

1 EINLEITUNG

Im Jahr 1969 führte WILLIAM LABOV das Konzept der „variablen Regel“ ein, das es erstmals erlaubte, Variation systematisch als integralen Teil der Sprache eines Sprechers oder einer Sprechergemeinschaft zu beschreiben (LABOV 1969). Bis dahin war sprachliche Variation – vor allem von der generativistischen Schule – oft als reines Performanzphänomen ohne systematischen Charakter gesehen worden. Fünf Jahre nach LABOVs Vorstoß wurde das Konzept der variablen Regel durch HENRIETTA J. CEDERGREN und DAVID SANKOFF in ein mathematisches Modell überführt, das sprachliche Äußerungen als statistische Stichproben aus einer probabilistischen Sprecherkompetenz behandelt. CEDERGREN/SANKOFF (1974, 353) „distinguish rule probabilities from rule frequencies, assigning the former to competence and the latter to performance“. Ihr Modell beschreibt sprachliche Variablen als Wahrscheinlichkeitsräume, die das Auftreten ihrer Varianten probabilistisch steuern. Die Wahrscheinlichkeiten werden dabei von inner- oder außersprachlichen Bedingungsfaktoren (sogenannten *constraints*) beeinflusst und manifestieren sich an der beobachtbaren Oberfläche als messbare relative Variantenfrequenzen. Die Kompetenz eines Sprechers wird so durch bedingte Auftretenswahrscheinlichkeiten von Varianten beschrieben. Die anfänglich vor dem Hintergrund des von CHOMSKY begründeten generativistischen Paradigmas entstandene, von LABOV und CEDERGREN/SANKOFF maßgeblich geprägte Theorie der Sprachvariation entwickelte sich zu einem neuen Paradigma, der jungen Soziolinguistik, die sich bald dezidiert gegen das Axiom der Generativisten stellte, sprachliche Variation sei ein reines Performanzphänomen (vgl. CEDERGREN/SANKOFF 1974, 352; BRITAIN 2010, 74). In der Soziolinguistik wird sprachliche Variation als integraler Bestandteil der Sprecherkompetenz und der Sprachproduktion betrachtet.

Während die Annahme von systeminhärenter Variation zum wesentlichen Merkmal der Sozio- bzw. Variationslinguistik schlechthin wurde, lebt das mathematische Modell von CEDERGREN/SANKOFF vor allem in Form des Analyseprogramms Varbrul und seiner Derivate weiter, die den Einfluss von *constraints* auf Variantenwahrscheinlichkeiten berechnen und so bei der praktischen Analysearbeit wertvolle Einsichten in das Verhältnis von Variablen zu ihrem inner- und außersprachlichen Kontext liefern. Die Beschäftigung mit den theoretischen Implikationen eines probabilistischen Variationsmodells sowie der Ausbau seiner Anwendung in den variationslinguistischen Teilbereichen führte dagegen eher ein Schattendasein;¹ insbesondere im Bereich der Dialektologie, die v. a. im anglo-amerikanischen Raum mittlerweile als ein Teilbereich der Soziolinguistik behandelt wird und sich ihren Methoden auch im deutschsprachigen Raum im-

1 Es gibt Ausnahmen (z. B. MENDOZA-DENTON/HAY/JANNEDY 2003). Im Bereich der quantitativen Linguistik wurden einige Modelle erarbeitet, die zumindest implizit probabilistisch arbeiten (z. B. ALTMANN et al. 1983).

mer mehr öffnet, fehlt bislang eine probabilistisch ausgerichtete Theorien- und Modellbildung.

Die vorliegende Arbeit nimmt konkrete dialektologische Fragestellungen zum Anlass, um diesen in meinen Augen überfälligen Schritt zu unternehmen und die Dialektologie mit einem probabilistisch-variationalinguistischen Modell auszustatten. Als konkrete Untersuchungsgegenstände dienen dabei die geographischen Verteilungen lexikalischer Varianten. Mit den wortgeographischen Karten des „Sprachatlas von Bayerisch-Schwaben“ (SBS) steht ein umfangreiches Kartenkorpus als Datengrundlage bereit (vgl. Kapitel 3), das insofern aus einem ‚klassischen‘ Dialektatlas stammt, als der SBS in erster Linie den zur Zeit der Erhebung ältesten erreichbaren Stand des Basisdialekts dokumentiert (vgl. SBS, Bd. 1, 15, 26). Die Methoden zur Vorverarbeitung und Auswertung der Daten wurden im Rahmen des DFG-finanzierten, interdisziplinären Forschungsprojekts *Neue Dialektometrie mit Methoden der stochastischen Bildanalyse* entwickelt, einer Kollaboration des Lehrstuhls für Deutsche Sprachwissenschaft, Universität Augsburg, und des Instituts für Stochastik, Universität Ulm (weitere in Zusammenhang mit dem Projekt stehende Publikationen sind: RUMPF et al. 2009; 2010; PICKL/RUMPF 2011; 2012; MESCHENMOSE/PRÖLL 2012a; 2012b; PRÖLL 2013; PICKL 2013; PRÖLL/PICKL/SPETTL [i. E.]; PICKL et al. [i. V.]).² Mit eigens implementierten und/oder entwickelten geostatistischen Methoden werden die Daten zunächst unter Berücksichtigung empirisch-statistischer und variationslinguistischer Grundlagen vorverarbeitet. Die so aufbereiteten Daten werden sodann mittels statistischer Tests ausgewertet, um den Einfluss von inner- und außersprachlichen Faktoren auf geographische Variantenverbreitungen im Korpus zu erhellen. Im Anschluss wird ein Data-Mining-Verfahren zur Identifikation von latenten Strukturen in den Daten angewandt. Auf der Grundlage der Ergebnisse folgt schließlich ein Versuch der probabilistischen Modellierung geolinguistischer Variation.

1.1 MOTIVATION

[...] dialectologists should not be content simply to *describe* the geographical distribution of linguistic features. They should also be concerned to *explain* – or perhaps, more accurately, to adduce reasons for – this distribution. Only in this way will we be able to arrive at an understanding of the sociolinguistic mechanisms that lie behind the geographical distribution of linguistic phenomena, the location of isoglosses, and the diffusion of linguistic innovations. (TRUDGILL 1974, 216–217)

In diesem Zitat fasst PETER TRUDGILL die Essenz dessen zusammen, was das Erkenntnisinteresse der Dialektologie bzw. der Geolinguistik³ ausmacht: die Identifikation von Mechanismen und Faktoren der räumlichen Verbreitung sprachlicher

2 Ich danke meinen Kollegen aus Augsburg und Ulm herzlich für die hervorragende Zusammenarbeit.

3 Obwohl ein nicht zu vernachlässigender Unterschied zwischen den Termini *Dialektologie* und *Geolinguistik* besteht, sind sie in dieser Arbeit an vielen Stellen austauschbar.

Varianten. Der Beitrag, den diese Arbeit auf diesem Gebiet leisten soll, lässt sich in drei Hauptpunkten zusammenfassen:

1. Auf der methodischen Ebene die Bereicherung der quantitativen Analyseverfahren der Geolinguistik um geeignete und sinnvolle Ergänzungen,
2. auf praktischer Ebene die Auswertung von empirischen Daten mit den entwickelten Methoden, um diese zu erproben und Erkenntnisse über die sprachgeographischen Verhältnisse im untersuchten Bereich zu gewinnen, und
3. auf theoretischer Ebene die Bereitstellung eines Deutungsrahmens für die Ergebnisse und die Methoden, der sie in den größeren Rahmen geo- und soziolinguistischer Theorie stellt. Besonderes Augenmerk liegt dabei auf der Anpassung probabilistischer Modelle aus der Soziolinguistik für die Geolinguistik.

Die Methoden müssen dabei so beschaffen sein, dass sie auf die Theorie bezogen und mit ihr vereinbar sind. Insbesondere wird sich zeigen, dass die diatopische Ausprägung des probabilistischen Variationsbegriffs bei CEDERGREN/SANKOFF, der die theoretische Seite dieser Arbeit bestimmt, ihr praktisches und geradezu natürliches Gegenstück in geostatistischer Methodik hat, die bisher in der Sprachgeographie wenig Anwendung fand (vgl. Kapitel 4). Im Jahr 1993 schrieben JAY LEE und WILLIAM A. KRETZSCHMAR JR. in ihrem Aufsatz „Spatial analysis of linguistic data with GIS functions“:

It is somewhat surprising that up to this point linguistic researchers have not maintained closer relations with other geographers except for cartographic assistance. During the 1980s geographers have achieved major advances in two areas: development of powerful GIS software and development of statistical models for analysis of geographical patterns of data, such as point pattern analysis, spatial autocorrelation, spatial filtering, and many others. Using this progress, a bridge between dialectologists and professional geographers can be built for mutual benefit. For those interested in language variation, finally there will be tools available for validation of areal boundaries and patterns to parallel those now in use to discuss social patterns. Geographers will be able to consider language data on the same terms as data from other fields. Finally, there is an opportunity here to use validated empirical evidence to improve our notions of dialects as well as to study differences in geographical distribution of single features. (LEE/KRETZSCHMAR 1993, 541)

Umso überraschender ist es, dass diese Idee in den 20 Jahren, die seit Erscheinen dieses programmatischen Artikels vergangen sind, praktisch nicht weiterverfolgt wurde und nur sehr sporadisch dialektologische Arbeiten erschienen sind, die sich geostatistische Methoden zunutze machen (vgl. auch HOCH/HAYES 2010, 24).⁴ Der Fokus der quantitativ arbeitenden Dialektologie scheint in den letzten Dekaden fast ausschließlich auf aggregative, klassifikatorische Verfahren gerichtet gewesen zu sein, bei denen die Geographie oft erst bei der kartographischen Darstellung eine Rolle spielt (vgl. 2.2). Das primäre Erkenntnisinteresse war es, auf quantitativem Wege Dialektregionen und -grenzen zu ermitteln. Sogar in dem oben zitierten Aufsatz wird die Einteilung des Raums in Sprachregionen als ultimatives Ziel der

4 Ausnahmen sind etwa LIGHT/KRETZSCHMAR (1996) oder GRIEVE (2009).

geostatistischen Methoden dargestellt, obwohl diese primär für die Analyse einzelner Variablen gedacht sind.⁵

Es lässt sich also konstatieren, dass in der Sprachgeographie Untersuchungen mit (geo-)statistischen Mitteln weiterhin unterrepräsentiert sind. Ein Schritt, die Anregung von LEE/KRETZSCHMAR aufzugreifen und fortzuführen, wurde mit dem Augsburg/Ulmer Forschungsprojekt *Neue Dialektometrie* (vgl. 1) unternommen. Die dort entwickelten Verfahren sind darauf zugeschnitten, die individuelle Variation einzelner Variablen zu bewahren, gesondert auszuwerten und zu vergleichen, um den durch Aggregation herbeigeführten Variationskollaps (vgl. 1.2.2, 4) zu vermeiden. Dadurch eröffnen sich völlig neue Möglichkeiten, eine Dimension der geolinguistischen Variation zu untersuchen, die bisher von quantitativen Methoden weitgehend unberührt geblieben ist: Die ‚Variation in der Variation‘, d. h. die Unterschiede, Gemeinsamkeiten und Besonderheiten der geographischen Verteilungen individueller sprachlicher Variablen und Varianten.

1.1.1 Die Variation in der Variation

Jede sprachliche Veränderung und mithin auch die Entstehung jeder dialektischen Eigentümlichkeit hat ihre besondere Geschichte. Die Grenze, bis zu welcher sich die eine erstreckt, ist nicht massgebend für die Grenze der anderen.

– HERMANN PAUL,
„Prinzipien der Sprachgeschichte“

Wenn man in erster Linie mit der traditionellen Dialektometrie, wie sie in Abschnitt 1.2.2 umrissen wird, vertraut ist, so ist man wahrscheinlich vor allem an der Abgrenzung und Einteilung von Dialekten interessiert und hat von Haus aus wenig Interesse an der Variation zwischen den räumlichen Verteilungen sprachlicher Variablen. Die Beschäftigung mit der Dialektgeographie erschöpft sich dann im Messen der Ähnlichkeit von Ortsdialekten, in der Feststellung von stärkeren und schwächeren Grenzlinien zwischen ihnen oder von Dialektgebieten verschiedener Größe. Verschiedenste Analysetechniken erlauben es, die aggregierten Dialektdaten zu visualisieren und auszuwerten. Damit, so könnte man meinen, ist die diatopische Variation im betrachteten Gebiet quantitativ umfassend beschrieben.

Jedoch übersieht man so folgenden Umstand, durch den überhaupt erst stärkere und schwächere Grenzen zwischen Dialekten, ähnlichere und unähnlichere Ortsdialekte zustande kommen können: Nicht alle sprachlichen Variablen, die im

5 „Once a significant areal pattern is detected for a combination of given words, one can attempt to draw boundaries between informants and communities to express agglomerations of occurrence and thus construct linguistic regions. In this fashion, one can look forward to boundaries that segregate different linguistic regions based on results of quantitative analysis which are statistically validated.“ (LEE/KRETZSCHMAR 1993, 554)

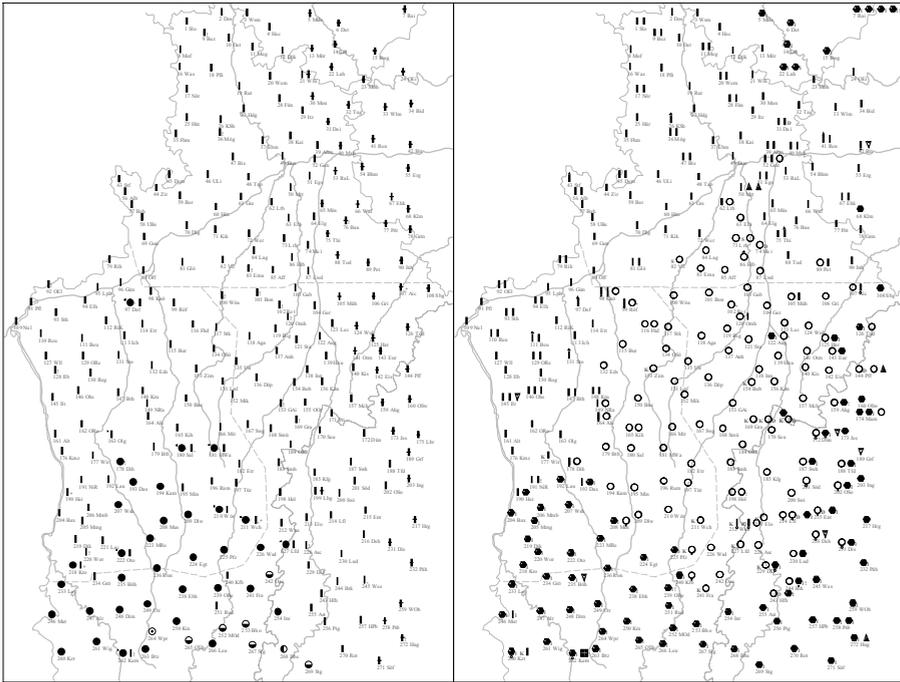


Abb. 1a: 'niesen' (K. 2.47)

Abb. 1b: 'dreieckige Papiertüte' (K. 10.53)

Abb. 1: Die Karten für 'niesen' und für 'dreieckige Papiertüte' aus dem SBS, entnommen aus dem „Digitalen Sprachatlas von Bayerisch-Schwaben“ (DSBS).

aggregierten Zustand für dialektometrische Grenzlinien verantwortlich sind, sind gleich im Raum verteilt; ansonsten gäbe es nur scharfe Grenzen, und links und rechts von ihnen gegenseitig komplett unverständliche Varietäten. Der Blick in einen Dialektatlas wie den SBS offenbart ein anderes Bild: Die meisten der Karten weisen völlig unterschiedliche Verteilungen auf, die Variation zwischen ihnen mutet geradezu chaotisch an (s. z. B. Abb. 1; vgl. FISCHER 1895, 80). Nur einige wenige wiederkehrende Strukturen lassen sich erkennen, z. B. bestimmte Verbreitungsareale einzelner Varianten, die mit auffälliger Häufigkeit wiederkehren, und die in ihrer Gesamtheit zu den Dialektgebieten beitragen, die die klassische Dialektometrie quantitativ erfasst. Der Normalfall ist jedoch (v. a. im Bereich der Lexik), dass jede Sprachkarte ihre eigene, individuelle Struktur aufweist. Diese Art der Variation, die geradezu unüberschaubare Variabilität zwischen den Verbreitungen einzelner sprachlicher Merkmale, stellt einen bisher vernachlässigten Forschungsgegenstand dar.

Doch allein die Tatsache, dass überhaupt wiederkehrende Muster auftreten, zeigt, dass Struktur im scheinbaren Chaos existiert. Dass zwischen den geographischen Verteilungen einzelner sprachlicher Variablen manchmal Zusammenhänge bestehen, ist seit Langem bekannt. Vor allem im Bereich der Lautgeographie weisen die Verbreitungen von systemisch verwandten Lauten oft fast identische Gren-

zen auf. Der Grund hierfür, der in gekoppeltem, sich im Raum manifestierendem Lautwandel ganzer Lautgruppen (beispielsweise bestimmter Diphthonge) besteht, wurde früh erkannt und 1918 von ANTON PFALZ erstmals unter dem Terminus *Reihenschritte* beschrieben (PFALZ 1918; vgl. auch WIESINGER 1982). In der strukturalistisch ausgerichteten Dialektologie ist die Untersuchung von sprachgeographisch relevanten, systembedingten Beziehungen zwischen sprachlichen Merkmalen ein Hauptbeschäftigungsfeld (vgl. GOOSSENS 1969; 1977, 89–101), wird jedoch bislang praktisch nicht mit quantitativen Methoden untersucht.

Solche im Sprachsystem begründeten Beziehungen zwischen sprachlichen Variablen sind am einfachsten festzustellen und am besten untersucht im Bereich der Phonetik bzw. Phonologie. Die artikulatorische Beschreibung von Lauten impliziert schon durch Kategorien wie *Artikulationsort* oder *Lippenrundung* systematische Zusammenhänge.⁶ Für die Relevanz solcher Beziehungen für die geographische Verteilung mag die neuhochdeutsche Diphthongierung als Beispiel dienen, in der sich alle hohen Langvokale des Mittelhochdeutschen geographisch konsistent zu den jeweils entsprechenden steigenden Diphthongen entwickelt haben. Dass andere, ebenfalls parallel laufende Entwicklungen systemisch verwandter Laute jedoch nicht zur geographischen Konsistenz prädisponiert sind, illustrieren in eindrucksvoller Weise die Ergebnisse der Zweiten Lautverschiebung, was die Frage aufwirft, wieso sich bei diesen speziellen Beispielen die Vokale einheitlich verhalten haben, die Konsonanten jedoch nicht.

Ohne die Unterschiede der geographischen Entwicklung in diesen Fällen kommentieren oder gar erklären zu wollen, bleibt festzuhalten, dass die Identifikation solcher Zusammenhänge bisher weitgehend auf die Phonologie beschränkt ist. Für das Lexikon galt im Allgemeinen bislang die Annahme, dass es im Vergleich zum Lautsystem zu wenig strukturiert sei, als dass sinnvolle Kategorien gefunden werden könnten, die für die Verteilungsgeographie von Bedeutung wären (vgl. auch KURATH 1972, 25): „Since the lexicon, with its thousands of items, is the least structured part of language, it is more susceptible to incidental variation than to systematic“ (FRANCIS 1983, 20). Diese Meinung ist offenbar vorherrschend (vgl. u. a. ULLMANN 1953, 225; GOOSSENS 1969, 53, 69; WAKELIN 1984, 70; VIERECK 1986, 725; LABOV/ASH/BOBERG 2006, 41, 119), obwohl oder möglicherweise gerade weil es in einigen Fällen einen geradezu offensichtlichen Zusammenhang zwischen der sprachlichen Relation zwischen zwei Variablen und der geographischen Ähnlichkeit ihrer Verbreitung gibt, wie etwa in Abb. 2. Die linguistische Beziehung zwischen den Variablen ‘Kamm’ und ‘kämmen’ ist offensichtlich, die Karten sind fast identisch. Es ist klar, dass die Beziehung hier so eng ist, dass sich die Verteilungen der beiden Variablen gegenseitig bedingen oder sogar als zwei Versionen

6 Ob es tatsächlich diese artikulatorischen Kategorien sind, die für die parallele Entwicklung von Lauten verantwortlich sind, ist hingegen fraglich; WALTER HAAS (1978, 21) weist darauf hin, dass die „Symmetrie der Phonemsysteme [...] nicht auf die Anatomie der Sprechwerkzeuge, sondern auf eine Besonderheit der menschlichen psychophysischen Struktur zurückgeführt werden“ muss. Er fährt fort: „Für den Linguisten genügt es, seine ‚Erklärungskette‘ bei der Feststellung abzubrechen, dass symmetrische Organisation besteht und im Sprachwandel eine entscheidende Rolle spielt“.

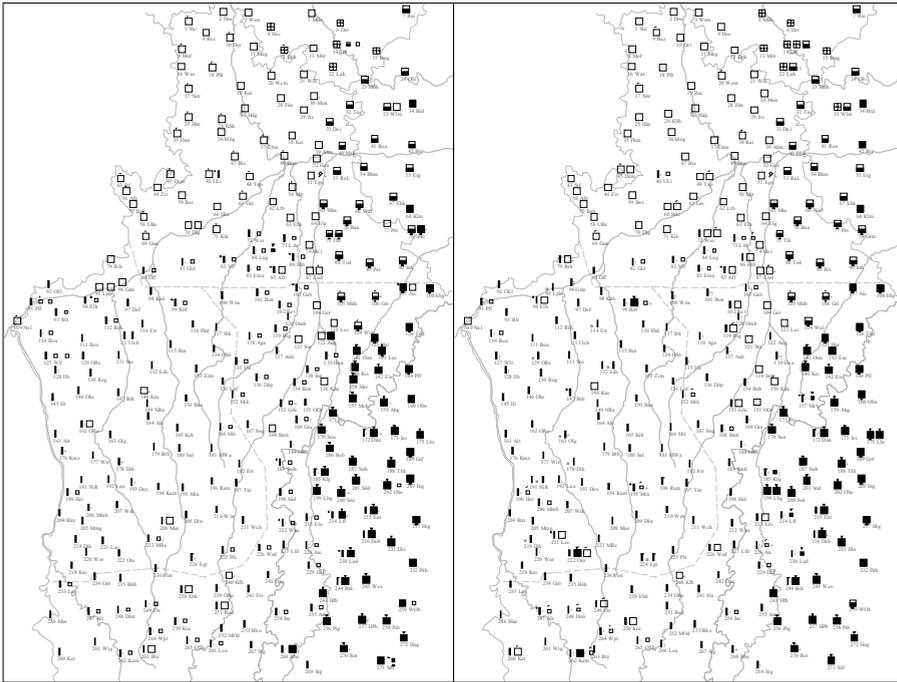


Abb. 2a: 'Kamm' (K. 2.6)

Abb. 2b: 'kämmen' (K. 2.7)

Abb. 2: Die Karten für 'Kamm' und für 'kämmen' aus dem SBS, entnommen aus dem „Digitalen Sprachatlas von Bayerisch-Schwaben“ (DSBS).

derselben Verteilung, nämlich der des Morphems {kamm} und seiner Allomorphe, anzusehen sind. Dies ist anscheinend so banal, dass andere Zusammenhänge als Wortfamilienzugehörigkeit bisher nur selten als Erklärungsfaktoren für ähnliche Verteilungen in Betracht gezogen worden sind. Dabei gibt es zahlreiche Karten, die auch ohne offensichtliche sprachliche Beziehung frappierende Ähnlichkeit aufweisen, wie etwa die Karten für 'Wiesenland umbrechen' und 'Erde (vornehmlich im Acker)' (vgl. Abb. 3, nächste Seite), die zwar keinen so deutlichen Bezug zueinander haben wie 'Kamm' und 'kämmen', aber dennoch eine in gewisser Hinsicht ähnliche Verteilung aufweisen. Ein thematischer Zusammenhang besteht jedoch auch hier (beide Karten gehören zum Themengebiet *Boden und Ackerbau* des SBS), doch ob dieser Zusammenhang tatsächlich mit der Ähnlichkeit der Karten in Verbindung steht oder ob die Ähnlichkeit dem Zufall zuzuschreiben ist, ist ohne eine breiter angelegte Untersuchung, die weitere Variablen einbezieht, nicht zu klären.

Auf eine weitere mögliche Art des Zusammenhangs, die sich auf eine etwas allgemeiner gefasste Ähnlichkeit zwischen einzelnen Sprachkarten bezieht, weist ANDREAS LÖTSCHER hin:

Innerhalb ein und desselben Dialektgebiets können Erscheinungen vergleichbaren Typs sprachgeografisch sehr verschieden stark aufgesplittet sein, ein unterschiedliches Maß an

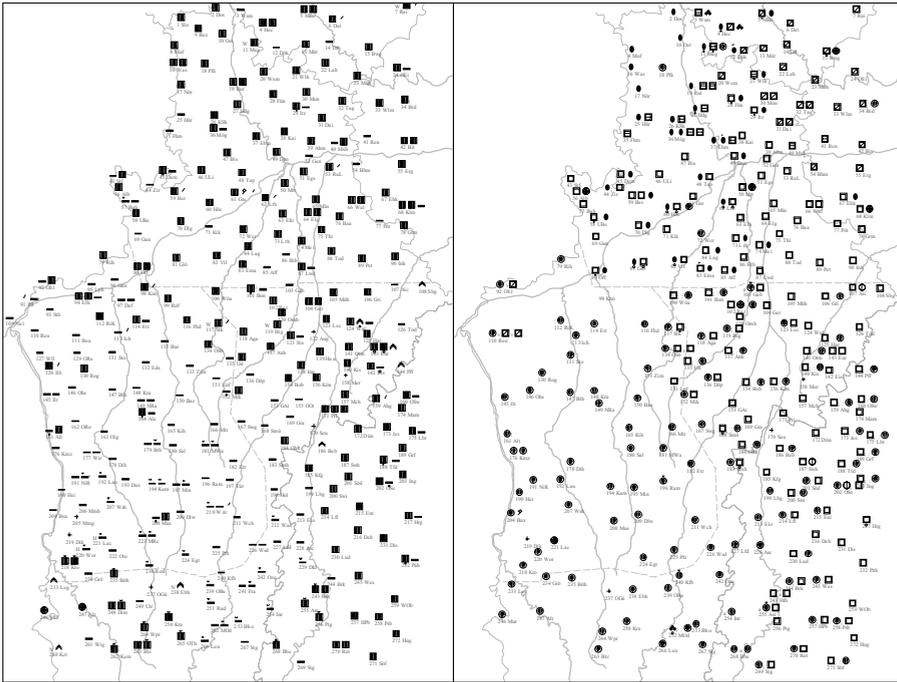


Abb. 3a: 'Wiesenland umbrechen' (K. 12.17)

Abb. 3b: 'Erde (vornehmlich im Acker)'
(K. 12.29)

Abb. 3: Die Karten für 'Wiesenland umbrechen' und für 'Erde (vornehmlich im Acker)' aus dem SBS, entnommen aus dem „Digitalen Sprachatlas von Bayerisch-Schwaben“ (DSBS).

„arealer Diversität“ aufweisen. [...] In jedem Sprachatlas ist diese Tatsache unmittelbar erkennbar. Da sie so evident ist, erscheint sie möglicherweise auch banal. Wenn man nach den Gründen solcher Divergenzen fragt, wird der Sachverhalt etwas problematischer. (LÖTSCHER 2006, 141)

LÖTSCHER spricht hier von Räummustern, die nicht in der Übereinstimmung der Verbreitungsareale von Varianten bestehen. Es geht vielmehr um die Art der Strukturiertheit ganzer Karten. So können manche etwa eine klare Gliederung in nur zwei, einander räumlich kaum durchdringende Variantengebiete aufweisen, andere dagegen viele kleine Gebiete, die immer noch klar voneinander abgegrenzt sein können; bei wieder anderen sind kaum Gebiete zu erkennen, da sich die Varianten räumlich sehr stark vermischen, usw. Erklärungsversuche für solche Unterschiede gibt es viele (vgl. 1.2.1.1), doch steht eine Überprüfung dieser Annahmen bis heute aus: Die Dialektometrie konnte bislang keine quantitativ begründeten Aussagen zu solchen Phänomenen machen, da sie durch die Aggregation die Variation zwischen den Verteilungen der Variablen unberücksichtigt lässt.

1.1.2 Ziele

Die Zielsetzung der vorliegenden Arbeit ist es, auf dem Wege der Einführung (geo-)statistischer, nicht-aggregativer Methoden und Analyseverfahren zu neuen Einsichten in die Struktur und die Gesetzmäßigkeiten diatopischer lexikalischer Variation zu gelangen. Konkret geht es um die statistisch gestützte Analyse der lexikalischen Variation, die in den wortgeographischen Bänden des SBS (Bd. 2; Bd. 8; Bd. 10–13) dokumentiert ist.

Die Lexik wurde dabei aus zweierlei Gründen ausgewählt: Erstens versprach sie, leichter handhabbar zu sein als etwa die Phonetik, da man es hier (fast) ausschließlich mit nominalskalierten Daten zu tun hat (vgl. 3.2), während dort mit einer Mischung aus nominal- und intervallskalierten Daten zu rechnen ist. Zweitens wird dem Lexikon meist geringerer Systemcharakter zugesprochen als etwa dem Lautinventar (vgl. 1.1.1), so dass bei Ersterem mehr Variation (d. h. mehr Unabhängigkeit zwischen den Variablen) zu erwarten ist (vgl. auch WIESINGER 2005, 1108). Dies erschwert zwar die Identifikation von Gesetzmäßigkeiten, doch die Suche nach solchen Regularitäten gerät wesentlich interessanter, da sie weniger offensichtlich sind als systemische Zusammenhänge zwischen lautlichen (oder auch morphologischen) Varianten. Da durch den Fokus auf die Lexik die eingeführten Analysemethoden zur Struktur geolinguistischer Variation an dem Bereich des sprachlichen Systems erprobt werden, der als am wenigsten strukturiert gilt, ist bei erfolgreicher Anwendung ihre Anwendbarkeit auf andere sprachliche Teilbereiche automatisch gewährleistet.

Es wurde bereits erwähnt, dass es schon einige Ansätze zur Erklärung von Verbreitungsmustern von Varianten in der Sprachgeographie gibt (für Details s. 1.2.1). Die dabei herangezogenen Argumentationen gehen jedoch in den allermeisten Fällen nicht über – wenngleich plausible – Erklärungsversuche hinaus; eine tatsächliche Bestätigung ihrer Richtigkeit steht aber weiterhin aus. Desiderate bestehen also unter anderem in der statistischen Überprüfung von seit Längerem in der Literatur zu findenden Behauptungen, die bislang nicht durch quantitative Studien bestätigt werden konnten. Hierfür werden statistische Hypothesen formuliert, um sie dann mit adäquaten Mitteln zu testen (s. 4.6). Über die Überprüfung dieser Annahmen hinaus sollen auch Strukturen in der lexikalisch-diatopischen Variation aufgedeckt werden, die bisher nicht postuliert worden sind. Zu diesem Zweck werden explorative Methoden eingesetzt, die keine vorformulierten Hypothesen testen, sondern aus den Daten bislang unbekannte Strukturen extrahieren, die dann plausibel zu interpretieren sind.

Das Vorhaben, die wortgeographische Variation in Bayerisch-Schwaben zu untersuchen, ist in drei aufeinander aufbauende Teilziele gegliedert:

1. Zunächst soll der Versuch unternommen werden, statistische, computergestützte Verfahren zu entwickeln, die in der Lage sind,
 - a) Hypothesen über die Zusammenhänge zwischen sprachgeographischen Verteilungen und möglichen inner- oder außersprachlichen Faktoren zu überprüfen und