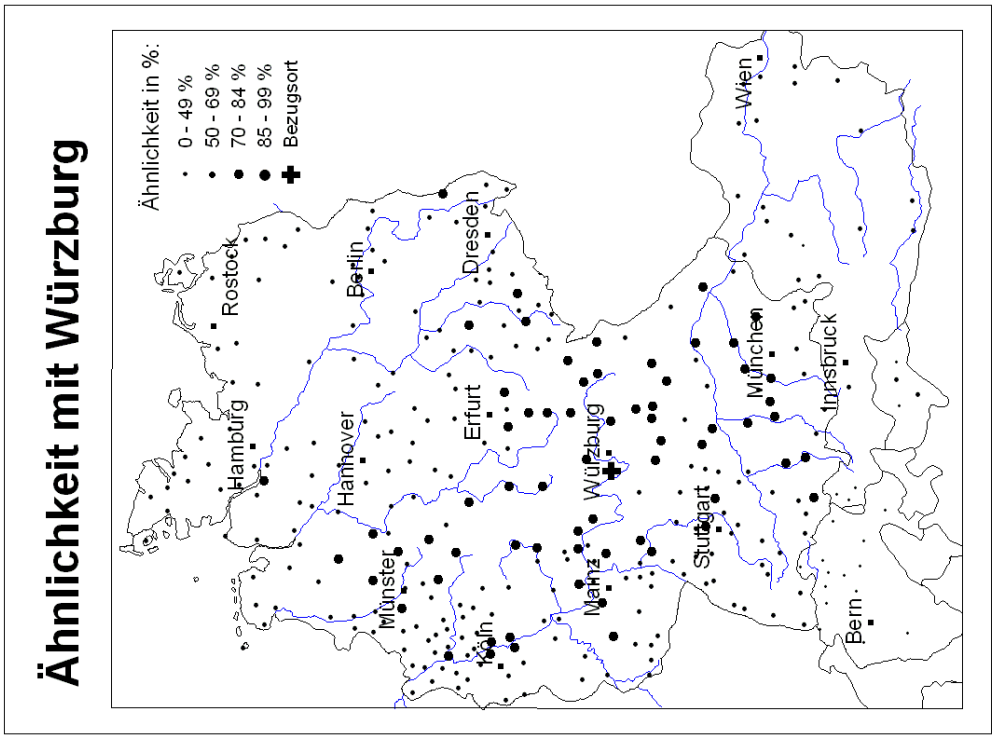


Abb. 9: Die Eifel als Grenzgebiet



Karte 3: Ähnlichkeitskarte „Leipzig“

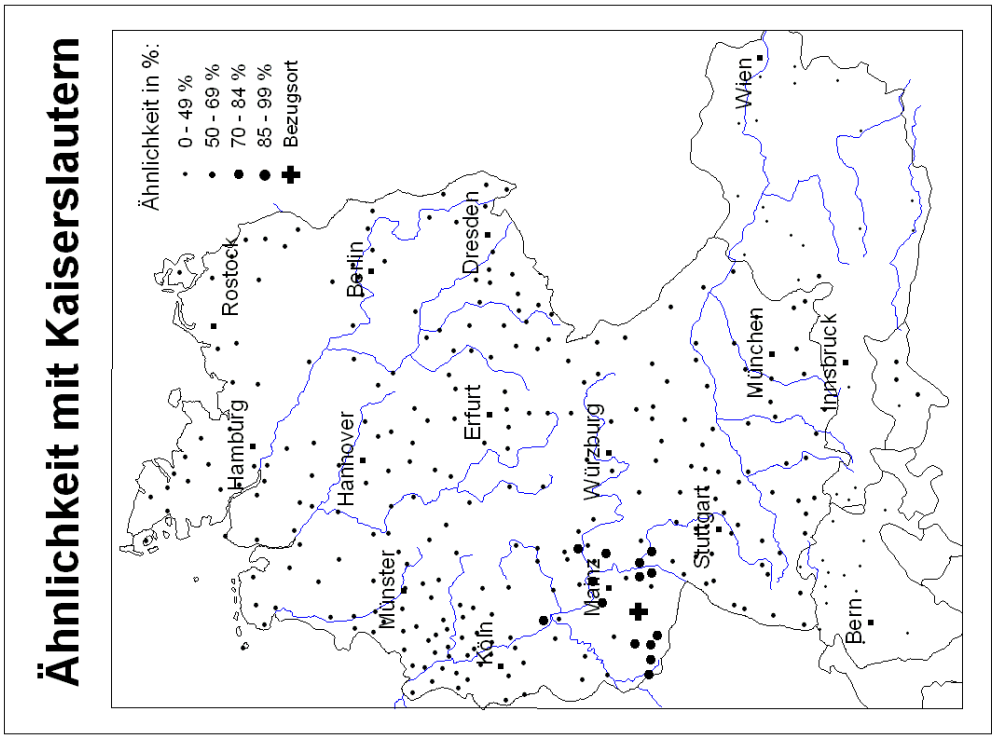


Karte 4: Ähnlichkeitskarte „Würzburg“

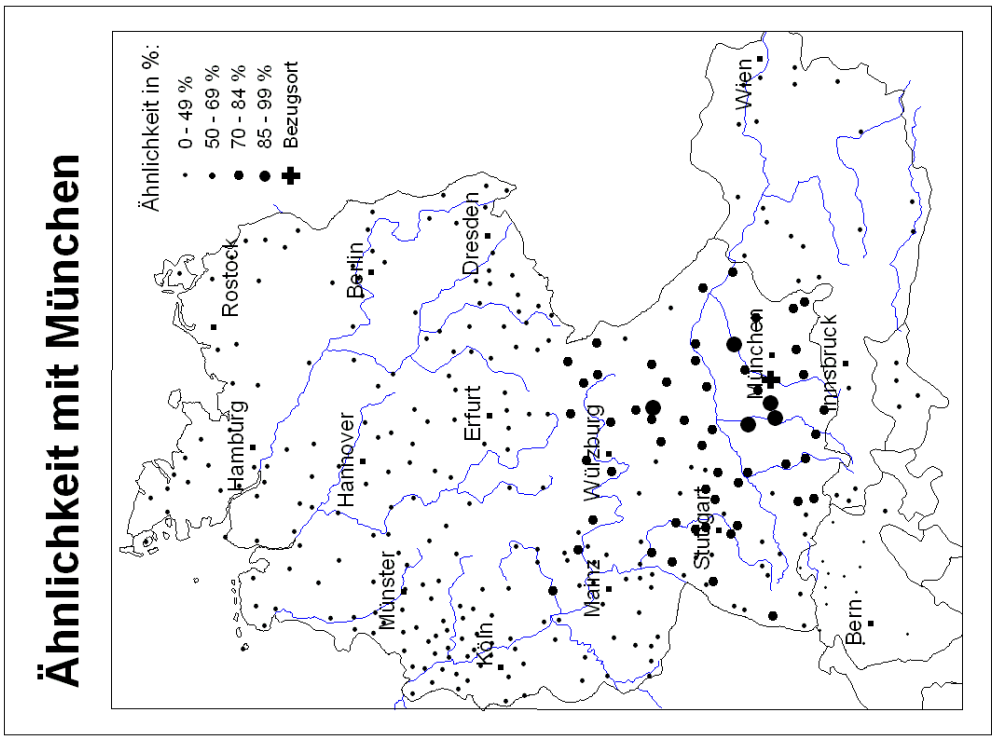
Auch andere hier nicht dargestellte Karten (z. B. von Bremen, Hamburg, Berlin, Wuppertal) bestätigen diese Nord-Süd-Teilung mit zwei markanten Sprachgrenzen und dem homogenen norddeutschen Gebiet. Dasselbe zeigt auch die Karte für Leipzig als Repräsentant des östlichen Deutschlands (s. Karte 3). Fast alle Orte nördlich des Mains zeigen Ähnlichkeitswerte über 70 % in Bezug zu Leipzig, während die Übereinstimmung mit dem Süden fast durchgehend niedriger ist. Das norddeutsche Sprachgebiet weist wiederum eine hohe Homogenität auf, dennoch lässt sich eine Differenzierung vornehmen. Die Zone der höchsten Übereinstimmung ($\geq$ 85) befindet sich im näheren Umfeld zu Leipzig und reicht im Norden bis an die Ostsee. Die Ausdehnung dieser Zone nach Westen ist eingeschränkt, sodass insgesamt eine West-Ost-Teilung des norddeutschen Sprachraums erkennbar ist. Die Grenze, die die beiden Räume voneinander trennt, ist zwar nicht so eindeutig wie diejenige entlang des Mains, man kann aber eine tendenzielle Orientierung an der ehemaligen innerdeutschen Grenze feststellen.

Ein Grenzgänger stellt in dieser Hinsicht der Ortspunkt Würzburg dar (s. Karte 4), der sich genau auf jener beschriebenen Sprachgrenze entlang des Mains befindet. Diese Tatsache spiegelt auch das Kartenbild wider. Es findet sich kein Ort, der eine über 85 %-ige Übereinstimmung zu Würzburg besitzt, allerdings weisen viele Orte zu beiden Seiten der Grenze Ähnlichkeitswerte zwischen 70 und 84 % auf. Die Orte mit dem größten Bezug befinden sich um Würzburg, in ganz Bayern und in einem Keil bis nach Niedersachsen ragend. Insgesamt entsteht bei dem Ortspunkt somit der Eindruck der fehlenden Zugehörigkeit zu einem Sprachgebiet bzw. die Vermutung, dass sich Würzburg in einem Übergangsbereich zwischen einem norddeutschen und einem süddeutschen Wortgebrauchsraum befindet.

Untersucht man nun das Gebiet südlich der aus den bisherigen Karten vermuteten Sprachgrenze entlang des Mains, so fällt auf, dass eine dem norddeutschen Sprachgebiet vergleichbare Homogenität fehlt und sich mehrere Unterteilungen vornehmen lassen. Eine erste ist aus dem Kartenbild mit dem Bezugspunkt Kaiserslautern erkennbar (s. Karte 5). Einerseits ist die fehlende Homogenität dadurch ersichtlich, dass im Gegensatz zum norddeutschen Wort-



Karte 5: Ähnlichkeitskarte „Kaiserslautern“



Karte 6: Ähnlichkeitskarte „München“

gebrauchsraum kein Ort einen Ähnlichkeitswert von über 85 % gegenüber Kaiserslautern besitzt, andererseits beschränkt sich der Raum mit dem größten Bezug zu Kaiserslautern auf das pfälzische Sprachgebiet und ist damit recht klein. Zu den übrigen deutschen Sprachlandschaften, sogar zu dem räumlich nicht weit entfernten ripuarischen Dialektraum, ist eine deutliche Abgrenzung erkennbar. Verstärkt wird diese wiederum an der deutschen Staatsgrenze, ab der die weiter südlich gelegenen Orte lediglich Ähnlichkeitswerte von unter 50 % aufweisen.

Eine weitere Differenzierung des Wortgebrauchsraums südlich des Mains ist aus der München-Karte zu entnehmen (s. Karte 6). Die Grenze zum homogenen norddeutschen Sprachraum befindet sich auch hier entlang des Mains. Allerdings ist im Gegensatz zu den Karten mit Bezugspunkt Hannover oder Kiel keine Ausbuchtung des Grenzverlaufs nach Norden bis zur Eifel erkennbar, da dieses Gebiet, wie die Karte Kaiserslautern gezeigt hat, einen eigenen Wortgebrauchsraum darstellt. In Karte 6 sind die Orte der höchsten Übereinstimmungskategorie, mit Ausnahme des Ortspunkts Nürnberg, rund um München gruppiert und die der zweithöchsten Kategorie im gesamten süddeutschen Gebiet zu finden, ohne eine klare innere Differenzierung zu zeigen. Mehrere Orte weisen hier Ähnlichkeitswerte von 70-84 % oder 50-69 % auf, wobei sich die Orte mit geringerer Übereinstimmung tendenziell im Übergangsbereich zu anderen Dialekträumen (z. B. zum Alemannischen) befinden. Eine Abgrenzung zeigt sich wiederum im Süden zum einen zu Österreich (alle Orte mit Ähnlichkeitswerten zwischen 50 und 69 %) und zum anderen zur Schweiz (alle Orte mit einer Ausnahme 0-49 % Ähnlichkeit).

Vergleicht man die München-Karte mit denen der anderen bayerischen Großstädte Nürnberg und Augsburg als Bezugspunkte, so fällt jeweils eine hohe Übereinstimmung der drei Ortspunkte auf. Was schon auf Karte 6 mit der überra-

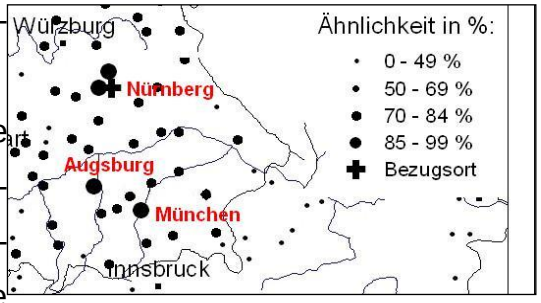


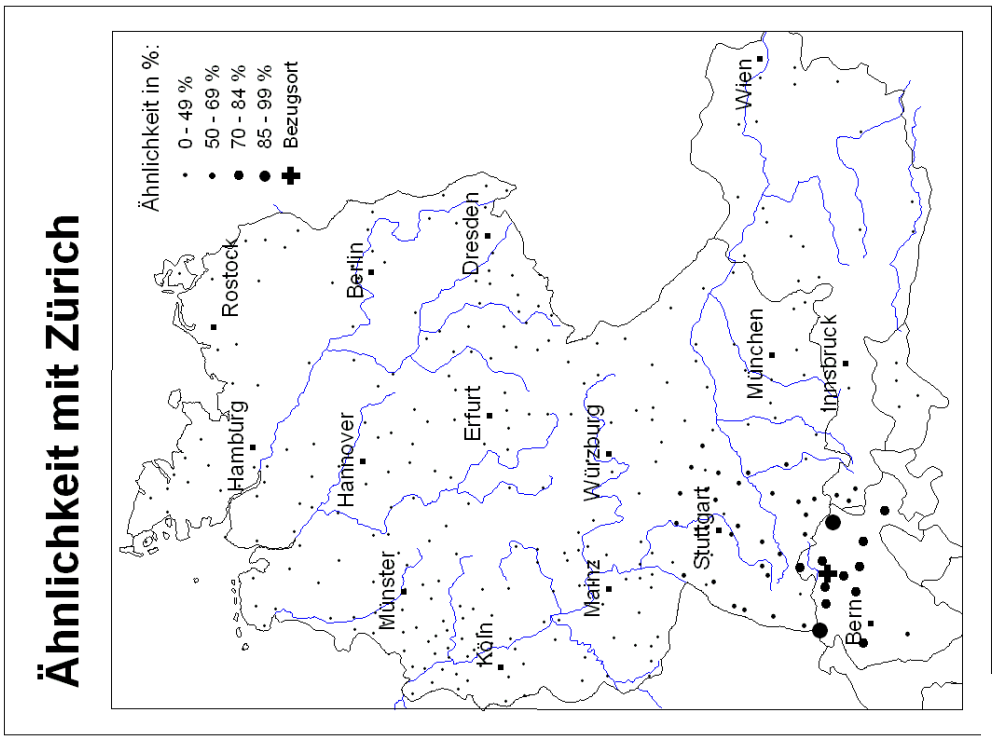
Abb. 10: Ähnlichkeit der bayerischen Großstädte

schend hohen Übereinstimmung des räumlich weiter entfernten Nürnberg in Bezug zu München erkennbar ist, fällt vor allem bei der Karte mit Nürnberg als Bezugsort auf (s. Abb. 10). Neben zwei direkt um Nürnberg gelegenen Ortspunkten weisen lediglich München und Augsburg Werte der höchsten Übereinstimmungskategorie auf. Auf dieser Grundlage läst sich also die Aussage formulieren, dass die drei Großstädte Bayerns durch einen sehr ähnlichen alltagssprachlichen Wortgebrauch gekennzeichnet sind. Diese Ähnlichkeit unter den drei Großstädten ist teilweise sogar höher als diejenige zu unmittelbar benachbarten Ortspunkten.

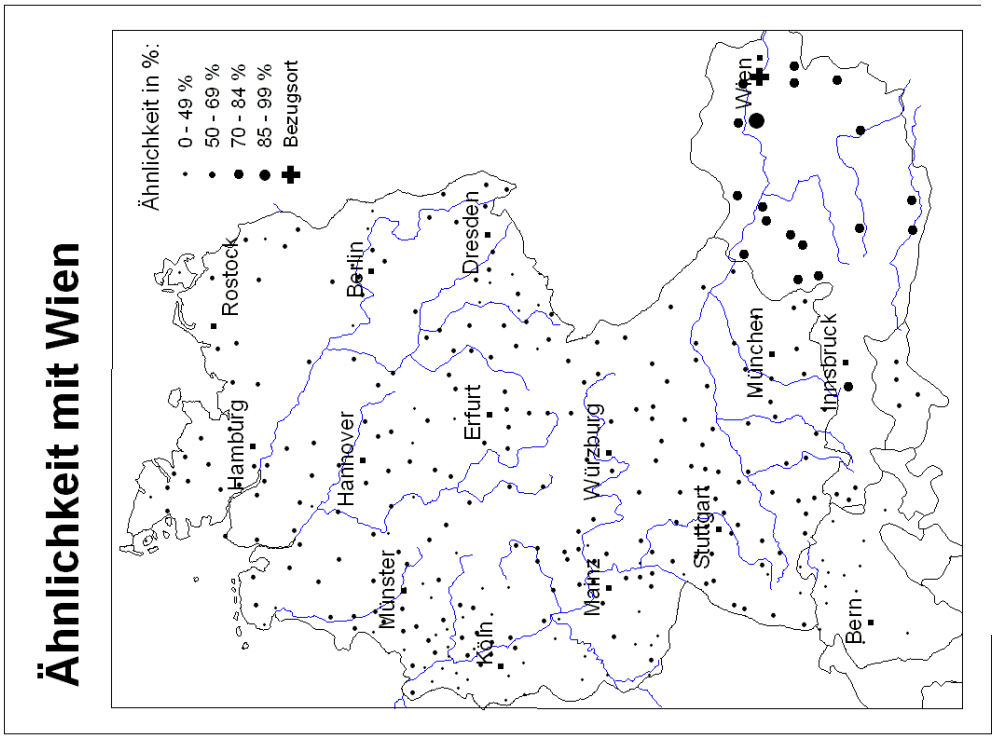
Die nächsten beiden Karten mit Wien bzw. Zürich als Bezugsorte geben Aufschluss über das Sprachgebiet, das auf den bisherigen Karten jeweils mehr oder weniger scharf abgegrenzt wurde. Leider sind aufgrund der Datenlage nur wenige Ortspunkte für diesen Raum vorhanden, dennoch lassen sich eindeutige Tendenzen ablesen. Auf der Wien-Karte (s. Abb. 7) erkennt man ein relativ großes homogenes Sprachgebiet, das fast ganz Österreich einnimmt. Zwar ist die Homogenität nicht so stark ausgeprägt wie die in Norddeutschland, dennoch erreichen die Ortspunkte Ähnlichkeitswerte von über 70 %. Dieses Gebiet erstreckt sich von Wien bis nach Innsbruck, wobei lediglich die Orte in Vorarlberg eine geringere Übereinstimmung zu dem Bezugspunkt zeigen. Eine Abgrenzung des Sprachgebiets erfolgt nach Norden wie nach Süden entlang der Staatsgrenze zu Deutschland bzw. Italien. Das deutschsprachige Südtirol wird demnach als eigenständiger Wortgebrauchsraum abgegrenzt. Zudem ist eine scharfe Grenze zur schweizerdeutschen Sprachlandschaft erkennbar. Die Ortspunkte weisen hier lediglich Ähnlichkeitswerte von unter 50 % auf.

Dass das schweizerdeutsche Sprachgebiet ein deutlich abgegrenztes und eigenständiges ist, hat sich nicht nur auf den bisher besprochenen Karten angedeutet, sondern zeigt sich auch in der Karte mit Bezugsort Zürich (s. Karte 8). Rund um Zürich ist ein kleines aber relativ homogenes Sprachgebiet erkennbar, das Ähnlichkeitswerte über 70 % (mit Ausnahme des Ortspunkts Frutigen) aufweist. Die sehr scharfe Abgrenzung zum restlichen deutschen Sprachgebiet erfolgt entlang der politischen Grenze der Schweiz. Fast alle anderen Ortspunkte weisen zum Bezugsort Zürich (und auch zu anderen hier

nicht dargestellten schweizerischen Orten) Ähnlichkeitswerte unter 50 % auf. Der geringste Wert liegt sogar bei niedrigen 32 %, was die Ausnahmestellung



Karte 8: Ähnlichkeitskarte „Zürich“



Karte 7: Ähnlichkeitskarte „Wien“

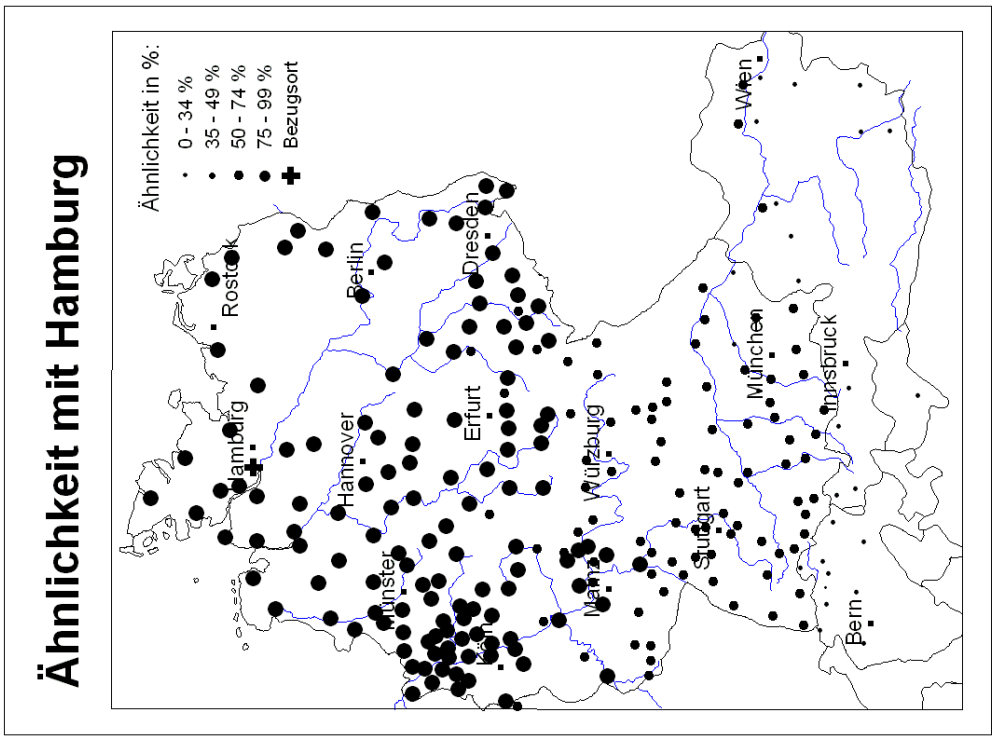
des schweizerdeutschen Wortgebrauchsraums unterstreicht. Lediglich zu einigen Orten im südlichen Baden-Württemberg und in Vorarlberg ist ein stärkerer Bezug erkennbar. Vorarlberg kann somit, da es sowohl eine geringe Übereinstimmung zu Wien und zu Zürich aufweist als eingeständiges Sprachgebiet im Übergangsbereich der beiden homogenen Wortgebrauchsräume angesehen werden.

Fasst man die Ergebnisse der im Detail besprochenen einzelnen Ähnlichkeitskarten zusammen, kommt man zu folgenden Schlussfolgerungen. Das deutsche Sprachgebiet ist nicht einheitlich, sondern weist mehrere unterschiedlich große und unterschiedlich homogene Sprachlandschaften auf, die durch Isoglossen voneinander getrennt sind. Ein großes und sehr homogenes norddeutsches Sprachgebiet, das eine geringe innere Differenzierung besitzt (z. B. ostdeutsche Sprachlandschaft) wird durch eine scharfe Nord-Süd-Grenze entlang des Mains bzw. der Eifel von einem süddeutschen abgegrenzt. Dieses besitzt im Gegensatz zu Norddeutschland eine stärkere Inhomogenität, wobei die Grenzen teilweise deutlich (z. B. Abgrenzung des Pfälzischen) zu erkennen sind, teilweise aber nur erahnt werden können (z. B. Abgrenzung bayerisches, schwäbisches und nieder- / oberalemannisches Sprachgebiet). Durch eine weitere Nord-Süd-Grenze entlang der deutschen Staatsgrenze werden zudem zwei deutlich voneinander getrennte Sprachlandschaften abgegrenzt. Neben dem österreichischen Sprachgebiet, das den politischen Grenzen folgt, nimmt vor allem das überraschend homogene, aus vielen Untermundarten bestehende schweizerdeutsche Sprachgebiet wegen seiner Eigenständigkeit eine Sonderstellung in der deutschen Sprachlandschaft ein. Zu erwähnen sind weiterhin Vorarlberg als Übergangsbereich und Südtirol als eigenständiges Sprachgebiet in Italien.

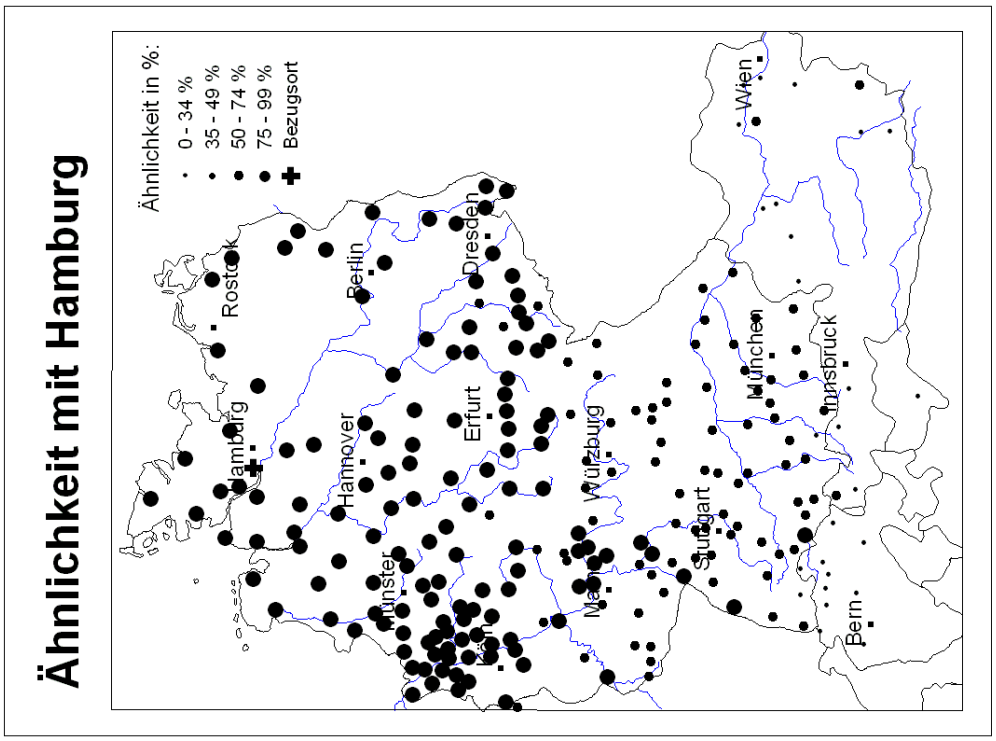
4.2 Merkmalsdifferenzierte Wortgebrauchsräume

Wie bereits erwähnt wurde, besteht der Vorteil der Online-Erhebung des AdA in der Möglichkeit, den Datensatz nach bestimmten Merkmalen zu differenzieren. Im Folgenden werden die geschlechtsspezifischen und altersspezifischen Ergebnisse im jeweiligen Vergleich vorgestellt.

Zu beachten ist dabei, dass die Stufen des Ähnlichkeitswerts anders angelegt wurden (s. die folgenden Karten). Während bei den Karten in Kapitel 4.1 die Intention in der Herausarbeitung ähnlicher Gebiete zum Bezugsort lag, sollen nun Unterschiede auch im unteren Ähnlichkeitsbereich aufgezeigt werden. Deshalb sind die Stufen der Karten in Kapitel 4.2 im unteren Bereich der Skala feiner differenziert. An diesem Beispiel kann die in Kapitel 3.3 theoretisch beschriebene methodische Vielfalt an einem praktischen Beispiel erläutert werden. Hamburg (Karte 9 und 10), das wie Hannover (Karte 2) und Kiel (Karte 1) zum homogenen norddeutschen Wortgebrauchsraum gehört, weist im Kartenbild Unterschiede zu den beiden anderen Karten auf (z. B. keine innere Differenzierung des norddeutschen Sprachgebiets). Die Ursache liegt in der soeben genannten Änderung der Ähnlichkeitsstufen. Aber trotz dieses methodischen Eingriffs, der eine leichte Variation im Kartenbild verursacht, sind die grundlegenden Aussagen wie z. B. die Nord-Süd-Teilung des deutschen Sprachgebiets mit den Grenzen entlang des Mains und der deutschen Staatsgrenze weiterhin enthalten. Der mitunter gefährlichen methodischen Vielfalt (in diesem Fall bezüglich der Klassendefinition), die bei falscher Anwendung zur Beliebigkeit ausarten kann, steht die statistische Aussagekraft gegenüber. Wendet der Bearbeiter also die methodischen Eingriffe in Anhängigkeit von der Fragestellung sinnvoll an, kann die Dialektometrie zum Erkenntnisgewinn in der Dialektologie beitragen.



Karte 10: Ähnlichkeitskarte „Hamburg“ (weiblich)



Karte 9: Ähnlichkeitskarte „Hamburg“ (männlich)

4.2.1 Geschlecht

Werden nach Geschlechtern getrennte Ähnlichkeitsmatrizen erstellt und diese Werte mittels einer Karte visualisiert, lautet die zentrale Aussage, dass keine bzw. kaum Unterschiede im Kartenbild zu erkennen sind. Exemplifiziert dargestellt wird dies an den Karten 10 (weiblich) und 9 (männlich) mit Bezugsort Hamburg. Beide Karten weisen ein sehr ähnliches Kartenbild mit nur geringen, unsystematischen Unterschieden auf. Dieselbe Aussage kann auch für andere Orte wie z. B. Frankfurt/Main, Basel oder Stuttgart getroffen werden, die hier nicht gesondert dargestellt werden. Aus den Daten des Ada können also keine geschlechtsspezifischen, diatopisch relevanten Unterschiede hinsichtlich des Wortgebrauchs festgestellt werden.[45]

4.2.2 Alter

Die deutsche Alltagssprache ist – wie bereits erwähnt – durch einen ständigen Wandel geprägt. Die ursprüngliche Intention des „Atlas zur deutschen Alltagssprache“ liegt in dem Versuch, diesen Wandel im Vergleich mit EICHHOFFs „Wortatlas der deutschen Umgangssprachen“, dessen erster Band 1977 erschien, aufzuzeigen.[46] Durch den Vergleich der Kartenbilder für eine ausgewählte Frage zwischen AdA und WDU kann deshalb die Aussage getroffen, inwiefern sich der Wortgebrauch innerhalb von 30 Jahren, also ungefähr einer Generation, verändert hat und ob sich das Geltungsareal einer bestimmten Variante vergrößert oder verkleinert hat. Mit diesem auf einzelne Fragen und Karten beschränkten Vorgehen können allgemeine Tendenzen, wie z. B. die *Konvergenz*, d. h. der Rückgang kleinräumigerer Varianten zu Gunsten großräumi-

---

[45] Natürlich darf daraus nicht die Schlussfolgerung gezogen werden, dass es hinsichtlich des Wort- und Sprachgebrauchs keine Unterschiede zwischen Frauen und Männern gibt, was allein schon die Vielzahl an Publikationen in diese Richtung verbietet. Die Aussagen beziehen sich auf diatopische Unterschiede des Wortgebrauchs und beruhen auf Daten eines theoretisch erstellten Fragebogens. Wie Sprache in der konkreten Situation angewendet wird, hängt von vielen Faktoren wie der spezifischen Situation, der sozialen Rolle oder natürlich dem Geschlecht ab.

[46] Vgl. Elspaß (2005).

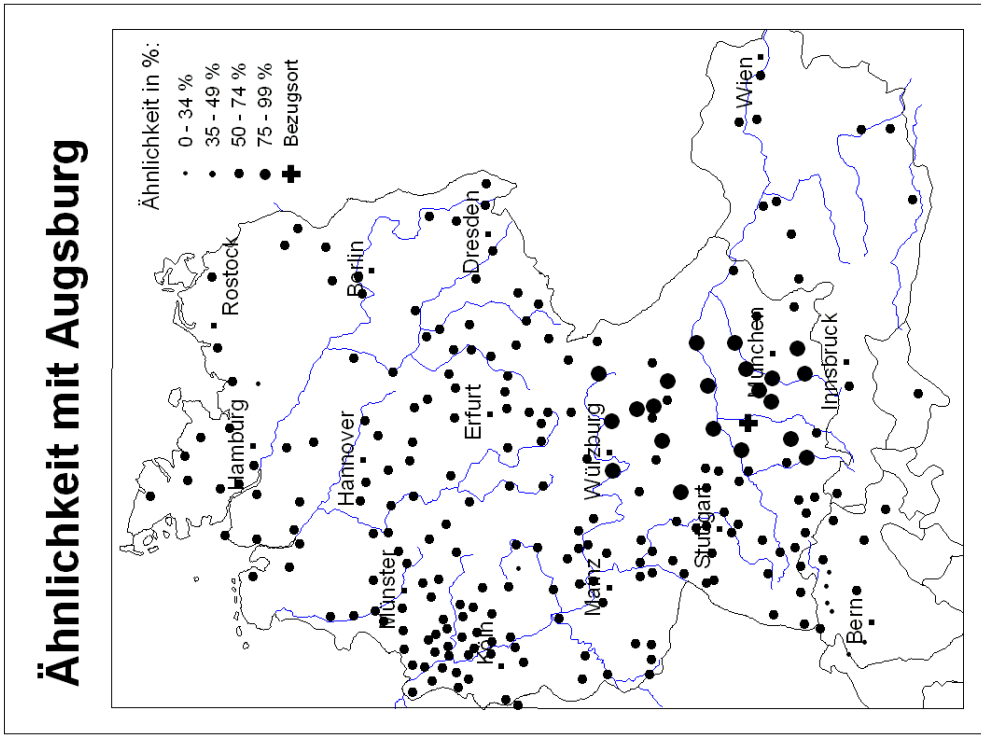
ger, aufgezeigt werden.[47] Allerdings ist eine räumliche Beschreibung dieser Tendenzen, also in welche Richtung sich der Wortgebrauchwandel für ein gesamtes Sprachgebiet (z. B. das Schweizerdeutsche) vollzieht, nur für einzelne Karten, nicht aber in einer Gesamtschau möglich. Hier wäre es aber von großem Interesse, die hinter den einzelnen Karten zugrunde liegenden Strukturen und Muster des Wandels zu kennen.

Darauf können die Daten des „Atlas zur deutschen Alltagssprache“ Auskunft geben. Durch die Differenzierung des Datensatzes in zwei unterschiedliche Altersgruppen (D_{alt}, D_{jung}) kann der Stand für zwei unterschiedliche Generationen angegeben werden. Bezogen auf die daraus erzeugten altersdifferenzierten Ähnlichkeitskarten bedeutet es, dass auf einen Wandel in der regionalen Umgangssprache geschlossen werden kann, wenn sich die Ähnlichkeitswerte eine Bezugsorts zu anderen Orten signifikant zwischen den Generationen verändern. Erhöhen sich die Werte eines regional geprägten Bezugsorts v. a. zu weiter entfernten, der Standardsprache nahe stehenden Vergleichsorten, deutet dies auf eine Entregionalisierung der Umgangssprache durch den Abbau regionaler Varianten hin. Erniedrigen sich die Ähnlichkeitswerte im umgekehrten Fall, würde dies hingegen für eine zunehmende Regionalisierung der Umgangssprache sprechen.

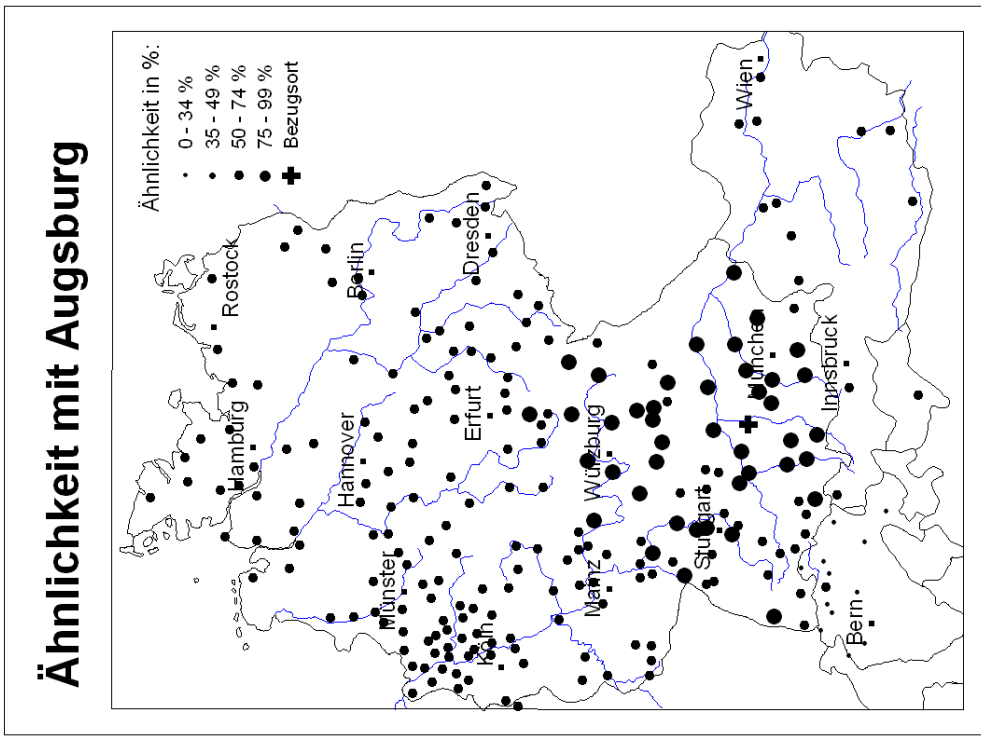
Die altersspezifischen Karten mit Bezugspunkt Augsburg deuten eine erste Tendenz an. Karte 11 mit den Antworten der über 30-jährigen zeigt das gewohnte Bild: Rund um Augsburg befinden sich die Orte mit den höchsten Ähnlichkeitswerten und die gewohnten Grenzverläufe sind zu sehen. Im Vergleich dazu ist im Kartenbild der jüngeren Generation (s. Karte 12) eine Ausdehnung des Gebiets der höchsten Übereinstimmungskategorie erkennbar. Neben mehreren Orten nördlich des Mains weisen jetzt auch Orte des schwäbischen und alemannischen Sprachraums hohe Ähnlichkeitswerte auf. Für Augsburg besteht also die Tendenz der Entregionalisierung vor allem in Bezug auf den süddeutschen Sprachraum, wohingegen die Sprachgrenze entlang des Mains weiterhin besteht, wenn auch leicht nach Norden verschoben.

---

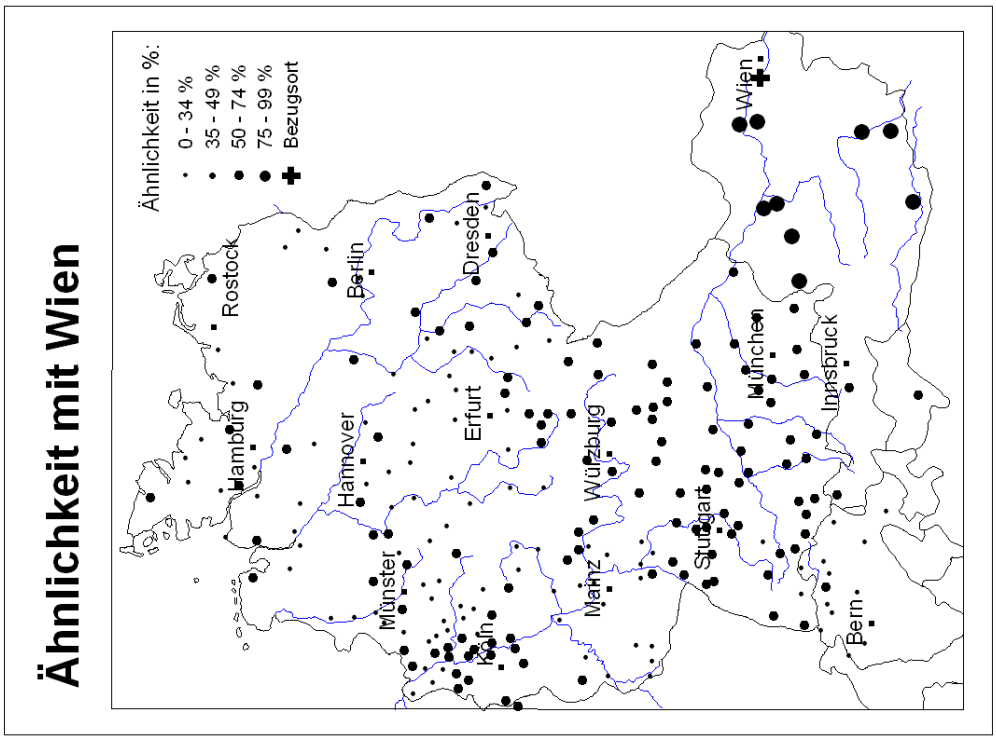
[47] Für weitere Informationen s. Elspaß (2005, 6 ff.).



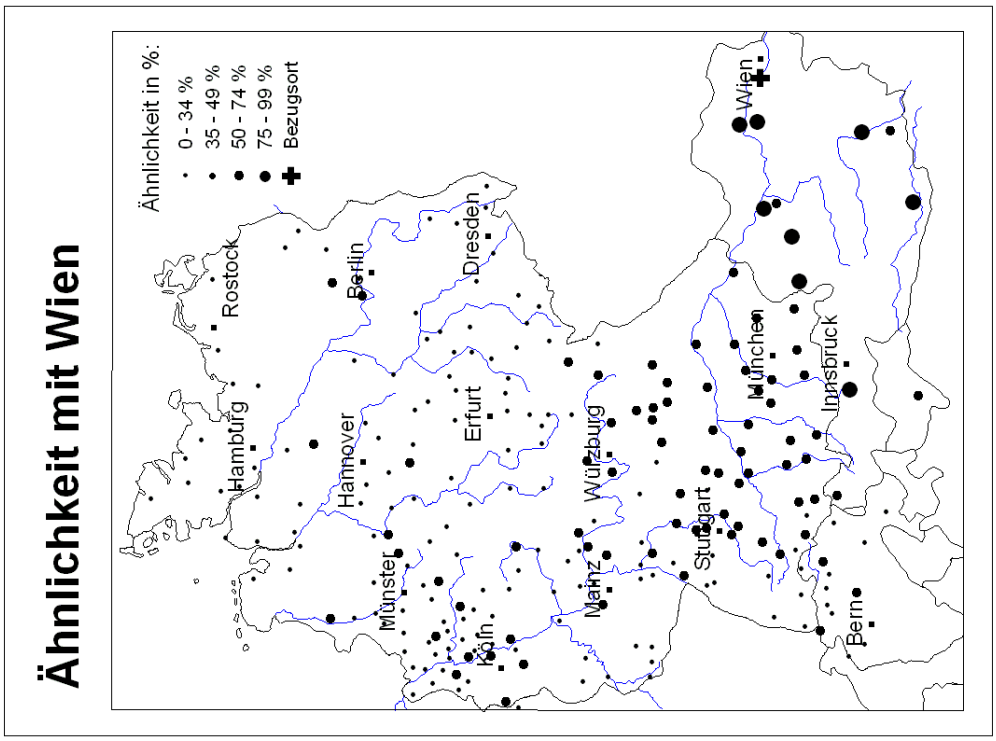
Karte 11: Ähnlichkeitskarte „Augsburg“ (ältere Generation)



Karte 12: Ähnlichkeitskarte „Augsburg“ (jüngere Generation)



Karte 14: Ähnlichkeitskarte „Wien“ (jüngere Generation)

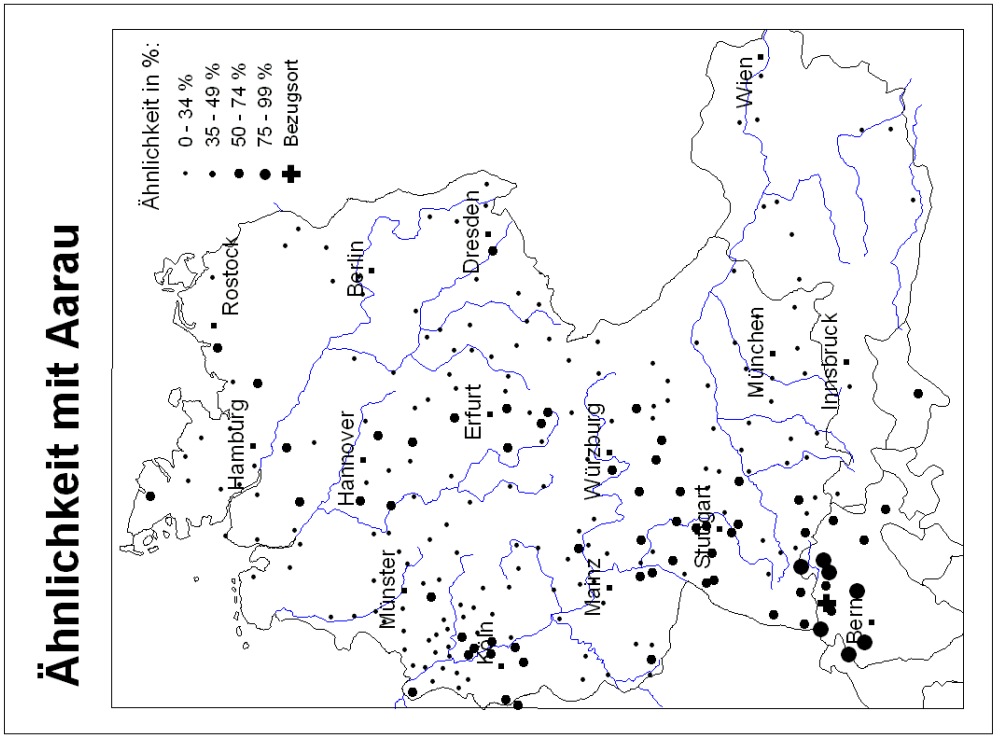


Karte 13: Ähnlichkeitskarte „Wien“ (ältere Generation)

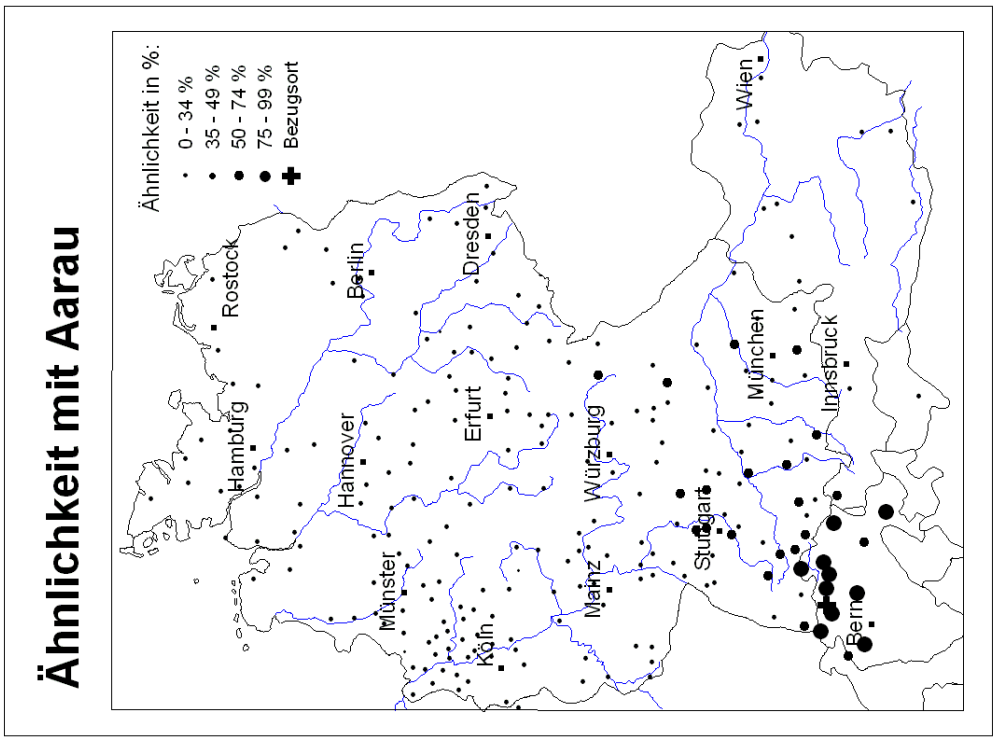
Die deutlichsten Anzeichen zur Entregionalisierung der Umgangssprache zeigen von den untersuchten Karten diejenige mit Bezugsort Wien. Die Charakteristika der bereits besprochenen Wien-Karte (s. Karte 7) sind auch in den altersdifferenzierten enthalten. So erkennt man in der Karte der älteren Generation (s. Karte 13) die relative Homogenität Österreichs mit dem Übergangsbereich Vorarlberg, die scharfe Grenze zum Schweizerdeutschen und der Bezug zum süddeutschen Sprachraum, vornehmlich zum Bairischen. Bei der Karte der jüngeren Generation (s. Karte 14) ist dieser Bezug deutlich verstärkt. Orte der zweithöchsten Übereinstimmungskategorie sind nun auch im Niederalemannischen, im Ostmitteldeutschen und mit deutlicher Zunahme im Westfälischen, bzw. im gesamten norddeutschen Sprachraum zu finden. Wien stellt in dieser Hinsicht ein sehr ausgeprägtes Beispiel für den Entregionalisierungsprozess dar, der in anderen Karten österreichischer Orte nicht so deutlich ersichtlich ist. Die Karten mit Bezugsort Salzburg, die nicht dargestellt sind, offenbaren zwar auch diese Tendenz, allerdings nicht so deutlich wie diejenigen mit Bezugsort Wien. Eine Generalisierung der Aussage der Entregionalisierung würde also nicht stimmen.[48]

Eine Sonderstellung im deutschen Sprachraum nimmt diesbezüglich wiederum das Schweizerdeutsche ein, für das differenzierte Aussagen getroffen werden müssen. Während es einerseits eine Gruppe von Orten gibt, bei denen eine Entregionalisierung feststellbar ist, weisen andere Orte diese Tendenz nicht auf. Für die beiden Gruppen sind die Beispiele mit Luzern und Aarau als jeweiligem Bezugsort herausgegriffen. Im Kartenbild von Aarau zeigt sich, ähnlich wie in dem von Wien, eine eindeutige Tendenz zum Regionalabbau. Orte, die bezüglich Aarau Ähnlichkeitswerte von über 50 % aufweisen, sind, wenn lediglich die ältere Generation berücksichtigt wird (s. Karte 15), nur in der Schweiz, in Baden-Württemberg und vereinzelt in Bayern zu finden. In der Karte der unter 30-jährigen (s. Karte 16) sind diese Orte im gesamten deutschen Sprachraum mit Ausnahme Österreichs nachweisbar. Vor allem in Baden-

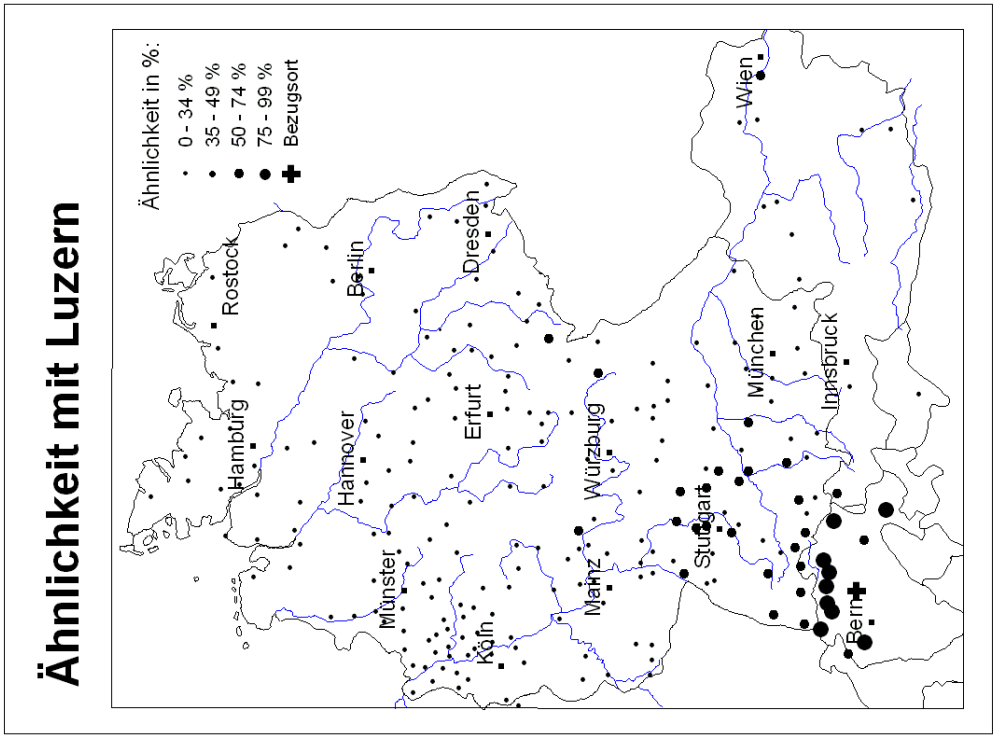
[48] Zur Untersuchung der These bedürfte einer aufwändigen Auswertung der Ähnlichkeitskarten und -matrizen, die an dieser Stelle nicht geleistet werden kann.



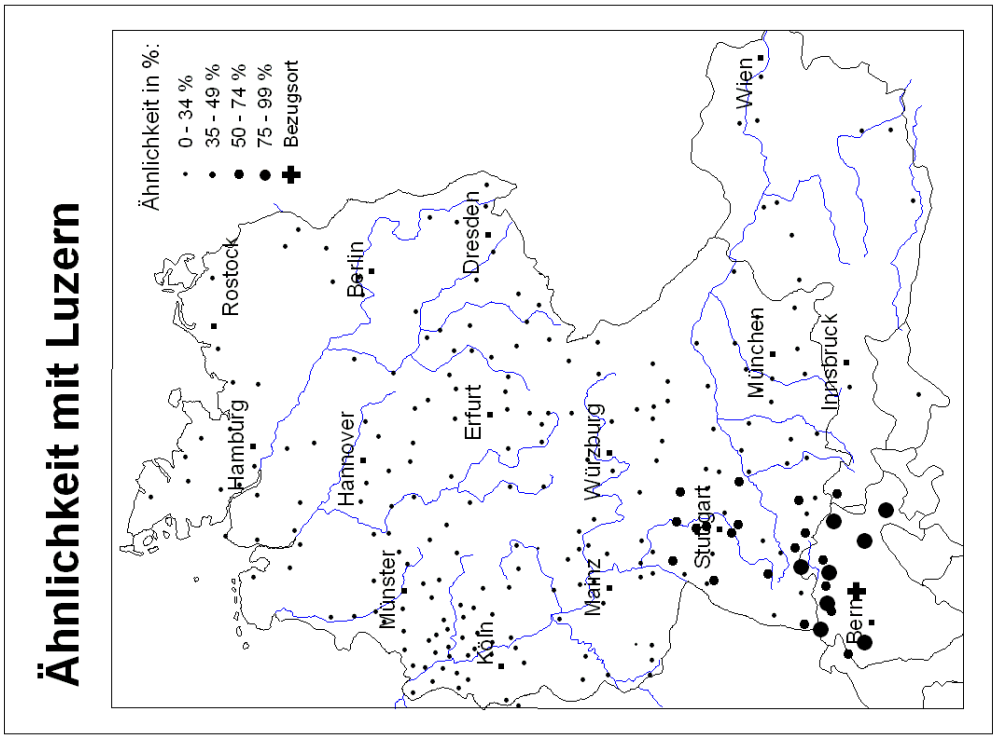
Karte 16: Ähnlichkeitskarte „Aarau“ (jüngere Generation)



Karte 15: Ähnlichkeitskarte „Aarau“ (ältere Generation)



Karte 17: Ähnlichkeitskarte „Luzern“ (ältere Generation)



Karte 18: Ähnlichkeitskarte „Luzern“ (jüngere Generation)

Württemberg und im norddeutschen Sprachgebiet (um Erfurt, in Westfalen und um Hannover) ist eine deutliche Zunahme zu verzeichnen. Nicht nur das Kartenbild offenbart diese Tendenz, auch beim Vergleich der Ähnlichkeitsmatrizen der jüngeren ($Ä_{jung}$) mit derjenigen der älteren Generation ($Ä_{alt}$) ist eine Zunahme der Ähnlichkeitswerte von durchschnittlich 10 %, maximal sogar über 20 % zu beobachten. Für den Ort Aarau kann also ein deutlicher Trend zur Entregionalisierung der regionalen Umgangssprache festgestellt werden.

Anders sieht das im Fall von Luzern aus. Vergleicht man die beiden Kartenbilder miteinander, kann eine Tendenz zum Regionalabbau nicht erkannt werden. Hier ist – auch wenn nur geringfügig – der gegenteilige Prozess, d. h. eine zunehmende Regionalisierung, zu beobachten. Die Orte mit unter 50 % Übereinstimmung sind in der Karte der jüngeren Generation (s. Karte 18) nicht nur seltener vertreten, sondern umfassen auch einen kleineren Raum als in der Karte der älteren Generation (s. Karte 17). Luzern stellt dabei kein Einzelbeispiel dar, z. B. auch die hier nicht dargestellten Karten mit Zürich als Bezugsort zeigen ein ähnliches Kartenbild mit fehlenden Entregionalisierungsprozessen. Über die Ursachen kann an dieser Stelle nur gemutmaßt werden. Möglicherweise spielt die Nähe Aaraus zum deutschen Staatsgebiet im Gegensatz zu Luzern eine Rolle, vielleicht auch eine damit verbundene fehlende Identität der Sprecher zum Schweizerdeutschen.

Auf alle Fälle kann das Fazit gezogen werden, dass der in der Literatur beschriebene Prozess zur *Konvergenz* und damit zur *Entregionalisierung* der Umgangssprache auch in den Ähnlichkeitskarten des AdA nachweisbar ist. Mehrere Beispiele belegen mehr oder weniger diesen Trend. Eine Ausnahme stellt dabei die Schweiz dar, wo ein differenzierteres Bild angenommen werden muss. Um genauere Aussagen treffen zu können, wäre allerdings eine intensivere Analyse der altersdifferenzierten Ähnlichkeitskarten und -matrizen vonnöten.

4.3 Gesamtgliederung

Versuche, das deutsche Sprachgebiet anhand bestimmter Merkmale zu gliedern und zu klassifizieren, sind in der Dialektologie weit verbreitet. Vor allem die allgemein bekannten, traditionellen Dialektgliederungen aufgrund phonologischer (und tlw. morphologischer) Merkmale sind hier zu nennen.[49] Es liegen aber bereits auch Klassifizierungsversuche der deutschen Umgangssprachen vor. Zu erwähnen sind hier DURRELL[50] mit einer manuellen und MÖLLER[51] mit einer computativen Auswertung der Daten von EICHHOFFs „Wortatlas der Umgangssprachen“.

Mit Hilfe der Clusteranalyse als statistisches Klassifizierungsinstrument wurden im dritten Schritt der dialektometrischen Verfahrenskette (s. Kapitel 3.3) Karten mit dem Ziel erstellt, eine Gesamtgliederung der regionalem Umgangssprachen unabhängig von einem Bezugsort zu erstellen (ein Cluster repräsentiert dabei jeweils einen bestimmten Sprachraum). Die Ergebnisse sollen nun im Folgenden vor allem unter der Frage vorgestellt werden, ob auf Grundlage der Daten des AdA mit seiner besonderen Erhebungsmethode plausible und mit den Ergebnissen der dialektometrischen Auswertungen des WDU (vgl. Abb. 11) vergleichbare Lösungen erzielt werden können.

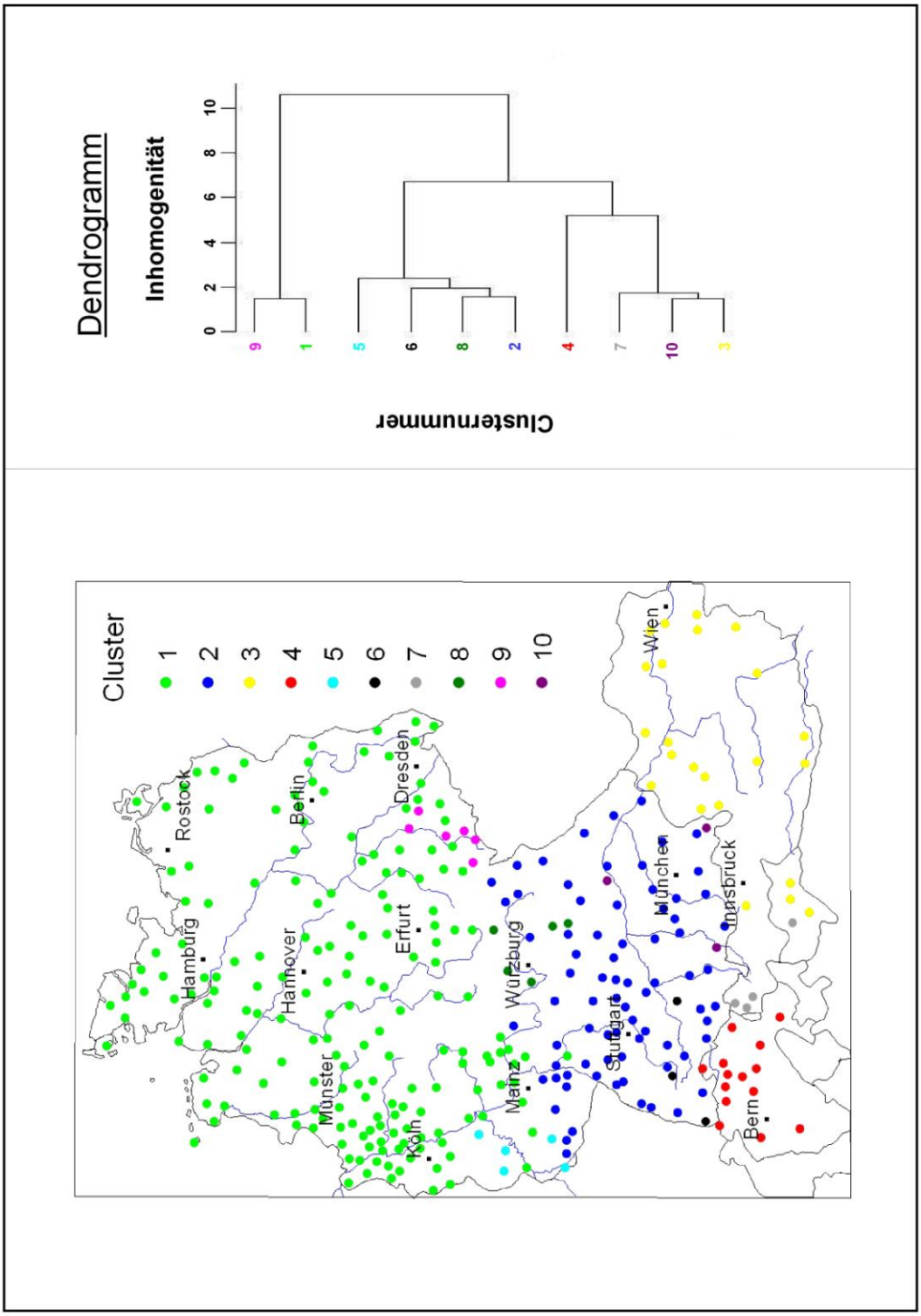
Karte 19 zeigt das Klassifizierungsergebnis der auf $Ä_{gesamt}$ angewendeten hierarchischen Clusteranalyse. Grundlage der Klassifizierung ist die alle Fragen und damit alle in diesem Zusammenhang unterschiedenen Kategorien (lexikalisch, morphologisch, phonologisch, syntaktisch) umfassende Datenmatrix des AdA. Nach den bereits erläuterten Kriterien wurde eine 10-Clusterlösung[52] gewählt, wobei jedes Cluster mit einer unterschiedlichen Farbe dargestellt ist. Der weitere Verlauf des Clusterns kann dem Dendrogramm entnommen werden. Es zeigt an, welche Cluster in den nächsten Schritten nach dem Prinzip der größten Ähnlichkeit zusammengefasst werden. Der x-Achse

[49] Für eine gute Übersicht s. z. B. König (2005, 230).

[50] S. Durrell (1989).

[51] S. Möller (2003).

[52] Die weiteren Clusterlösungen sind im Anhang enthalten.



Karte 19: 10-Clusterlösung der auf alle Variablen angewendeten hierarchischen Clusteranalyse

ist der Grad der Inhomogenität einer Fusionierung zu entnehmen. Sie gibt somit Aufschluss über die Ähnlichkeit bzw. Unterschiedlichkeit von Clustern: Eng zusammenstehende „Äste“ des Baums deuten auf die Zusammenlegung sehr ähnlicher Cluster hin, weit auseinander stehende auf die Zusammenlegung sehr unterschiedlicher Cluster, was logischerweise vor allem gegen Ende des Fusionsprozesses auftritt. In jedem Schritt wird das Cluster mit der jeweils höchsten Nummer mit einem weiteren, das dem Dendrogramm entnommen werden kann, zusammengefügt. Das entstandene Cluster trägt fortan die kleinere Nummer der beiden in die Fusionierung eingegangen Cluster.

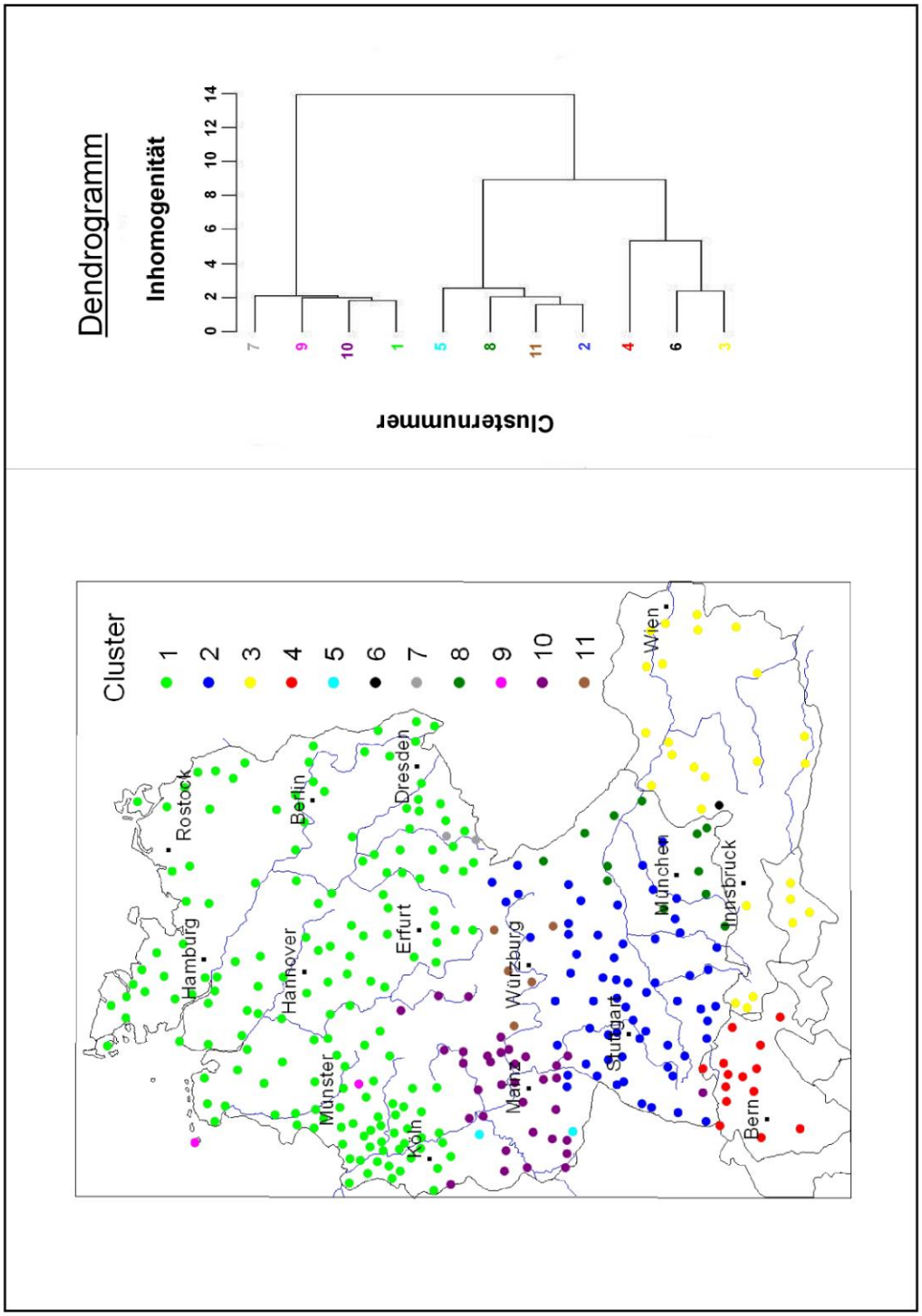
Im ersten Schritt wird also Cluster 10 mit Cluster 3 (Lösung s. Anhang 1) fusioniert. In den nächsten Schritten werden die noch relativ ähnlichen Cluster 9 und Cluster 1 (s. Anhang 2), Cluster 8 und Cluster 2 (s. Anhang 3), Cluster 7 und Cluster 3 (s. Anhang 4), usw. (s. Anhang 5-8) zusammengefügt. Die Vorteile der Darstellung sind sofort ersichtlich. Es kann nicht nur die nach statistischen und fachlich-logischen Kriterien ermittelte „optimale Lösung“ veranschaulicht werden. Gleichzeitig kann der weitere Fusionierungsprozess nachvollzogen und die Stabilität, d. h. der Grad der Eigenständigkeit eines Clusters (= Sprachgebiet), abgelesen werden: Je niedriger die Clusternummerierung eines Sprachgebiets, desto eigenständiger und zu den anderen Sprachgebieten unterschiedlicher ist es.

Interpretiert man das Ergebnis (s. Karte 19), erkennt man auf einen Blick die Strukturen, die anhand der einzelnen Ähnlichkeitskarten mühevoll herausgearbeitet wurden. Neben dem einheitlichen und räumlich sehr ausgedehnten norddeutschen und einem etwas differenzierteren süddeutschen Cluster südlich der Mainlinie ist vor allem ein österreichischer und schweizerischer Cluster zu erkennen, die exakt den politischen Grenzverlauf wiedergeben. Zudem erkennt man noch kleinere, auf bestimmte Regionen beschränkte Cluster. Hier sind vor allem ein Cluster rund um die Mosel, einer südwestlich von Dresden und ein weiterer in Vorarlberg zu nennen. Zusätzliche Cluster sorgen für eine Differenzierung des süddeutschen Sprachraums, wobei Cluster 6 und 10 unsystematisch über das Gebiet streuen und nur Cluster 8 räumlich einheitlich einen Übergangsraum zum Norddeutschen andeutet.

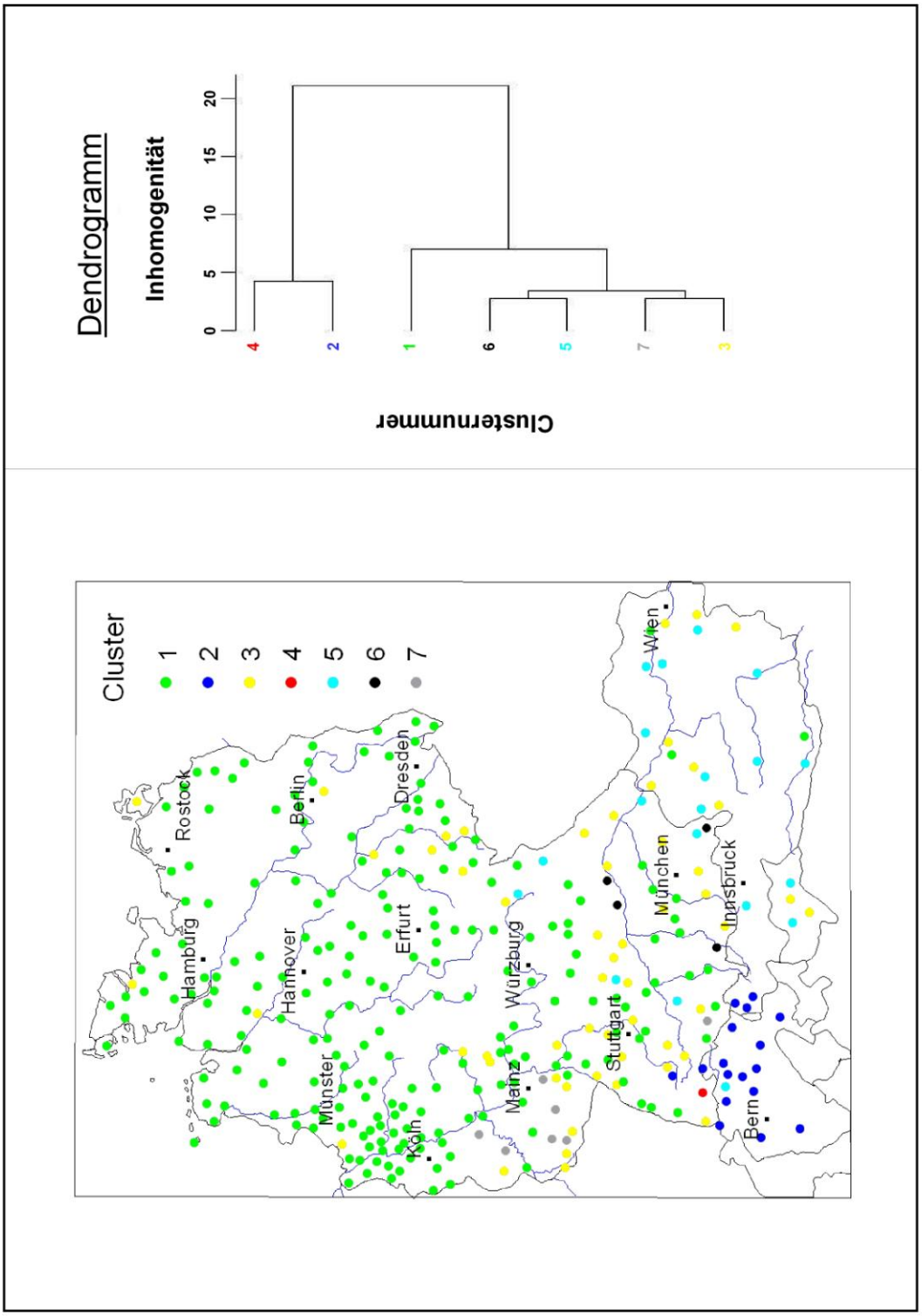
Im weiteren Verlauf des Klassifizierungsprozesses wird, wie dem Dendrogramm entnommen werden kann, Cluster 10 überraschenderweise mit Cluster 3, d. h. mit dem österreichischen Sprachgebiet und nicht dem süddeutschen, zusammengefügt. Von den auf die räumliche Ausdehnung bezogenen kleineren Clustern ist das Mosel- / Rheinfränkische am stabilsten und wird erst im fünftletzten Fusionierungsprozess mit dem süddeutschen Cluster zusammengefügt. In der 4-Clusterlösung (s. Anhang 6) erkennt man dann die vier großen, sehr unterschiedlichen Wortgebrauchsräume: der norddeutsche, der süddeutsche, der österreichische und der schweizerische. Die beiden letztgenannten werden im nächsten Schritt zu einem Cluster zusammengefasst (s. Anhang 7), der dann wiederum mit dem süddeutschen fusioniert wird (s. Anhang 8). Als Schlussfolgerung kann also festgestellt werden, dass die schärfsten Wortgebrauchsraumgrenzen zum einen entlang des Mains („Mainlinie“), zum anderen entlang der Staatsgrenzen zwischen Deutschland, Österreich und Schweiz zu finden sind. Eine besondere Aufmerksamkeit muss außerdem dem mosel- / rheinfränkischen Fächer zugestanden werden.

Im Folgenden sollen nun die Ergebnisse der nach den sprachwissenschaftlichen Kategorien differenzierten Datensätze im Einzelnen kurz vorgestellt und auf mögliche Abweichungen eingegangen werden. Dabei muss erwähnt werden, dass vor allem bei den morphologischen, syntaktischen und phonologischen Ergebnissen aufgrund der geringen Anzahl der extrahierten Variablen wenig differenzierte und nur bedingt statistisch gesicherte Aussagen getroffen werden können.

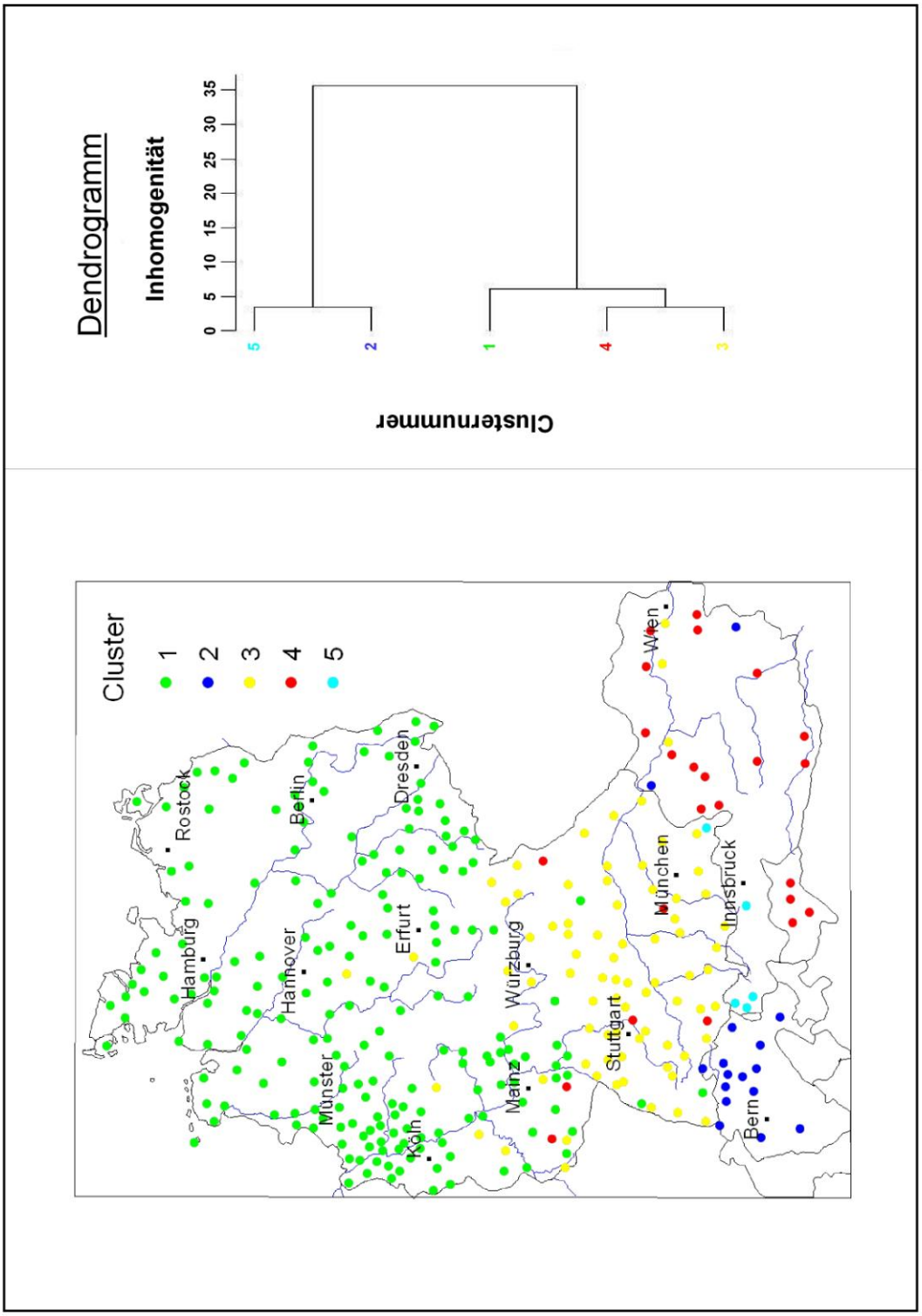
Fließen in den Datensatz nur lexikalische Variablen ein (s. Karte 20), zeigt das Ergebnis der Clusteranalyse ein ähnliches Bild im Vergleich zu Karte 19. Abgesehen von ein paar Ausnahmen (Cluster 9, Cluster 6, Ableger von Cluster 10 im südlichen Baden-Württemberg) ist eine erwartbare diatopische Gliederung erkennbar. Auffallende Strukturen sind die Abtrennung Ostbayerns vom restlichen süddeutschen Sprachgebiet und der ausgeprägte westmitteldeutsche Fächer (mit Ausnahme des Ripuarischen), der weit nach Nordosten vordringt und so an die traditionelle (auf phonologischen Kriterien basierende) Dialektgliederung erinnert.



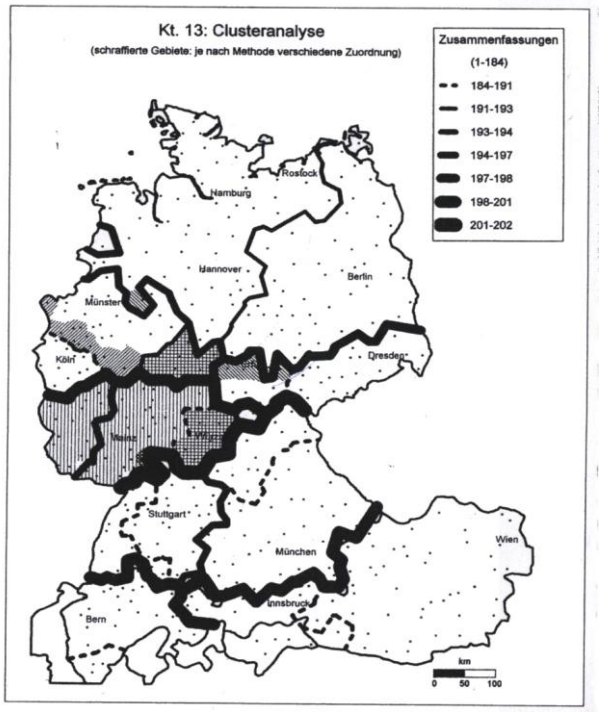
Karte 20:11-Clusterlösung der auf die lexikalischen Variablen angewendeten hierarchischen Clusteranalyse



Karte 21: 7-Clusterlösung der auf die syntaktischen Variablen angewendeten hierarchischen Clusteranalyse



Karte 22: 5-Clusterlösung der auf die morphologischen Variablen angewendeten hierarchischen Clusteranalyse



<u>Abb. 11</u>: Ergebnis der dialektometrischen Auswertung des WDU (Quelle: Möller (2003, 288))

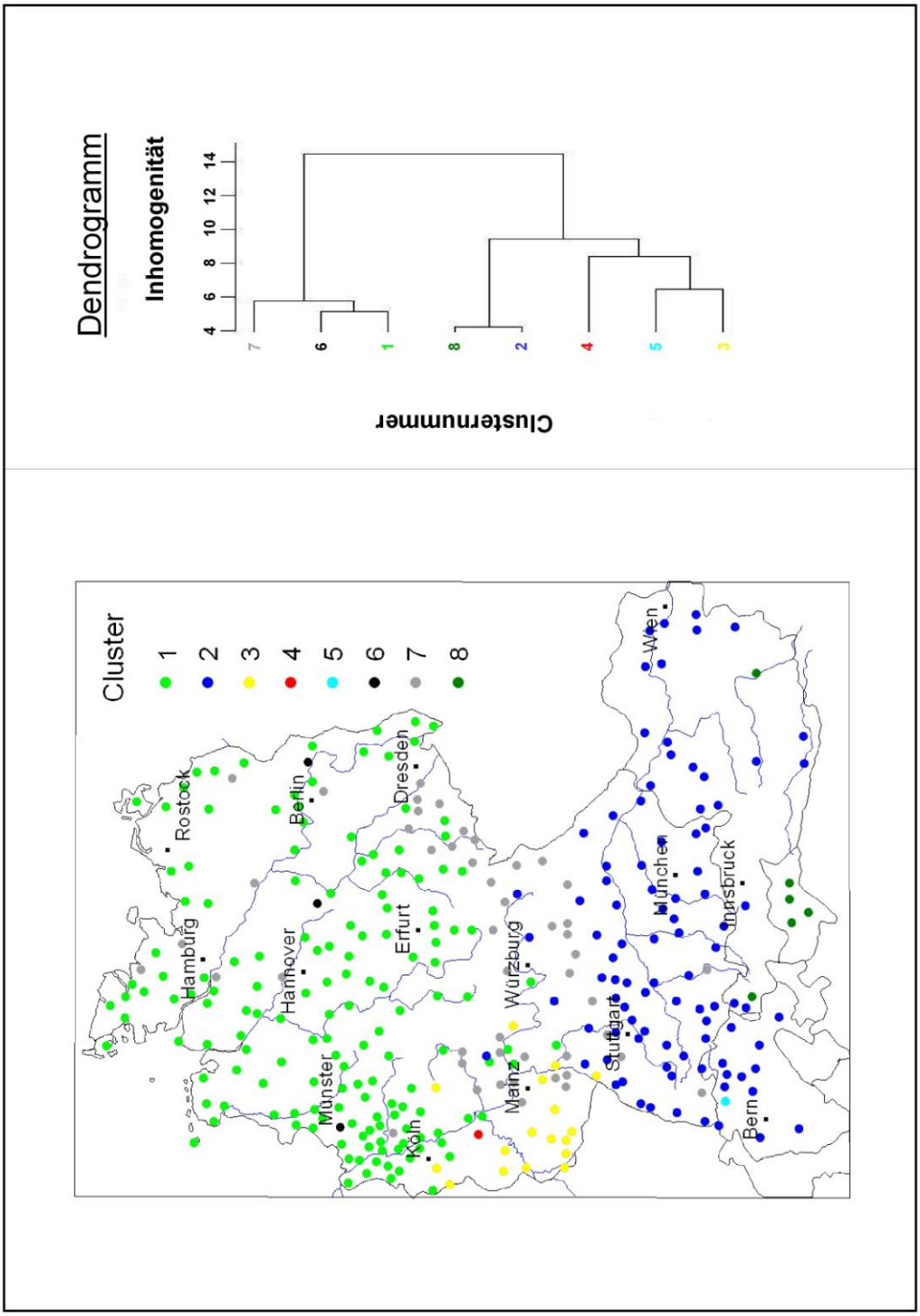
Durch die Berücksichtigung lediglich lexikalischer Variablen ist ein Vergleich mit der dialektometrischen Analyse des „Wortatlas der deutschen Umgangssprachen“ möglich, in die ebenfalls nur lexikalische Variablen eingehen.[53] Abbildung 11 zeigt das Ergebnis der auf die Daten des WDU angewendeten hierarchischen Clusteranalyse. Die Strichstärke gibt dabei Auskunft über die Ähnlichkeit von Clustern: Je dicker die Grenze, desto unterschiedlicher sind die abgetrennten Sprachgebiete und desto später werden sie zusammengefasst. Vergleicht man die beiden Kartenbilder miteinander, kann festgestellt werden, dass das dialektometrische Ergebnis des AdA das des WDU im Wesentlichen bestätigt. Nord- und Süddeutschland (mit stärkerer Differenzierung), das Westmitteldeutsche und Österreich sowie die Schweiz als wichtigste einheitliche Sprachgebiete auf einer oberen Stufe sind erkennbar. Verfeinert man die Differenzierung, lassen sich jedoch auch Unterschiede ausmachen. Zum einen weisen die Ergebnisse des WDU eine höhere Detailgenauigkeit sowie eine höhere Konsistenz (bezüglich Ausreißern) auf, zum anderen könne auch auf der inhaltlichen Ebene Unterschiede festgestellt werden. Auffallende Strukturen sind vor

[53] Vgl. Möller (2003).

allem die auf der WDU-Karte erkennbare Grenze nördlich der Mainlinie bzw. des Westmitteldeutschen, die von Münster bis nach Dresden reicht, und die Teilung des süddeutschen Sprachraums entlang der bayerisch-baden-württembergischen Landesgrenze. Diese Aspekte lassen sich in den Auswertungen des AdA nicht nachweisen.

Im Kartenbild der auf syntaktischen und morphologischen Variablen basierenden Ergebnisse zeigt sich, wie wichtig für eine hohe statistische Aussagekraft eine möglichst große Anzahl an Variablen ist. Zwar erkennt man in beiden Fällen noch Muster, allerdings prägen viele nicht erklärbare Ausnahmen das Kartenbild, v. a. in der syntaktischen Karte (s. Karte 21). Österreich, die Schweiz, das Mosel-/Rheinfränkische, Süd- und Ostdeutschland sind zwar als eigenständige Wortgebrauchsräume erfassbar, allerdings fehlt das bisher gesehene konsistente Bild. Eindeutige räumliche Muster sind noch eher in der morphologischen Karte zu finden (s. Karte 22). Auffallend ist in diesem Fall, dass Österreich einen größeren Bezug zu Süddeutschland und Vorarlberg einen größeren zum Schweizerdeutschen besitzt. Aber auch in dieser Karte sind viele nicht zu erklärende Ausnahmefälle zu beobachten.

Besondere Aufmerksamkeit verdienen die auf phonologischen Variablen beruhenden Ergebnisse der Clusteranalyse (s. Karte 23), da sich in diesem Fall bisher noch nicht beobachtbare Strukturen zeigen. Auch hier sind zwar aufgrund der geringen Variablenanzahl mehrere nicht erklärbare Phänomene (z. B. Cluster 4, 5 und 6) enthalten, klassifiziert man diese aber gedanklich zu den jeweils größeren Sprachgebieten, so ergibt sich ein relativ konsistentes Bild. Im Gegensatz zu den bisherigen Karten bildet das Gebiet südlich der Mainlinie, die nun leicht nach Süden verschoben ist, einen einheitlichen Sprachraum mit Ausnahme von Südtirol (wahrscheinlich durch den Einfluss der italienischen Aussprache). Auffällig ist auch der breite Streifen rund um den Main, der zur einer Dreiteilung des gesamten Sprachgebiets führt, die – wenn auch mit einer Südverlagerung – an die klassische Dreiteilung in Nieder-, Mittel- und Oberdeutsch erinnert.



Karte 23: 8-Clusterlösung der auf die phonologischen Variablen angewendeten hierarchischen Clusteranalyse

Fasst man die einzelnen Ergebnisse der Clusteranalysen zusammen, kann aus diatopischer Perspektive das Fazit gezogen werden, dass die Karten zumeist ein plausibles Bild wiedergeben. Der Vergleich mit den Ergebnissen der dialektometrischen Analyse des WDU verstärkt diesen Eindruck. Auch wenn nicht dessen detaillierte Differenzierung erreicht wird, so sind dennoch die wichtigsten Strukturen erkennbar. Allerdings müssen die in manchen Karten vorhandenen, nicht erklärbaren „Ausnahmefälle“ erwähnt werden, die aber in einer unsystematisch gewonnen Datenmatrix wohl unumgänglich sind.

Kapitel 5

Ausblick

Dass die Dialektometrie zu einem Erkenntnisgewinn in der sprachwissenschaftlichen Forschung beitragen kann, haben die vorliegenden Ausführungen gezeigt. Viele Anwendungsmöglichkeiten sind vorstellbar, von denen ein paar Beispiele in Ansätzen aufgezeigt wurden. Dass diese auf einem mit geringen Kosten und Anstrengungen durch die Methode der Online-Befragung gewonnen Datensatz beruhen, ist dabei genauso erstaunlich wie die plausiblen Ergebnisse dessen dialektometrischer Auswertung. Um zu noch detaillierteren und differenzierteren Aussagen zu gelangen, vergleichbar mit denen der dialektometrischen Auswertung von EICHHOFFs „Wortatlas der deutschen Umgangssprachen“, bedarf es noch weiterer methodischer Verbesserungen.

Anzusetzen wäre hierbei schon bei der Datengewinnung. Die Ergebnisse einer dialektometrischen Auswertung werden umso besser, je mehr Variablen (d. h. Fragen) und je mehr Orte in die Analyse eingehen. Letzteres kann beim „Atlas zur deutschen Alltagssprache“ durch eine neue Zuweisungsmethode der Meldungen ermöglicht werden. Da bei der bisherigen Methode (s. Kapitel 2.2) viele Antworten der Informanten nicht verwertet werden können, muss über eine Neugestaltung (z. B. Zuweisung basierend auf Postleitzahl) nachgedacht werden. Die Anzahl der Variablen hingegen erhöht sich mit der Anzahl der Runden, sodass hier in Zukunft mit immer besseren und statistisch gesicherteren Ergebnissen zu rechnen ist. Möchte man zu den einzelnen sprachwissen-

schaftlichen Kategorien Aussagen treffen, muss die Anzahl der diesbezüglichen Variablen erhöht werden.

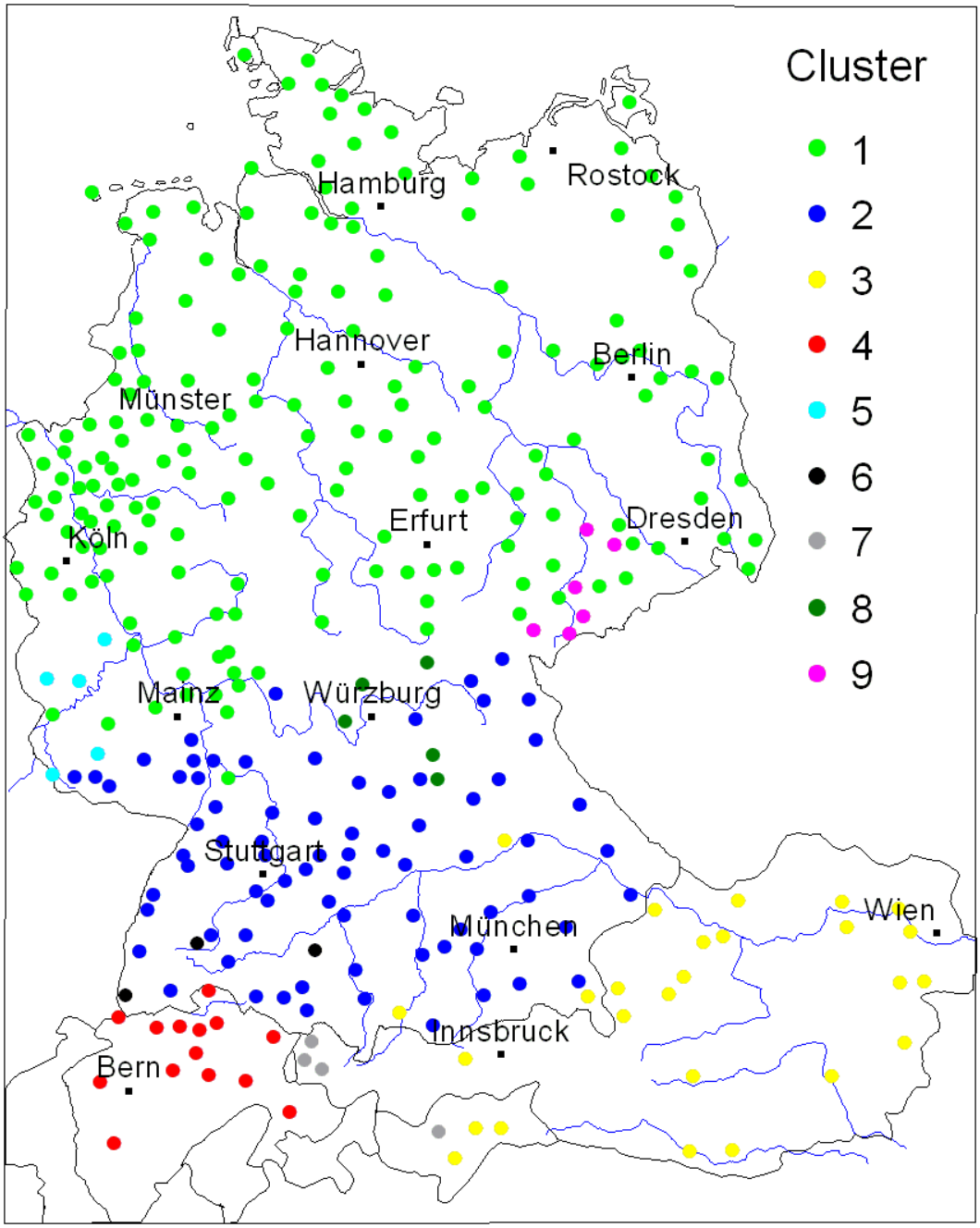
Auch im dialektometrischen Analyseschritt sind methodische Verbesserungen und zusätzliche Anwendungen möglich.[54] Um die Anzahl der Ortspunkte zu erhöhen, besteht eine Möglichkeit in der Programmierung eines dynamischen Programms, das Fehlwerte automatisch erkennt und nicht in die Analyse eingehen lässt. So können auch Ortspunkte, die nicht in allen Fragerunden vertreten sind, in der Auswertung berücksichtigt werden. Zudem stellt die Statistik als Klassifizierungsverfahren nicht nur die hierarchische Clusteranalyse bereit.[55] Ein Vergleich der unterschiedlichen Methoden kann zu einem weiteren Erkenntnisgewinn beitragen.

Ingesamt enthält die Dialektometrie also noch ein erhebliches Entwicklungspotential, das es in Zukunft auszuschöpfen gilt.

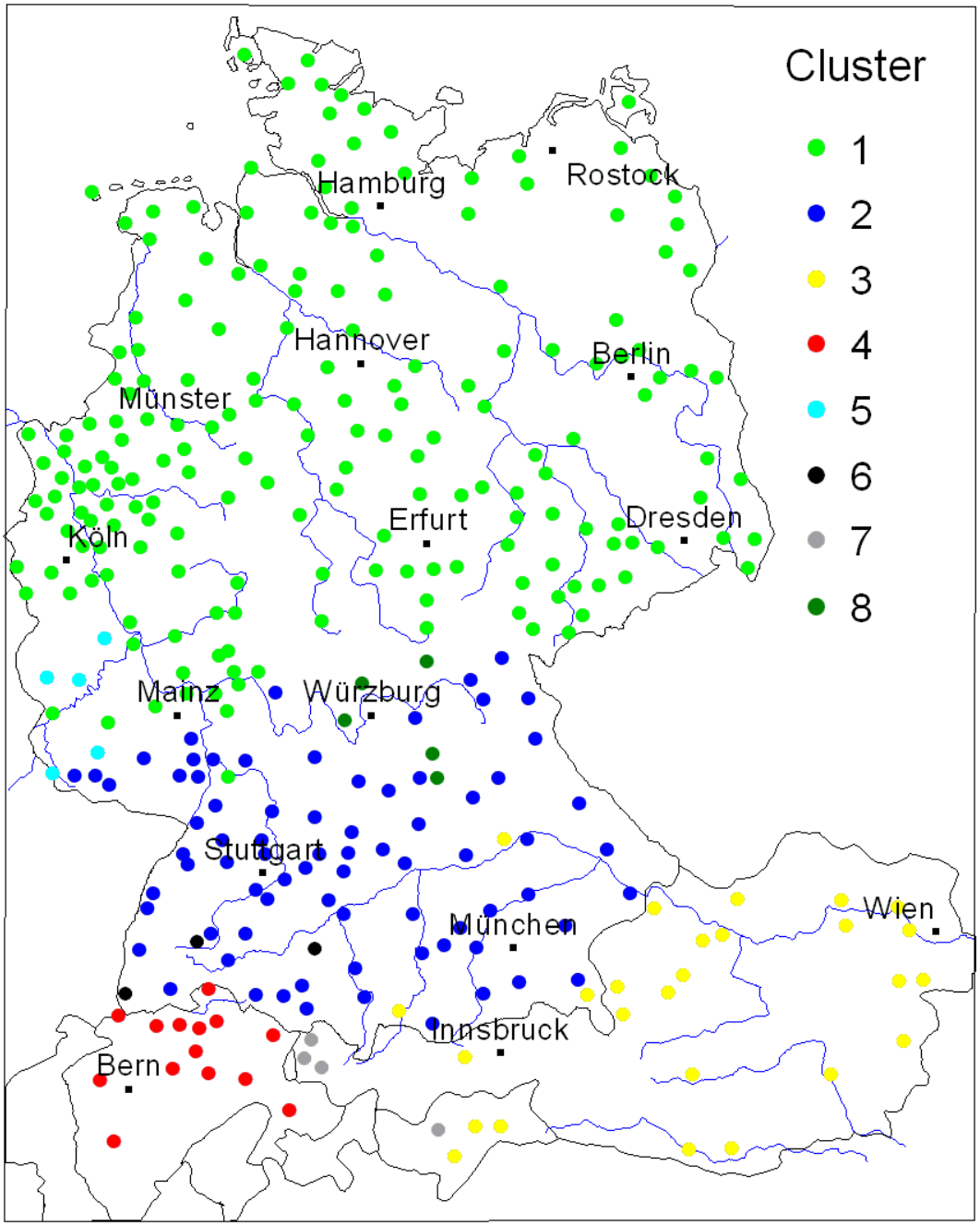
---

[54] Eine ausführliche Darstellung weiterer statistische Anwendungsmöglichkeiten, wie z. B. Distanz- oder Wabenkarten, bietet Goebl (1982).

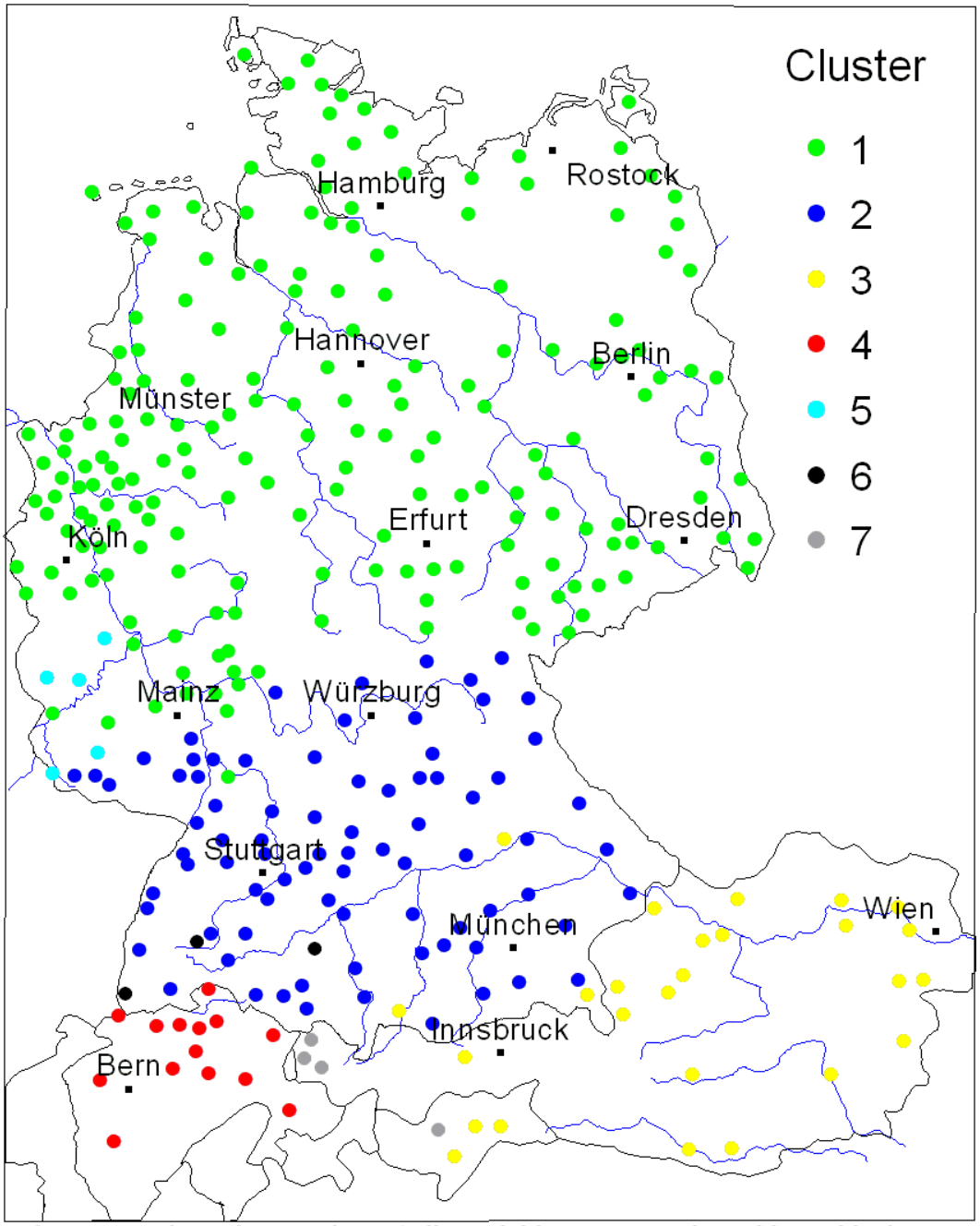
[55] Weitere Methoden sind z. B. die komplexere, aber zu – statistisch gesehen – besseren Lösungen führende nicht-hierarchische Clusteranalyse oder die Hauptkomponentenanalyse (s. Bahrenberg (2003)).



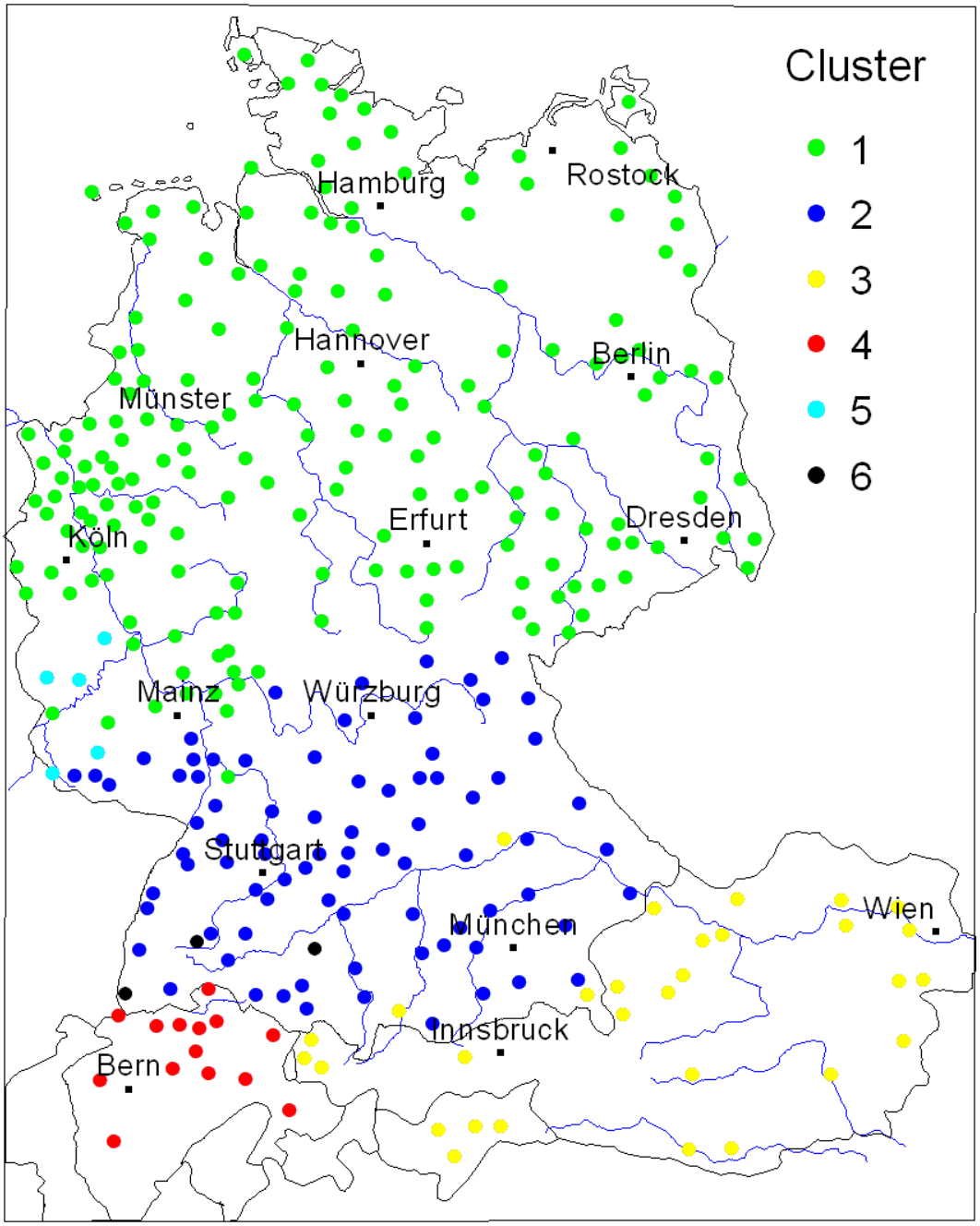
<u>Anhang 1</u>: 9-Clusterlösung der auf alle Variablen angewendeten hierarchischen Clusteranalyse



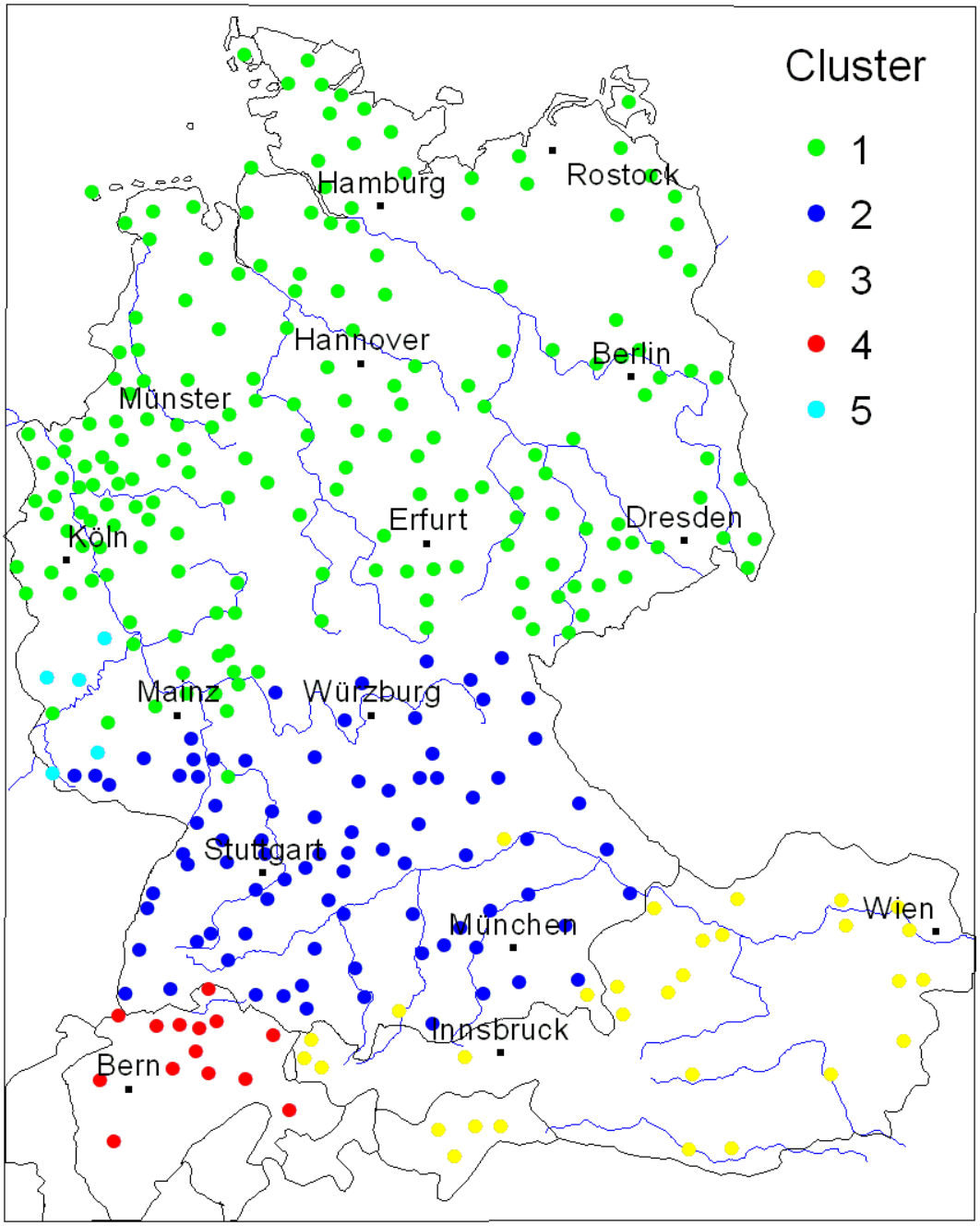
<u>Anhang 2</u>: 8-Clusterlösung der auf alle Variablen angewendeten hierarchischen Clusteranalyse



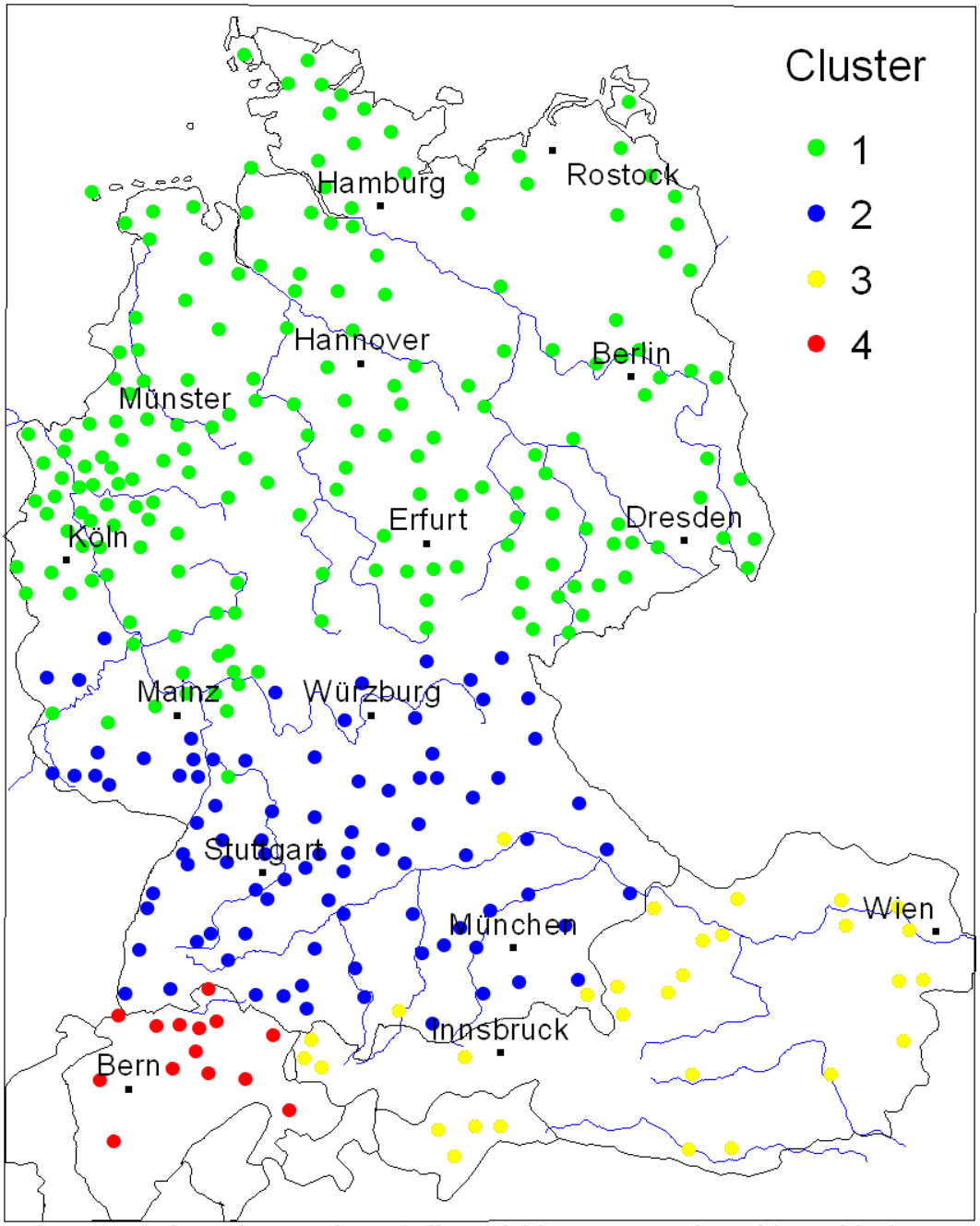
Anhang 3: 7-Clusterlösung der auf alle Variablen angewendeten hierarchischen Clusteranalyse



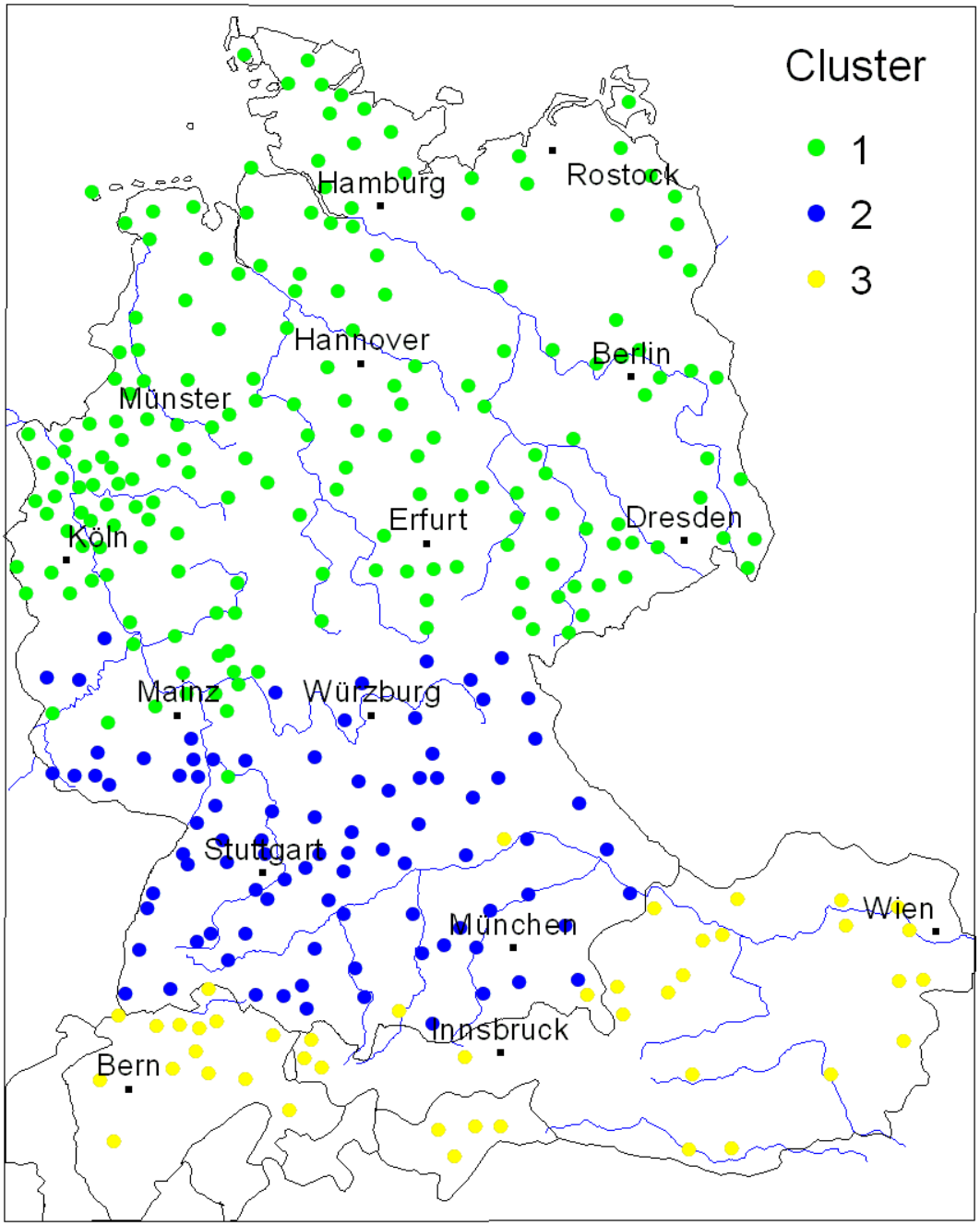
Anhang 4: 6-Clusterlösung der auf alle Variablen angewendeten hierarchischen Clusteranalyse



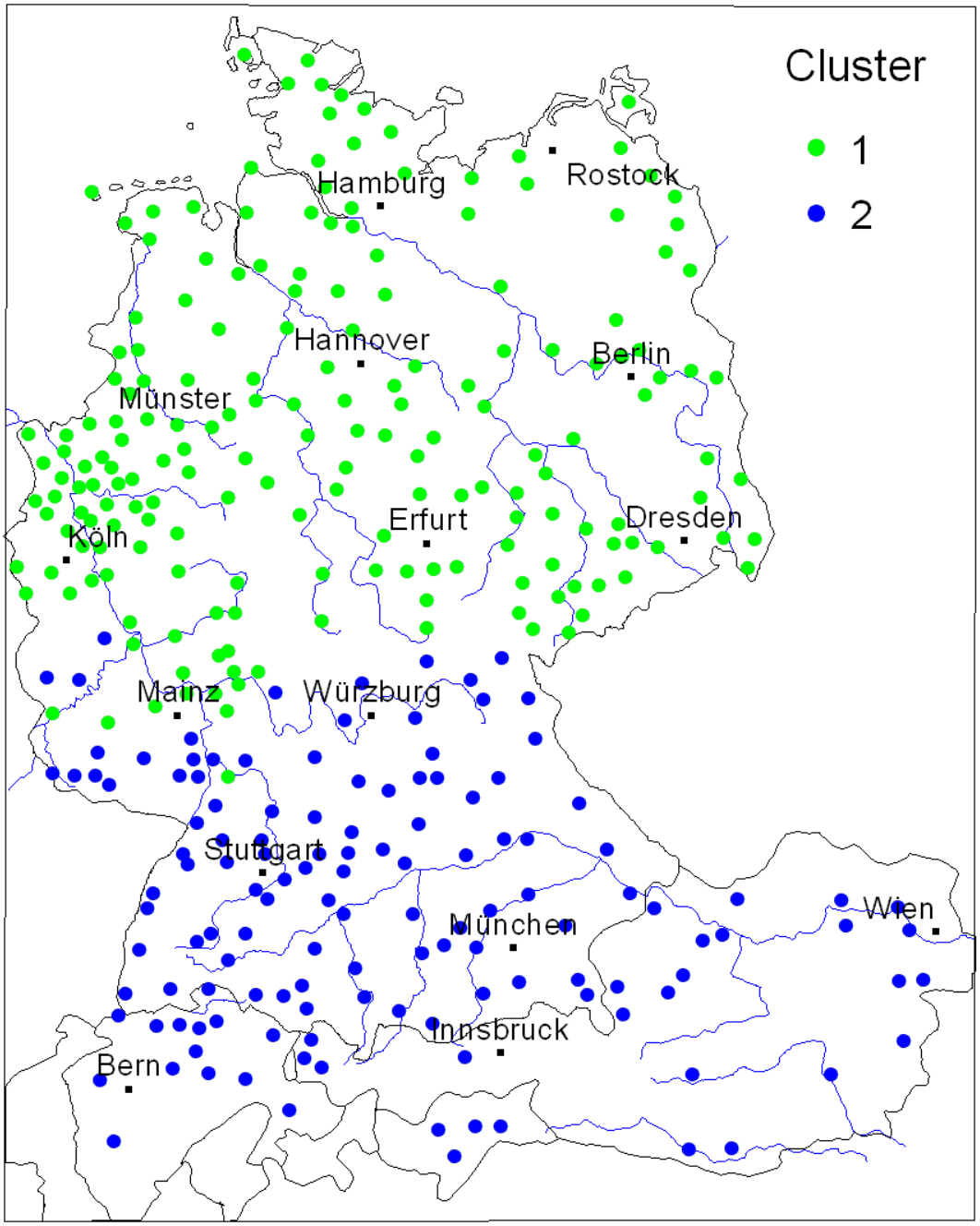
Anhang 5: 5-Clusterlösung der auf alle Variablen angewendeten hierarchischen Clusteranalyse



Anhang 6: 4-Clusterlösung der auf alle Variablen angewendeten hierarchischen Clusteranalyse



Anhang 7: 3-Clusterlösung der auf alle Variablen angewendeten hierarchischen Clusteranalyse



Anhang 8: 2-Clusterlösung der auf alle Variablen angewendeten hierarchischen Clusteranalyse

Literaturverzeichnis

Bahrenberg, Gerhard / Giese, Ernst / Nipper, Josef (1999): Statistische Methoden in der Geographie 1. Stuttgart: Teubner.

Bahrenberg, Gerhard / Giese, Ernst / Nipper, Josef (2003): Statistische Methoden in der Geographie 2. Stuttgart: Teubner.

Bauer, Roland (2002/2003): Dolomitenladinische Ähnlichkeitsprofile aus dem Gadertal. Ein Werkstattbericht zur dialektometrischen Analyse des ALD-I. In: Ladinia XXVI-XXVII, S. 209-250.

Besch, Werner (1983): Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung. Berlin: de Gruyter.

Besch, Werner (1983): Entstehung und Ausprägung der binnensprachlichen Diglossie im Deutschen. In: Besch (198), S. 1399-1410.

Durrell, Martin (1989): Die „Mainlinie“ als sprachliche Grenze. In: Putschke, Wolfgang / Veith, Werner / Wiesinger, Peter (Hrsg.) (1989): Dialektgeographie und Dialektologie. Marburg: Elwert.

Eichhoff, Jürgen (1977ff.): Wortatlas der deutschen Umgangssprachen. Bern, München: Francke.

Elspaß, Stephan (2005): Zum Wandel im Gebrauch regionalsprachlicher Lexik. Ergebnisse einer Neuerhebung. In: Zeitschrift für Dialektologie und Linguistik 72, S. 1-51.

Elspaß, Stephan / König, Werner (Hrsg.) (2008): Sprachgeographie digital. Die neue Generation der Sprachatlanten. Hildesheim [u.a.]: Olms.

Elspaß, Stephan / Möller, Robert (2006): Internet-Exploration: Zu den Chancen, die eine Online-Erhebung regional gefärbter Alltagssprache bietet. In: Osnabrücker Beiträge zur Sprachtheorie 71 [Thematischer Band „Dialekt im Wandel. Perspektiven einer neuen Dialektologie“], S. 143-158.

Goebl, Hans (1981): Isoglossen, Distanzen und Zwischenpunkte. Die dialektale Kammerung der Rätoromania und Oberitaliens aus dialektometrischer Sicht. In: Ladinia V, S. 23-55.

Goebl, Hans (1982): Dialektometrie. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Goebl, Hans (1984): Dialektometrische Studien (Band I-III). Tübingen: Niemeyer.

Goebl, Hans (1992): Dendrogramme im Dienst der Dialektometrie. Zwei hierarchisch-agglomerative Klassifikationen von Daten des Sprachatlasses AIS. In: Klenk, Ursula (1992): Computatio linguae. Aufsätze zur algorithmischen und quantitativen Analyse der Sprache. Stuttgart: Steiner, S. 54-73.

Goebl, Hans (1993): Probleme und Methoden der Dialektometrie: Geolinguistik in globaler Perspektive. In: Viereck, Wolfgang (Hrsg.) (1993): Verhandlungen des internationalen Dialektologenkongresses (Band I). Stuttgart: Steiner. S. 37-81.

Hummel, Lutz (1993): Dialektometrische Analysen zum Kleinen Deutschen Sprachatlas (KDSA) (Teil I-II). Tübingen: Niemeyer.

König, Werner (2005): dtv-Atlas zur deutschen Sprache. 14.. durchges. und aktual. Aufl. München: dtv.

Löffler, Heinrich (2003): Dialektologie. Eine Einführung. Tübingen: Gunter Narr Verlag

Möller, Robert (2001): „Wortgebrauchsräume“ in der nord- und mitteldeutschen Alltagssprache – eine Untersuchung zum Wortatlas der deutschen Umgangssprachen. In: Niederdeutsches Wort 41, S. 1-21.

Möller, Robert (2003): Zur diatopischen Gliederung des alltagssprachlichen Wortgebrauchs. Eine dialektometrische Auswertung von Jürgen Eichhoff: Wortatlas der deutschen Umgangssprachen (Bd. 1-4; 1977, 1978, 1993, 2000). In: Zeitschrift für Dialektologie und Linguistik 70, S. 259-297.

Möller, Robert / Elspaß, Stephan (2008): Erhebung dialektgeographischer Daten per Internet: Ein Atlasprojekt zur deutschen Alltagssprache. In: Elspaß / König (2008), S. 115-132.

Putschke, Wolfgang (1993): Zur Kritik dialektologischer Einteilungskarten. In: Viereck, Wolfgang (Hrsg.) (1993): Verhandlungen des internationalen Dialektologenkongresses (Band I). Stuttgart: Steiner, S. 421-443.

Schiltz, Guillaume (1996): Der dialektometrische Atlas von Südwest-Baden (DASB) (Teil I-IV). Marburg: Elwert.

Schuppenhauer, Claus / Werlen, Iwar (1983): Stand und Tendenzen in der Domänenverteilung zwischen Dialekt und deutscher Sprache. In: Besch (1983), S. 1411-1427.

Videsott, Paul (2003): Das Dendrogramm: Eine auch in der Toponomastik geeignete Heuristik? In: Beiträge zur Namensforschung 38, S. 43-74.