

BEIHEFTE

Simon Pröll

Raumvariation zwischen Muster und Zufall

Geostatistische Analysen am Beispiel
des Sprachatlas von Bayerisch-Schwaben

Germanistik

Franz Steiner Verlag

ZDL

ZEITSCHRIFT
FÜR DIALEKTOLOGIE
UND LINGUISTIK

BEIHEFTE

160

Simon Pröll
Raumvariation zwischen Muster und Zufall

ZEITSCHRIFT FÜR DIALEKTOLOGIE UND LINGUISTIK
BEIHEFTE

In Verbindung mit Michael Elmentaler und Jürg Fleischer

herausgegeben von Jürgen Erich Schmidt

BAND 160

Simon Pröll

Raumvariation zwischen Muster und Zufall

Geostatistische Analysen am Beispiel
des Sprachatlas von Bayerisch-Schwaben



Franz Steiner Verlag

Bibliografische Information der Deutschen Nationalbibliothek:

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

Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist unzulässig und strafbar.

© Franz Steiner Verlag, Stuttgart 2015

Druck: Hubert & Co., Göttingen

Gedruckt auf säurefreiem, alterungsbeständigem Papier.

Printed in Germany.

ISBN 978-3-515-11055-6 (Print)

ISBN 978-3-515-11056-3 (E-Book)

INHALTSVERZEICHNIS

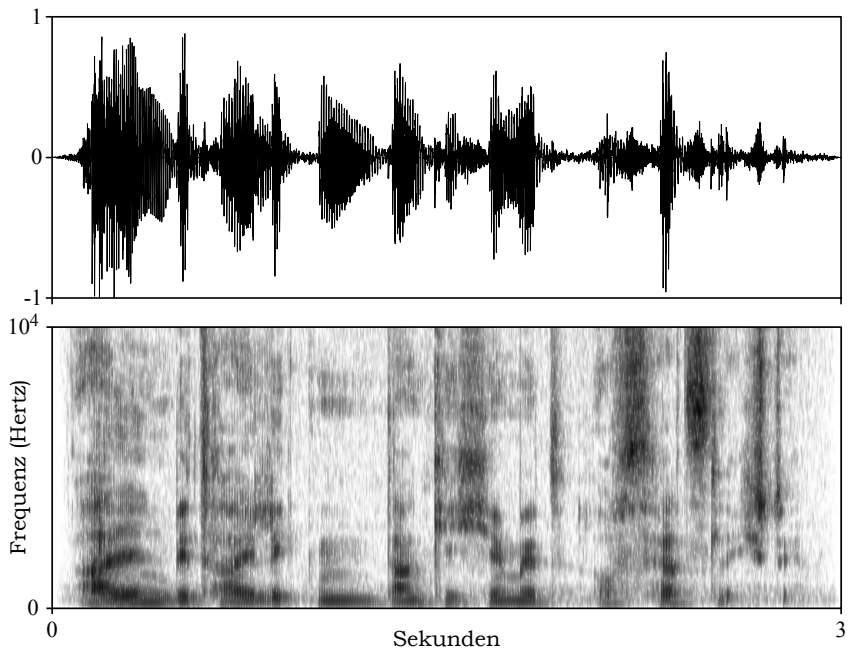
VORWORT	9
1 EINLEITUNG UND ÜBERSICHT	11
2 HERANFÜHRUNG: VOM EINZELBELEG ZUR VARIETÄT	13
2.1 Grundlagen	13
2.1.1 Empirische und historische Ausgangspunkte	13
2.1.2 Prinzipien der Dialektometrie	16
2.1.3 Grenzen des dialektometrischen Variationsbegriffs	17
2.2 Multivariate Sichtweisen	19
2.2.1 Erweiterung des Modells	19
2.2.2 Dynamische Theorien	20
2.2.3 Komplexität, Emergenz und Zufall	22
2.2.4 Genese und Wandel des Systems	25
2.2.5 Illusionen der Trennschärfe und der Reinheit	28
2.2.6 Rekurs: Die Bedeutung der Einzelvariablen	29
2.3 Chancen einer stochastischen Perspektive	31
2.3.1 Desiderat	31
2.3.2 Probabilisierung der Variationsdimensionen	32
2.3.3 Varietätenkonstruktion	36
3 TECHNIK: STOCHASTISCHE BILDANALYSE	39
3.1 Datengrundlage und vorbereitende Schritte der Datenverarbeitung	39
3.1.1 Quellen und Inhalte der Beispieldaten	41
3.1.2 Datenbankdesign	43
3.1.3 Gewichtung	44
3.1.4 Skalenniveau und Kategorisierung der Varianten	44
3.1.5 Distanzmaße	48
3.1.6 Dichteschätzung	51
3.1.7 Generierung und Visualisierung verschiedener Kartentypen	53
3.1.8 Zusammenfassung: Gesamtprozess und Nutzen	58
3.2 Statistische Werte der Karten	59
3.2.1 Komplexität, Kompaktheit und Homogenität	60
3.2.2 Kovarianz	63

3.3 Clustern von Sprachkarten basierend auf räumlicher Ähnlichkeit ..	65
3.3.1 Was ist Clustern (und welchem Zweck dient es)?	65
3.3.2 Typen von Clusteralgorithmen	66
3.3.3 Technisches Vorgehen beim unscharfen Clustern von Sprachkarten	71
3.3.4 Interpretationsmöglichkeiten der Ergebnisse	74
3.4 Faktorenanalyse	79
3.4.1 Grundlagen	80
3.4.2 Faktorenzahl und Projektimplementierung	81
3.4.3 Resynthese aus Faktoren	83
4 SUBKORPORA	84
4.1 Wortschatz	84
4.1.1 Kennwertanalysen	84
4.1.2 Raumstrukturelle Ähnlichkeit der Kategorien (Cluster)	85
4.1.3 Geografische Grundstrukturen (Faktoren)	91
4.1.4 Fazit	98
4.2 Lautung	99
4.2.1 Kennwertanalysen	99
4.2.2 Raumstrukturelle Ähnlichkeit der Kategorien (Cluster)	103
4.2.3 Geografische Grundstrukturen (Faktoren)	106
4.2.4 Fazit	116
4.3 Formen	116
4.3.1 Kennwertanalysen	118
4.3.2 Raumstrukturelle Ähnlichkeit der Kategorien (Cluster)	121
4.3.3 Geografische Grundstrukturen (Faktoren)	124
4.3.4 Fazit	132
5 GESAMTSYSTEM UND VARIETÄTENKONSTRUKTION	133
5.1 Vergleich der Subkorpora	133
5.1.1 Kennwerte	133
5.1.2 Systematische Verzerrungen	137
5.1.3 Relationen zwischen den Subkorpora	138
5.2 Konstruktion und Gestalt eines Gesamtsystems	141
5.2.1 Perspektivenabhängigkeit der Dialektkonstruktion	141
5.2.2 Globale Faktorenanalyse	144
5.2.3 Frequenzkorrigierte Faktorenanalyse	147
5.3 Tiefenvielfalt	149
5.3.1 Latente Variation	149
5.3.2 Mischverhältnisse der Orte	150
5.3.3 Typischste Variablen	152

6 ERGEBNISSE, KONSEQUENZEN UND PERSPEKTIVEN	156
6.1 Zum Sprachraum Bayerisch-Schwaben	157
6.1.1 Norden	157
6.1.2 Flüsse: Donau und Lech	158
6.1.3 Westen	159
6.1.4 Zentrum	160
6.1.5 Städte	160
6.1.6 Osten	161
6.1.7 Süden	162
6.2 Rückwirkungen auf Theorie und Methode	162
6.2.1 Der Abschied von der Isoglosse	162
6.2.2 Prototypen	163
6.2.3 Reihenschritte	164
6.2.4 Perspektiven für Simulationsstudien	164
6.2.5 Die kritische Masse	165
6.2.6 Selbstkritik: Der Nimbus der Zahl	166
6.3 Beiträge zur Qualitätssicherung in der Variationslinguistik	167
6.3.1 Skizzierung des Desiderats	167
6.3.2 Defektive Datenlage	168
6.3.3 Bestimmung der Güte der Verfahren	168
6.3.4 Probabilisierung von Störgrößen	170
6.3.5 Ermittlung und Subtraktion von Störfaktoren	172
6.3.6 Rauschgeminderte (Re-)Konstruktion der Gesamtvariation	176
7 FAZIT	178
LITERATURVERZEICHNIS	179
ANHANG 1: GRUNDKARTE UND ORTSLISTE DES SBS	189
ANHANG 2: FARBABBILDUNGEN	193
REGISTER	213

VORWORT

Von der Mühsal meiner Arbeit will ich nicht sprechen.
(Jorge Luis Borges, *Die Inschrift des Gottes*)



Augsburg, Sommer 2014

1 EINLEITUNG UND ÜBERSICHT

Praktisch alle modernen Wissenschaften suchen nach Mustern im Chaos, nach Regelmäßigkeiten in den Daten, die sie erheben oder erzeugen – das gilt im selben Maße für die Astrophysik, wie es für die Variationslinguistik gilt.

Diese Arbeit, ein Resultat des DFG-geförderten Projekts Neue Dialektometrie mit Methoden der stochastischen Bildanalyse (in Zusammenarbeit des Lehrstuhls für Deutsche Sprachwissenschaft der Universität Augsburg und des Instituts für Stochastik der Universität Ulm),¹ bildet keine Ausnahme. Sie thematisiert dies im Rahmen eines variationslinguistischen Grundproblems, der Konstruktion von ganzen Varietäten bzw. Dialekten, im Spannungsfeld zweier scheinbar unvereinbarer Blickwinkel: Wer sich mit Variation beschäftigt, beschäftigt sich – solange er auf empirisch abgesichertem Terrain bleiben will – mit einer Vielzahl an Einzelphänomenen, die sich idiosynkratisch, bisweilen chaotisch und damit in größerer Anzahl unüberschaubar zeigen. Demgegenüber ist der Blick auf Variation per se im Regelfall globaler Natur und will bzw. darf sich nicht auf Einzelphänomene einschränken; das Interesse gilt hoch- und höchstrangigen Ordnungen, wobei notwendigerweise hohe Grade an Abstraktion über die eigentlichen Daten gebildet werden müssen. Für sich genommen kann keiner dieser beiden Blickwinkel ein befriedigendes Gesamtbild von sprachlicher Variation zeichnen. Der Kompromiss beider Herangehensweisen, der in dieser Arbeit angestrebt wird, ist durch einen wahrscheinlichkeitsbasierten Zugang möglich, der weiterhin das Gesamtbild der Varietät(en) im Blick hat, dabei aber ihre jeweilige interne Variabilität transparent hält, indem er in jedem Analyseschritt die Ausprägungen der einzelnen Sprachphänomene klar registrieren und benennen kann.

Der Text gliedert sich grob in einen theoriegeleiteten Teil (Kapitel 2), einen technik- und methodenorientierten Teil (Kapitel 3) sowie einen anwendungsbezogenen Teil (Kapitel 4 und 5), gerahmt von diesem Einleitungskapitel sowie einem konzentrierten Ergebnis- und Konsequenzenkapitel (Kapitel 6) und einem knappen abschließenden Fazit (Kapitel 7).

Der Theorieteil bietet einen Forschungsüberblick über relevante Sprachvariations- und Sprachwandelmodelle nebst den jeweils damit verbundenen technischen Aspekten. Der Schwerpunkt liegt darauf, die Komplexität und innere Variabilität des Gegenstandes zu verdeutlichen sowie bestehende Ansätze zur Beschreibung und Erklärung seiner vielfältigen Gestalt zu präsentieren. Dabei ist es das Ziel, darzulegen, was eine geostatistische Herangehensweise dazu beitragen kann, bestehende Lücken in Theorie und Praxis zu schließen.

1 Mein herzlicher Dank gilt neben allen engagierten Beteiligten dieses Projekts und den Korrekturlesern älterer Fassungen dieses Texts den Herausgebern der Reihe sowie dem Bezirk Schwaben, der die Arbeit mit dem *Förderpreis des Bezirks Schwaben 2014* bedacht hat.

Im methodenorientierten Teil werden Prinzipien und Verfahren der stochastischen Bildanalyse vorgestellt, die als Grundlage für die später angewandten Techniken dienen und es erlauben sollen, Variationsmuster zu quantifizieren und miteinander zu vergleichen. Zunächst beinhaltet dies die Umlegung der für variationslinguistische Unternehmungen wie etwa Sprachatlanten punktuell erhobenen Informationen in eine wahrscheinlichkeitsbasierte, flächige Interpretation. Das wird durch Schätzung der Auftretenswahrscheinlichkeit jeder Variante an jedem Ort erzielt. Mittels exemplarischer Anwendung auf das Material des SBS („Sprachatlas von Bayerisch-Schwaben“) zeigt das Kapitel weiterhin, wie aus diesen Flächenkarten individuelle Werte ausgelesen werden können, die es ermöglichen, Karten in objektiverer Art und Weise miteinander zu vergleichen, als es der rein optische Eindruck leisten kann. Diese Werte bilden dann die Basis für weiterführende Analysen: So wird anschließend gezeigt, wie mit automatisierten Verfahren Gruppen von Sprachkarten gebildet werden, die eine ähnliche Raumstruktur aufweisen, und latente gemeinsame Strukturen, sogenannte Faktoren, aus großen Mengen an Karten ermittelt werden können.

Der anwendungsorientierte Teil der Arbeit ist in zwei Teilbereiche untergliedert: In Kapitel 4 werden jeweils einzeln Wortschatz, Lautung und Formen des Bayerisch-Schwäbischen Raums mittels der im Vorfeld erarbeiteten Techniken analysiert. Kapitel 5 zeigt das Gesamtvarietätengefüge, also Zusammenhänge und Unterschiede zwischen den einzelnen sprachlichen Systemteilen (Wortschatz, Lautung, Formen). Somit können Aussagen über die Gesamtgestalt einer Varietät getroffen bzw. die Einflüsse einzelner Teilmengen auf die Gesamtvariation ermittelt werden. Form, Kontur und Inhalt einer Varietät werden gleichzeitig als transparente, graduelle Struktur darstellbar und nicht als dichotomisches Konstrukt. Die Ergebnisse bestätigen die Vermutung, dass das klassische Konzept der „Dialekteinteilung“ selbst problematisch ist: Gestalt und Ertrag einer Einteilung unterscheiden sich zwischen den sprachlichen Teilsystemen (Wortschatz, Lautung, Wortbildung, Syntax) teils deutlich.

Vom Deskriptiven über das Inferierende hin zur Qualitätssicherung und Kritik fortschreitend werden zu guter Letzt in Kapitel 6 die Ergebnisse des Anwendungsteils für den Bayerisch-Schwäbischen Raum gebündelt dargestellt sowie Konsequenzen für die Theoriebildung im Bereich Variation und Wandel gezogen. Zum Abschluss bietet das Kapitel Ansätze zu einer variationslinguistischen Fehlertheorie: Einerseits kann die Intensitätsschätzung als ein Mittel zur Qualitätssicherung des Datensatzes interpretiert werden. Zusätzlich können einzelne Faktoren auch aus der Gesamtvariation entfernt werden – dies gilt sowohl für „echte“ Faktoren als auch für Störfaktoren, die während der Erhebung oder Bearbeitung der Daten entstehen.

2 HERANFÜHRUNG: VOM EINZELBELEG ZUR VARIETÄT

Dieses einführende Kapitel dient zweierlei Zwecken: Zum einen stellt es variationslinguistische Theorien, Modelle und Methoden dar, die genutzt wurden und werden, um den sowohl theoretischen als auch methodischen Spagat zwischen der empirischen Erhebung einzelner Sprachproben und der abstrahierenden Konstruktion kompletter Varietäten zu meistern. Der Schwerpunkt liegt auf dem Zusammenspiel zwischen einer eher qualitativen und einer eher quantitativen Forschungsperspektive, das nötig ist, um einen Gegenstand zu fassen, der im Laufe der Abstraktion vom Einzeldatum zum Gesamtsystem zunehmend komplexere Strukturen zeigt. Zum anderen wird mittels einer kritischen Beleuchtung dieser bestehenden Ansätze hergeleitet, was durch einen wahrscheinlichkeitsbasierten Blickwinkel auf Variation und Wandel für Theorie und Methode gewonnen werden kann.

2.1 GRUNDLAGEN

2.1.1 Empirische und historische Ausgangspunkte

Die grundlegende empirische Arbeit von Variationslinguisten bzw. Dialektologen ist die (direkte oder indirekte) Erhebung einzelner Variablen. Doch die Beschäftigung mit Einzelvariablen ist mühsam und im Regelfall lediglich ein Mittel zum Zweck: Man will klar voneinander unterscheidbare Gegenstände oder Klassen, hochrangige Strukturen, sich wiederholende Muster in seiner Umwelt erkennen. Die zentrale, hochrangigste Struktur der regionalen Sprachvariation, der Regiolekt oder Dialekt, existiert jedoch nicht als direkt messbare Entität – er muss aus empirischen Messungen vieler Einzelvariablen konstruiert werden.² Schon die Ortsgrammatiken in der Nachfolge der junggrammatischen Schule, mit denen das Studium der geografischen Sprachvariation in den deutschsprachigen Ländern im modernen Sinn beginnt, zeigen zwangsläufig diese Dualität: Die Forscher erheben einzelne Variablen, letztlich aber zu dem Zweck, ein Gesamtbild der Mundart eines ausgewählten Orts zu zeichnen (vgl. Abbildung 1).

- 2 Das rezente (Wieder-)Aufleben des wissenschaftlichen Interesses an Laienmeinungen zu Dialekten und Dialektgrenzen im Rahmen der *perzeptiven Dialektologie* oder *Wahrnehmungsdialektologie* (vgl. z. B. – mit jeweils unterschiedlichen Ausrichtungen – ANDERS 2010, KÖNIG 2010a und PURSCHKE 2011) kann ebenfalls als Lösungsvorschlag zu diesem Dilemma interpretiert werden: Statt Einzelvariablen werden von vornherein (mentale) Gesamtkonzepte ermittelt.

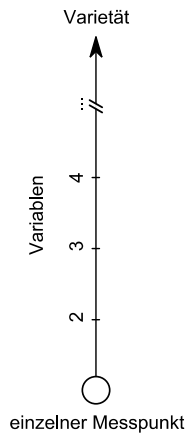


Abb. 1: Grundlegendes Variationsmodell (eine Dimension).

Die über den einzelnen Ortsdialekt hinausgreifende, traditionelle Dialektgeografie stellte – in Analogie zu anderen Disziplinen, die als Untersuchungsvariable (oder Projektionsfläche) den Raum nutzen – die Ergebnisse ihrer Erhebungen in der Form von Karten dar, die zunächst jeweils ein Einzelphänomen zeigten. Präferierte Methoden waren (und sind) dabei die Darstellung als a) Punktsymbolkarte und / oder b) Isoglossenkarte. Für jeden Ort ermittelt man also für jede Variable von Interesse die lokale(n) Variante(n) und stellt diese mittels eines jeweiligen Symbols direkt am Erhebungsort dar (siehe Abbildung 2a). Darstellung ist immer auch eine Form von Interpretation; diese Form der Darstellung ist dabei relativ treu zu den tatsächlich erhobenen Daten, sie visualisiert Belege direkt an den entsprechenden Erhebungsorten. Dennoch suggeriert sie bereits einiges – ähnliche Symbole legen Ähnlichkeit der Varianten nahe, das Gehirn konstruiert aus gleichen, eigentlich punktuellen Symbolen Flächen etc. (vgl. ausführlicher PRÖLL 2011). Die Isoglossenkarte (siehe Abbildung 2b) interpretiert stärker: Sie konstituiert eine Grenze zwischen einzelnen Varianten, hier steht also nicht mehr der empirisch ermittelte Einzelbeleg im Vordergrund, sondern die Flächenbildung und die angenommenen Grenzen der Verteilung der Varianten – *angenommen* daher, weil die jeweilige Grenze nicht gemessen wurde bzw. als Entität überhaupt nicht messbar ist. Die Isoglosse³ ist ein rein virtueller, dichotomischer Schnitt durch den Raum mit einer klaren Trennung beider Seiten voneinander.

3 RUOFF (1980, 94–95) legt stichhaltig dar, dass der Terminus selbst „erstens unpraktisch [...] und zweitens im Grunde falsch“ ist. Ich teile diese Ansicht, verwende ihn aber hier dennoch: a) aus Konsistenzgründen, weil sich seine Alternativen in der Literatur bislang nicht durchgesetzt haben und b) weil ich im Folgenden insgesamt gegen das dahinterstehende Konzept argumentiere.

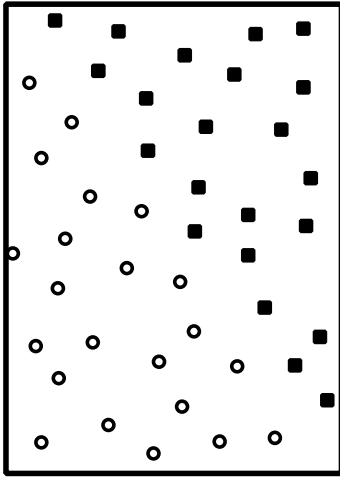


Abb. 2a: Idealierte Punktsymbolkarte.

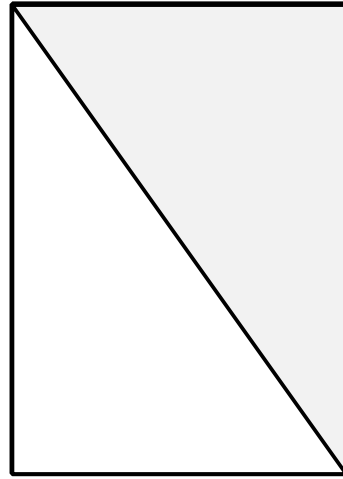


Abb. 2b: Idealierte Isoglossenkarte.

Schon früh ist Mundartforschern/Dialektologen zweierlei aufgefallen: Zum einen, dass Grenzen zwischen Phänomenen nicht immer absolut sind, sondern dass es Übergangsgebiete (welcher Natur auch immer; darauf wird später zurückzukommen sein) gibt, vgl. etwa die folgende, typische Aussage aus GLUTH / LOMPA / SMOLKA (1982, 487–488), dass „die Vorstellung einer Grenzlinie überhaupt unangemessen sein kann, nämlich dann, wenn der Gebietsrand eigentlich als Übergangszone angesehen werden müßte“. Zum anderen, dass die Darstellung von idiosynkratischen Grenzen einer einzelnen Variable keine relevante globale Aussage über den anvisierten Untersuchungsgegenstand bieten kann, oder anders gesagt, dass es ein massives Repräsentativitätsproblem dabei gibt, eine Variable als aussagekräftig für eine gesamte Varietät anzusehen. Das führt zu dem Zwang, eine oder mehrere Variablen auswählen zu müssen, die geeignet erscheinen, die Varietät als Ganzes zu repräsentieren (vgl. RUOFF 1980, 98–99). Ohne robuste Kriterien wäre diese Auswahl eine reine ad-hoc-Entscheidung – aber auch diese Kriterien selbst bleiben wohl grundsätzlich tautologischer Natur. In jedem Fall fällt der Großteil der Gesamtvariation unter den Tisch.

Zunächst wurde versucht, diesem Repräsentativitäts- / Selektionsproblem durch Kombinationskarten beizukommen, wobei die Isoglossen mehrerer Variablen in eine einzige Karte gezeichnet werden. Dadurch zeigen sich Isoglossenbündel (vgl. GIRNTH 2010, 112–116), was in der Gesamtvariation der kombinierten Phänomene Übergangsbereiche sichtbar macht.⁴ Diese Übergangsgebiete zeigen aber lediglich Überschneidungen zwischen Ausprägungen verschiedener Variablen, keine Übergangsgebiete von Varianten innerhalb einer Variablen: Die jeweiligen einzelnen Gebiete der kombinierten Varianten bleiben dichotomisch,

4 Das bekannteste Beispiel für eine derartige Karte stellt wohl die viel zitierte (bzw. abgebildete) Karte zur Einteilung der deutschen Dialekte in WIESINGER (1983a) dar.

was dazu führt, dass die durch die dichotomische Isoglossentechnik mangelnde Validität der Variantendarstellung nicht nur bestehen bleibt, sondern sich durch die Überlagerung sogar noch anhäuft (vgl. ähnlich PRÖLL 2014). Schon PAUL ([1920] 1995, 42–43) gibt überdies zu bedenken, dass die Gesamtvariation sich dergestalt nicht ökonomisch abbilden lässt:

Ziehen wir daher in einem zusammenhängenden Sprachgebiete die Grenzen für alle vorkommenden dialektischen Eigentümlichkeiten, so erhalten wir ein sehr kompliziertes System mannigfach sich kreuzender Linien. Eine reinliche Sonderung in Hauptgruppen, die man wieder in so und so viele Untergruppen teilt u. s. f., ist nicht möglich. (PAUL [1920] 1995, 42–43)

Für größere Datenmengen ist demnach eine deutliche Abstraktion bzw. Datenreduktion nötig. Konsequenz dieser Erkenntnis war die Ausprägung einer dezidiert quantitativen Variationslinguistik bzw. Dialektologie.

2.1.2 Prinzipien der Dialektometrie

Im vorliegenden Kontext sind hier zunächst insbesondere diejenigen Bereiche der quantitativen Dialektologie relevant, die seit den in der Nachfolge HAAGS (1898) entstandenen Arbeiten SÉGUYS (1971, 1973) und GOEBLS (1982, 1984) unter dem Terminus *Dialektometrie*⁵ rangieren.

Die große Innovationsleistung der Dialektometrie besteht darin, mittels Zusammenfassung von Einzeldaten ein Gesamtbild erzeugen zu können: Die Datengrundlage, auf der der Großteil an dialektometrischen Untersuchungen beruht, ist das *Aggregat* der Dialektdaten, die zusammengefasste Gesamtvariation.

Technisch umgesetzt wird dies unter anderem durch die Erstellung einer sogenannten *Ähnlichkeitsmatrix*. Dazu zählt man für jedes Paar von Orten, wie viele Variablen in beiden Orten die gleiche Variante haben. Dieser Grad an Übereinstimmung, in der Terminologie GOEBLS (1984) der *relative Identitätswert* (RIW), ist ein reziprokes Maß, skaliert zwischen 0 und 1 (bzw. zwischen 0 und 100 %): Er lässt sich dadurch problemlos kartieren.⁶ Entsprechend lässt sich auch eine *Distanzmatrix* gewinnen, die den Grad an Verschiedenheit der Orte bündelt. Weitere Verfeinerungen dieser grundlegenden Aggregationstechniken (wie etwa Gewichtungen der Matrix, vgl. z. B. GOEBL 2005) existieren und werden mit Erfolg eingesetzt – die Feinheiten sind hier im Moment aber nicht relevant.

Aus diesem Aggregat der Daten werden darüber hinaus durch sogenanntes *Clustern* Orte zu Gebieten zusammengefasst. Beim Clustern werden Daten (hier die aggregierte Gesamtvariation) mittels robuster Kriterien in Klassen aufgeteilt

5 Eine terminologische Anmerkung in eigener Sache: Die Methoden, die im nachfolgenden Kapitel 3 vorgestellt werden, sind im strengeren Sinn nicht durchgängig als Dialektometrie zu apostrophieren, fußen aber zum Teil auf entsprechenden Grundüberlegungen zum Nutzen quantitativer Techniken, wandeln typische Verfahren der Dialektometrie ab bzw. haben den Anspruch, Probleme zu überwinden oder zu vermeiden, die insbesondere die Dialektometrie aufwirft.

6 Siehe dazu genauer Abschnitt 3.1.5, ab Seite 48.

(siehe dazu insbesondere die detailliertere Einführung in Clustertechniken in Abschnitt 3.3.1, ab Seite 65; die genauen technischen Details spielen hier noch keine Rolle). Als Ergebnis erhält man „Dialektregionen“, Gruppen von Orten, die sich ähneln, weil sie sich jeweils in einer gewissen Anzahl an Merkmalen entsprechen.

Ziel der Analyse des Aggregats ist es, die „höchstrangigen Strukturen“ der gemessenen Daten aufzudecken und Strömungen und Grundmuster zu finden, die die Gesamtheit der Variabilität der untersuchten Daten adäquat erfassen. So stellt sich GOEBL (2005, 509) mittels Aggregation explizit gegen „Wortschatzstudien [...] im Zeichen kasuistischer Einzelanalysen“:

Demgegenüber belegen die hier gezeigten synthetischen Resultate, daß es möglich ist, durch die konsequente Addition vieler Einzel-„Geschichten“ zu einer Global-„Geschichte“ der lexikalischen Bewirtschaftung einer bestimmten Gegend durch die betreffenden Sprecher vorzustoßen bzw. diese zu rekonstruieren. (GOEBL 2005, 509)

Die Anzahl an Messpunkten ist endlich, die Gesamtzahl an Variablen im Sprachsystem dagegen unbegrenzt. GOEBL (2005, 500) ist davon überzeugt, dass sich ab einer Überlagerung von ca. 200–300 Variablen „die Gestalt (Musterung, Struktur etc.) von dialektometrischen Synthesen [...] nur mehr wenig verändert“. Die individuellen Strukturen einzelner Variablen würden also spätestens ab dieser Größenordnung keine große Rolle mehr spielen;⁷ die räumliche Gestalt des angenommenen Gesamtsystems könnte aus einer bewältigbaren Menge an Einzeldaten inferiert werden. Je nach Blickrichtung und Fragestellung kann es nun als Stärke oder als Schwäche aggregativer Methoden gesehen werden, dass wiederkehrende, dominante Strukturen sich gegen Ausreißer in den Daten durchsetzen (positiver Sichtwinkel) bzw. Einzelphänomene in der Gesamtvariation untergehen (negativer Sichtwinkel).

2.1.3 Grenzen des dialektometrischen Variationsbegriffs

Eine Varietätenkonstruktion nach diesem Vorgehen ist aber nicht frei von weiteren Reduktionen; so beschäftigt sich die ursprüngliche Art der dialektometrischen Analyse, wie sie etwa in GOEBL (1982, 1984, 2005) vertreten wird, nicht mit der Rolle des Individuums bzw. der Sprechergemeinschaft als Träger der Variation. Bei dialektologischen Erhebungen wird naturgemäß versucht, Störgrößen gering zu halten, indem man – soweit möglich – die Sozialparameter der Gewährspersonen kontrolliert. Hauptsächlich dient das (neben rein erhebungspraktischen Vorteilen, vgl. KÖNIG 1988b, 173) dazu, die Messungen der einzelnen Orte untereinander vergleichbar zu machen. Die Belege sind daher im Idealfall repräsentativ für die Alters- und Sozialstruktur, nach der die Gewährspersonen ausgewählt wurden.⁸ Das heißt aber noch nicht, dass die Personen eines Ortes, die den Aus-

7 Unerwähnt bleibt dabei jedoch, dass dazu Grundanforderungen an diese Daten zu stellen sind. Die Abschnitte 5.2.1 (ab Seite 141) und 6.2.5 (ab Seite 165) gehen darauf näher ein.

8 Aber sie sind natürlich mitnichten repräsentativ für den gesamten Ort, der eine völlig andere – und in keinem Fall homogene – Alters- und Sozialstruktur aufweisen kann. Ein solches Pos-

wahlkriterien entsprechen, auch sprachlich eine homogene Gruppe bilden müssen. Es folgt daraus auch nicht, dass jede dieser Personen nur Träger einer einzigen Variante sein kann. Laut GOEBL (1997, 23) interpretiert die Romanistik das Auftreten von mehreren Belegen an einem Ort meist als „the beginning of a sociolinguistic disintegration of the local dialect or geolinguistic stratum under consideration“. Dies führt nun aber zur Annahme einer Homogenität der Ortsdialekte, deren Grundüberzeugung es ist, dass es nur eine repräsentative Form pro Ort gäbe, die ausnahmslos von jedem repräsentativen Sprecher im Ort genutzt würde: ein Ort = eine Gewährsperson = ein Beleg.⁹ Sowohl die Repräsentativität der Gewährspersonen als auch die des Belegs wird absolut gesetzt; folglich findet sich z. B. in den Daten der romanischen Atlanten, auf denen die Analysen in GOEBL (1984) beruhen, praktisch nirgends ein Ort mit mehreren Belegen. Die dafür schon in der Erhebungsphase zwangsläufig nötige Vorselektion schafft zwar eindeutige Karten, zeichnet aber ein Gesamtbild der Variation, das weniger treu zur tatsächlich beobachtbaren Variation ist.¹⁰ Treten dennoch mehrere Antworten an einem Ort auf, müsste dies als „Verunreinigung“ des Datensatzes gesehen werden – die Gewährsperson wäre kein kompetenter Sprecher der ansässigen Varietät, der Explorator hätte die falschen Fragen gestellt – oder der Forscher beschäftige sich mit „a blending of sociolinguistic and geolinguistic information“ (GOEBL 1997, 28).

Diese Annahmen korrespondieren mit einer (implizit) statischen Sprachsystemanschauung sowie einer (wiederum oft impliziten) problematischen Dichotomie zwischen Synchronie und Diachronie.¹¹ Das System Sprache scheint über das Individuum hinaus homogen, stabil und außerdem als Objekt direkt messbar zu sein. Der Messpunkt ist „synchroner“ Natur, aber auf eine a historische Art, als statischer Zustand. Die Diachronie des Systems hätte man sich als lineare Abfolge a-historischer, synchroner Schnitte vorzustellen. Ein derartig konzipierter, statischer Variationsbegriff konstruiert aber ein System, das a) von seinen Nutzern unabhängig ist und b) als Zustand beschrieben wird. Ob man die individuelle Va-

tulatur würde auch die Ausrichtung der meisten Dialektatlanten verkennen, die gar nicht in der Absicht erstellt wurden, eine sozial repräsentative Sicht des aktuellen Sprachgebrauchs an einem Ort zu leisten, sondern den ältesten greifbaren, im Verschwinden befindlichen Dialekt zu dokumentieren.

- 9 In einigen Erweiterungsarbeiten zur grundlegenden Dialektometrie wird das nicht nur kritisch gesehen, sondern auch als Ausgangspunkt für die technische Berücksichtigung von Mehrfachbelegen an einem Ort herangezogen, vgl. NERBONNE / KLEIWEG (2003); PEREA / UEDA (2010); NERBONNE / KRETZSCHMAR (2013); AURREKOETXEA / FERNANDEZ / RUBIO / RUIZ / SÁNCHEZ (2013) (mit einem kurzen Forschungsüberblick). Das Repräsentativitätspostulat der Gewährsperson bleibt davon aber unberührt.
- 10 In diesem Zuge muss allerdings betont werden, dass im ursprünglichen Rahmen der Dialektometrie aus Gründen der Reliabilitätssicherung direkt die *Sprachatlanten* als Gegenstand fungieren, nicht ihr Interpretament (vgl. GOEBL 1982, 15; GOEBL 1984, 5, 18–19). BAUER (2009, 172) jedoch löscht auch bereits kartierte Mehrfachbelege für die Aggregation.
- 11 Man vergleiche dagegen HAAS (1978, 2): „[W]enn in irgendeiner Sparte der Sprachwissenschaft die prinzipielle Trennung von Synchronie und Diachronie sich als unangemessen erweist, dann in der Mundartforschung.“

riation und Interaktion im System nun als Soziolinguistik apostrophiert oder nicht,¹² es ist riskant, sie als eigene Variationsebene neben der geografischen Variation zu betrachten (vgl. dazu besonders BERTHELE 2004, mit einem wissenschaftshistorischen Überblick). Eine alternative Betrachtungsweise könnte etwa konstatieren, dass individuelle Variation erst dafür sorgt, dass geografische Variation entsteht: „The particular structure that the linguistic space has is the consequence of the language activity of people“ (AURREKOETXEA 2010, 208).

2.2 MULTIVARIATE SICHTWEISEN

2.2.1 Erweiterung des Modells

Spätestens an diesem Komplexitätsstatus sprachwissenschaftlicher Modellbildung hat man es demnach mit (mindestens) zwei Dimensionen von Variation zu tun, die zu berücksichtigen sind (siehe Abbildung 3). Erstens: Die Variation kann nur in Form von Einzelvariablen gemessen werden, der anvisierte Gegenstand dieser Dimension ist aber die Gesamtheit der Variablenausprägungen oder zumindest strukturell zusammenhängender Teilaspekte. Diese Gesamtheit spiegelt aber lediglich den Idiolekt eines einzelnen Sprechers wider. Zweitens: Messpunkte der Variation sind Individuen, Aussageziel aber das Gesamtsystem aller Sprecher einer Varietät.

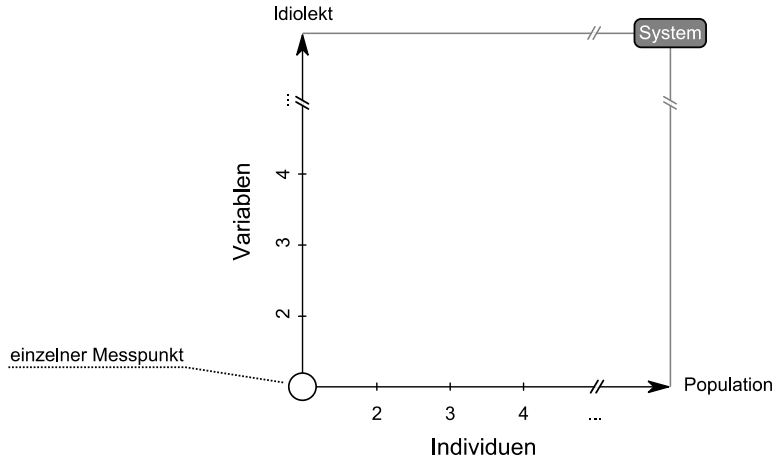


Abb. 3: Erweitertes Variationsmodell (zwei Dimensionen).

12 Vgl. etwa die Aussage von AURREKOETXEA / FERNANDEZ / RUBIO / RUIZ / SÁNCHEZ (2013, 26) in Juxtaposition zu GOEBL (1997): „Any speaker can hear more than one word or allophone for a concrete concept and can use it in any circumstance, without taking into account sociolinguistic factors or motivations“; grundsätzlich so auch WEINREICH / LABOV / HERZOG (1968, 101). Man vergleiche auch die differenzierte Diskussion des amerikanischen Arms der Dialektometrie, etwa durch KRETZSCHMAR / SCHNEIDER (1996, 5–16).

Das Gesamtsystem ist nicht direkt messbar, es wird von zwei Dimensionen aus erschlossen: Mittels der Erhebung vieler einzelner Variablen inferiert man den Idiolekt eines Individuums, mittels einer Vielzahl an Individuen inferiert man die Population. Das Gesamtsystem, den Sprachgebrauch der Population, inferiert man demnach mittels einer Vielzahl an von Individuen geäußerten Varianten.¹³

Diese Verknüpfung zwischen jeweils einer individuellen, qualitativ analysierbaren und einer höher strukturierten, nur quantitativ fassbaren Seite der Variation erfordert für den Empiriker eine entsprechende Berücksichtigung auf methodischer Ebene; es bedarf aber auch eines theoretischen Rahmens, der diese Aspekte angemessen würdigen kann. Dieses Desiderat führte zu den sogenannten *dynamischen* Theorien.¹⁴

2.2.2 Dynamische Theorien

Als repräsentative Vertreter der Familie der dynamischen Theorien zu Variation und Wandel greife ich hier exemplarisch CROFTS *evolutionären Sprachwandel* (CROFT 2000) sowie KELLERS *Theorie der unsichtbaren Hand* (KELLER 2003) heraus, da beide einen starken Fokus auf die Relation zwischen Individuum, Gesellschaft und System legen.

CROFTS Modell (am prominentesten dargestellt in CROFT 2000) adressiert exakt und pointiert die Relation zwischen Sprecher und System: „Languages don't change; people change language through their actions“ (CROFT 2000, 4); Wandel ist kein Prozess, der sich an einem isolierten Gegenstand (etwa der zum System verkürzten und abstrahierten „Sprache“) abspielt, sondern an einzelnen Äußerungen und individuellem sprachlichen Wissen (bzw. ihrer Weitergabe): „The evolution of both of these entities, particularly utterances, occurs through replication, not inherent change“ (CROFT 2000, 3; vgl. ähnlich auch PAUL [1920] 1995, 34). Auch KELLER will seine *Theorie der unsichtbaren Hand* explizit als evolutionäre Theorie verstanden wissen (vgl. KELLER 2003, 13–14), die sich allerdings im Gegensatz zu CROFTS *evolutionärem Sprachwandel* an sozialwissenschaftlich-ökonomischen Vorlagen orientiert, nicht an genetisch-biologischen. Die Anbindung an den Terminus der *Evolution* geschieht auch hier in klarer Abgrenzung zu SCHLEICHER (1873), der Sprache – ebenfalls in Analogie zu damals vorherrschenden wissenschaftlichen Strömungen – als Organismus betrachtet, dabei aber nicht

13 So schon PAUL ([1920] 1995, 24): „Das wahre Objekt für den Sprachforscher sind vielmehr sämtliche Äußerungen der Sprechfähigkeit an sämtlichen Individuen in ihrer Wechselwirkung auf einander.“

14 Der Übergang von „statischen“ zu „dynamischen“ Theorien ist natürlich fließend und nicht ohne Übergeneralisierungen zu treffen. ALTMANN (1985) sieht den Hauptunterschied in der Konzentration auf qualitative Konzepte einerseits und quantitative andererseits: Die statischen Theorien nutzen vermehrt lineare Modelle, um Zusammenhänge zu fassen, stützen sich auf Logik „or simply verbal descriptions“ (ALTMANN 1985, 183), während die dynamische Perspektive vermehrt Wahrscheinlichkeiten, stochastische Prozesse und systemtheoretische Überlegungen integriert (vgl. auch WILDGEN / PLATH 2005, 689).