

# LATENTE STRUKTUREN IN GEOLINGUISTISCHEN KORPORA<sup>1</sup>

*Simon Pröll / Simon Pickl / Aaron Spettl*

## 1 HINFÜHRUNG: ZWEI DIALEKTOMETRISCHE DESIDERATE

Den Ausgangspunkt dieses Texts bildet das Desiderat, große geolinguistische Korpora effizient auszuwerten, ohne dabei bedeutende Teile der Variation unberücksichtigt zu lassen. Ziel ist es, umfassende Datenmengen zur diatopischen Variation (d. h. komplette Atlasprojekte oder ihre Teilkorpora) aus einem Blickwinkel zu betrachten, der sowohl die Gesamtvariation als auch die Besonderheiten einzelner Varianten angemessen berücksichtigt, und dabei Strukturen aufzudecken, die dem bloßen Auge der Dialektologin/des Dialektologen aufgrund der Unüberschaubarkeit der Daten verschlossen bleiben.

Die Suche nach diesen latenten Strukturen ist im Zuge der quantitativen Dialektologie bzw. Dialektometrie in eine paradoxe Situation geraten. Zwar entstehen durch die Aggregation großer Datenmengen hierarchisch hochrangige Darstellungen der Variation; welche Variablen dabei für welche Strukturen verantwortlich gemacht werden können, bleibt aber unklar. Ebenso besteht das Problem, dass die Wahl der zugrundegelegten Datenbasis das Resultat prädeterminiert. Bezüglich des ersten Punkts findet sich etwa bei GOEBL (2005: 500) folgende Passage:

Es ist allerdings sehr schwer, für die auf einer Sprachatlaskarte aufscheinende Typen- bzw. Taxat-Vielfalt die verursachenden *Wirkfaktoren* (Kräfte, Synergien etc.) in eindeutig quantifizierbarer Form zu erfassen. Zwar haben viele Einzelanalysen von Sprachatlasdaten (vor allem im Rahmen der Onomasiologie) diesbezüglich eine Vielzahl anekdotischer Information beigebracht. Doch beschränkt sich deren Erklärungskraft jeweils nur auf eine einzige Sprachatlaskarte. Die Einbindung dieser disparaten Einzelphänomene in eine mathematisch faßbare Gesamtsynergetik ist bis jetzt noch nicht geglückt bzw. scheint zur Zeit auch außer Griffweite zu sein.

Der andere Punkt ist weniger explizit adressiert worden, ein Blick auf die Datenbasis quantitativer Studien lässt ihn aber klar hervortreten: Die Bedeutung der Datenselektion. NERBONNE (2006: 464) fasst die Problematik auf konzise Art so

1 Die Methoden und Resultate, die dieser Text präsentiert, sind im Rahmen des DFG-geförderten Projekts „Neue Dialektometrie mit Methoden der stochastischen Bildanalyse“, durchgeführt vom Institut für Stochastik der Universität Ulm und dem Lehrstuhl für Deutsche Sprachwissenschaft der Universität Augsburg, entstanden. Andere zentrale Ergebnisse des Projekts finden sich u. a. in RUMPF et al. (2009), PICKL (i.E.) und MESCHENMOSER / PRÖLL (2012).

zusammen: „Normally, no criteria are identified as to which features are to be studied for their geographic (or social) distribution, and yet the conclusions depend greatly on their choice.“ Während GOEBL (1984) etwa gezielt Lexik und Morphologie für die Analysen zum AIS wählt, nutzt HEERINGA (2004) zur Darstellung der dialektalen Variation der Niederlande vokalische und konsonantische Ausspracheunterschiede in 125 Wörtern. WIESINGERS (1983) vielrezipierte Einteilungskarte zu den deutschen Dialekten ist vornehmlich auf Daten zum Vokalismus gestützt (vgl. WIESINGER 1983: 810–814). Inwiefern die Wahl der zugrundeliegenden Daten die Gestalt der Analyseresultate beeinflusst, bleibt jedoch unklar.

Wir argumentieren dafür, diese Lücken mittels der sogenannten *Faktorenanalyse* zu schließen: Es handelt sich dabei um ein statistisches Data-Mining-Verfahren, das als ein Mittel zum Aufspüren interpretierbarer geografischer Faktoren dienen kann, räumlichen Grundmustern, die sich durch große Teile der analysierten Daten ziehen. Gleichzeitig bleibt aber auch transparent, welche Variablen im Aggregat wie stark zu welchem Grundmuster beitragen. Die Einblicke, die somit in die Struktur von Dialektgebieten und Dialektkontinua möglich werden, sind durch bisherige, rein qualitative Zugänge ebenso verdeckt geblieben wie durch rein quantitative.

Quasi nebenbei wenden wir uns damit auch von einem Varietätenbegriff ab, der seinem Gegenstand homogene Gebiete und scharfe Grenzen ‚oktroziert‘ (zur Kritik siehe AUER 2004; PICKL / RUMPF 2012) – stattdessen geben die Ergebnisse die grundsätzlich kontinuierliche Natur räumlicher Sprachvariation in angemessener Form wieder.

Die Anwendungsbereiche dieser Methode zeigen sich dabei vielgestaltig: Zum einen können auf methodisch sicherer Basis Aussagen zur diatopischen Konfiguration verschiedener linguistischer Ebenen (Phonologie, Morphologie und Lexik) sowie zu ihren Relationen untereinander getroffen werden. Zum anderen kann gezeigt werden, wie stark die Konstruktion von regionalen Varietäten von der Auswahl des zugrundeliegenden Materials abhängt.

Im Folgenden werden kurz die technisch-mathematischen Voraussetzungen erläutert; im Anschluss zeigen wir einige kursorische Ergebnisse zu Daten des *Sprachatlas von Bayerisch-Schwaben*.

## 2 TECHNIK: FAKTORENANALYSE

Im Rahmen einer Faktorenanalyse wird der Zusammenhang einzelner sprachlicher Varianten ermittelt: Treten Varianten häufig miteinander auf, heißt das, dass ihre Distributionen miteinander korrelieren. Den korrelierenden Anteil dieser Variantendistributionen kann man durch einen sogenannten *Faktor* ausdrücken. Statt nun eine große Anzahl Variablen einzeln analysieren zu müssen, kann man nach diesem Muster ihre jeweilige Gemeinsamkeit in nur einer Grundtendenz erfassen.

In der Psychologie, für die die Faktorenanalyse ursprünglich entwickelt wurde (vgl. WOTTAWA 1996: 813), dient sie der Ermittlung von latenten (d. h. nicht direkt beobachtbaren) Variablen, also Persönlichkeitsstrukturen wie Intelligenz oder

Aggression, aus manifesten (d. h. direkt messbaren) Variablen. Dieses Ziel ist praktisch unverändert auf die Sprachwissenschaft übertragbar. Einem größeren Kreis an Linguisten ist die Faktorenanalyse seit den korpuslinguistischen Arbeiten von BIBER (1985; 1991) ein Begriff, aber erst NERBONNE (2006) wendet sie gezielt auf räumlich aufgelöste Sprachdaten, den Vokalismus im *Linguistic Atlas of the Middle and South Atlantic States* (LAMSAS), an. Dort werden artikulatorische Unterschiede zwischen Realisierungen eines Vokals nach einem festgelegten Schema bepunktet, dann wird jeweils die Verschiedenheit der Realisierungen eines Vokals von Ort zu Ort als ‚Abstand‘ ermittelt. Damit ist das Verfahren in dieser Form nur auf phonetische Daten anwendbar; wir nutzen dagegen keine phonetische Abstandsmessung, sondern greifen direkt auf die Varianten zu 1673 Karten<sup>2</sup> des *Sprachatlas von Bayerisch-Schwaben* (SBS)<sup>3</sup> zurück.

Von technischer Seite her werden zunächst die Korrelationen der Variantenverteilungen in den gesamten zugrundeliegenden Daten untereinander ermittelt. Aus der dabei entstandenen Korrelationsmatrix werden im Anschluss die Faktoren extrahiert, die größere „Bündel“ korrelierender Variantendistributionen darstellen – sehr schwache Korrelationen, die wahrscheinlich zufälliger Natur sind, werden ignoriert; somit deckt sich die durch die Faktoren beschriebene Varianz nicht völlig mit der ursprünglich beobachteten (vgl. BACKHAUS et al. 2011: 356). Da durch die jeweiligen Faktorwerte der Varianten quantifiziert ist, welchen Anteil welche Variante an welcher Struktur hat, ist das Verfahren (trotz seiner verdichtenden Wirkungsweise) bis auf die Datenbasis zurück transparent.

Welche Anzahl an Faktoren extrahiert wird, ergibt sich nicht automatisch, sondern erfordert eine Entscheidung von außen. Wir nutzen hier zur Bestimmung das *Kaiser-Guttman-Kriterium*, das verlangt, dass nur Faktoren mit einem Eigenwert  $> 1$  berücksichtigt werden (vgl. BACKHAUS et al. 2011: 359).<sup>4</sup> Die Extraktion und damit Datenreduktion erfolgt anschließend „durch die Auswahl ‚wichtiger‘ Achsen“ (WOTTAWA 1996: 815) im mehrdimensionalen Achsensystem. BACKHAUS et al. (2011: 329–369) sowie TABACHNICK / FIDELL (2007: 607–651) schildern das genaue Vorgehen. Die ermittelten Faktoren stellen „hypothetische Größen, die das Zustandekommen von Korrelationen erklären sollen“ (BORTZ 1993: 473), dar: Jeder der Faktoren kann als Ausdruck eines Prozesses gesehen werden, der für den Zusammenhang zwischen den Variablen verantwortlich ist (vgl. auch TABACHNICK / FIDELL 2007: 582). Abschließend lassen sich die Ergebnisse als Karten visualisieren, indem man entweder

- a) die jeweilige Stärke eines Faktors pro Ort oder
- b) pro Ort den stärksten (oder „dominanten“) Faktor kartiert.

2 Insgesamt 736 Karten zur Lexik, 601 zur Phonologie sowie 336 zur Morphologie (inklusive einiger weniger Syntaxkarten).

3 Der SBS wurde zwischen Herbst 1984 bis Frühjahr 1989 an 272 Orten der Regierungsbezirke Schwaben, Oberbayern und Mittelfranken erhoben. Pro Ort und Variable fallen zwischen 0 und 3 Belege an.

4 Ein Eigenwert  $> 1$  bedeutet hier, dass der Faktor insgesamt mehr Varianz erklärt als ein Ortspunkt.

Ersteres ermöglicht für jede rekonstruierte latente Struktur eine Karte – alternativ erhält man eine Art „Dialekteinteilung“ durch den dominanten Einfluss einzelner Faktoren auf ihre entsprechenden Regionen.

### 3 PRAKTISCHE ANWENDUNG

Wir illustrieren die Möglichkeiten in dieser Fallstudie zunächst detaillierter anhand der Lautung und wählen dabei als Analysetiefe aus Platzgründen das gesamte System – feinere Untergliederungen des Materials sind natürlich problemlos möglich.<sup>5</sup> Im Anschluss zeigen wir zunächst, inwiefern die dabei aufgedeckten Strukturen von denen der Lexik und der Morphologie abweichen, und konstruieren schlussendlich ein Gesamtsystem auf der Basis aller Systemebenen.

#### 3.1 Einzelfaktoren der Lautung

In Abbildung 1 ist pro Karte ein Faktor kartiert worden, der sich durch die Daten zum SBS-Lautsystem zieht; je dunkler eine Region ist, desto stärker ist der Faktor dort. Wenn ein Gebiet von einem schwarzen Rahmen eingegrenzt ist, so ist der betreffende Faktor dort stärker als alle anderen. Wir zeigen hier die 12 stärksten von insgesamt 17 Faktoren, die das Kaiser-Guttman-Kriterium zulässt; sie erklären zusammen ca. zwei Drittel der Variation in den Ausgangsdaten.

Für jeden dieser Faktoren wird über die individuellen Faktorwerte der Varianten offenbar, wie stark jede Variante zum jeweiligen Faktor beiträgt. Für Faktor 1, der am besten als ‚Mittelostschwäbisch‘ zu apostrophieren ist, führt Tabelle 1 exemplarisch die 15 Varianten mit den höchsten Faktorwerten an.

5 Detailliertere Analysen zu Subsystemen der Lexik finden sich bereits bei PICKL (i.E.), tiefere Systemanalysen zu Phonologie und Morphologie sind derzeit in Arbeit.

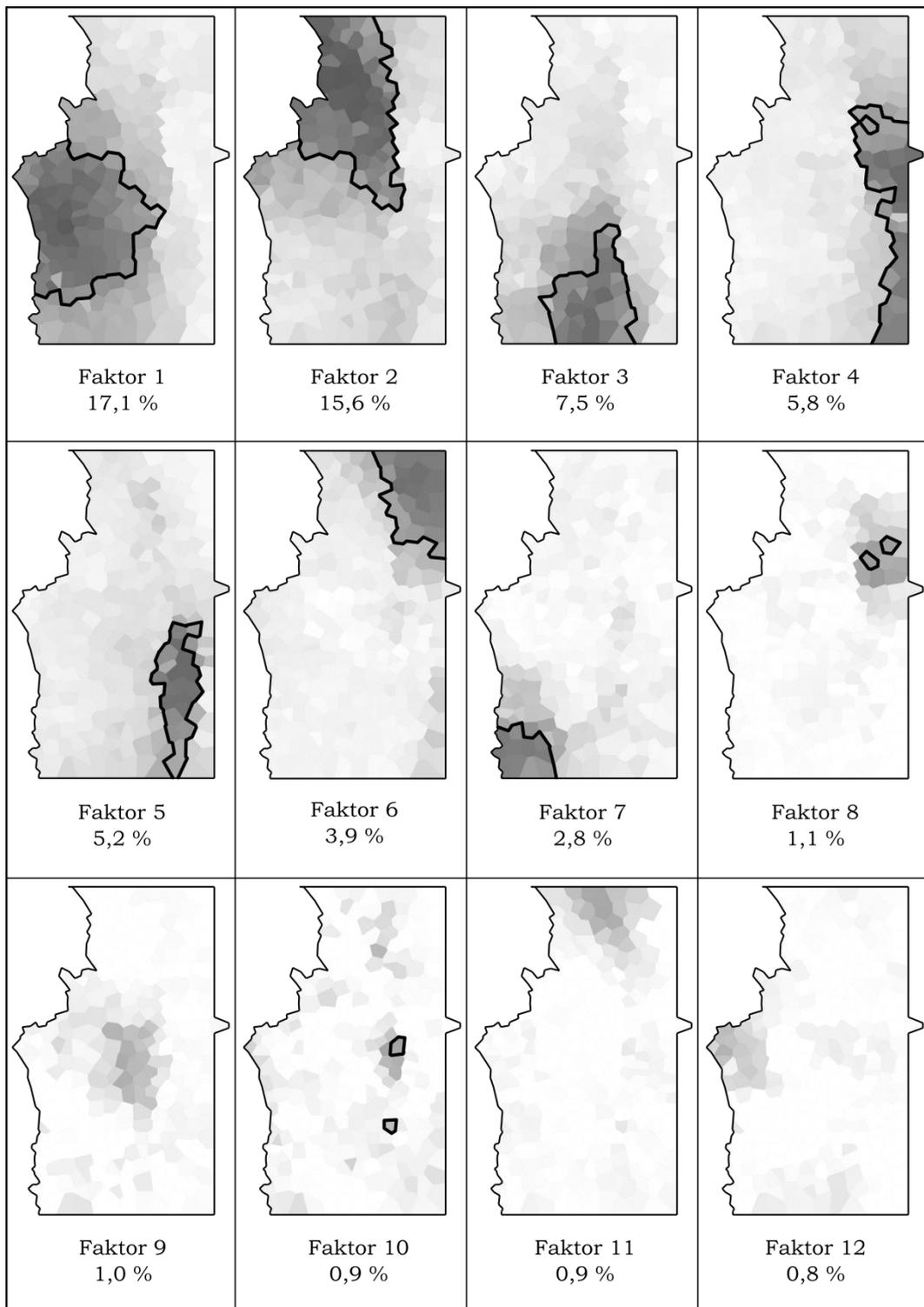


Abbildung 1: Die stärksten 12 Faktoren der SBS-Lautung.

Karte	Variante	Faktorwert
mhd. <i>z</i> (< germ. <i>t</i> ) in »heraußen«	mit Lenis-Frikativ, Typ <i>hyʒə</i>	5,64
mhd. <i>b</i> im In- und Auslaut	realisiert als Lenis-Plosiv	5,24
mhd. <i>ou</i> vor (ehemaligem) Fortisfrikativ in »(ein)kaufen«	Diphthonge vom Typ <i>âo</i>	5,21
mhd. <i>h</i> (< germ. <i>h</i> ) in »siehst du«	Form ohne Reflex von mhd. <i>h</i>	5,19
mhd. <i>ou</i> in »glauben/Glaube(n)«	Diphthonge vom Typ <i>âo</i>	5,06
mhd. <i>ch</i> im In- und Auslaut bei »Furche«	keine Realisierung von mhd. <i>ch</i> bei langem Stammvokal	5,05
mhd. <i>ë</i> in »Besen«	fallender Diphthong, Typ <i>bĕəsəm</i>	5,01
mhd. <i>k</i> in »Onkel«	Realisierung als unbehauchte Lenis	4,94
mhd. <i>ou</i> in »geglaubt«	Diphthonge vom Typ <i>âo</i>	4,81
mhd. <i>ou/(öu)</i> vor (ehemaligem) Fortisfrikativ in »Raufe(l)«	Diphthonge vom Typ <i>âo</i>	4,80
mhd. <i>g</i> im Inlaut bei »morgen«	keine Realisierung von mhd. <i>g</i>	4,80
mhd. <i>â</i> in »Mase«	Diphthonge vom Typ <i>ōũ</i>	4,74
mhd. <i>ou/öu</i> vor (ehemaligem) Fortisfrikativ in »taufen«	Diphthonge vom Typ <i>ei</i>	4,71
mhd. <i>öu</i> in »Äuglein«	Diphthonge vom Typ <i>âe</i>	4,70
mhd. <i>â</i> vor <i>l</i> in »malen«	Diphthonge vom Typ <i>âo</i>	4,68

Tabelle 1: Faktorwerte der 15 stärksten Varianten in Faktor 1.

Die stärksten Varianten innerhalb dieses Faktors stellen demnach die charakteristische mittelostschwäbische Diphthongrealisierung (vgl. KÖNIG 2010) sowie lenisierte / nullrealisierte Plosive und Frikative dar. Der benachbarte Faktor 2 (Nordostschwäbisch) ist dagegen maßgeblich zurückzuführen auf Lenis-Realisierungen von Dentalplosiven. Die Faktoren 3 und 7 entsprechen den Einzugsgebieten von Lech und Iller und decken zusammen die Region des Allgäus ab. Die Faktoren 4 und 8 belegen den mittelbairischen Raum des Untersuchungsgebiets (im Folgenden USG), scharf östlich des Lechs und südlich der Donau: Faktor 4 ist durch Formen gekennzeichnet, die im Gebiet des klassischerweise als „Mittelbairisch“ bezeichneten Dialekts vorkommen. Er wird durch den schwächeren Faktor 8 ergänzt (der fast nirgendwo dominant wird). Dieser zeigt den besonders starken Gegensatz zwischen bairischem und alemannischem Raum (die bis heute durch eine starke „mentale Grenze“ im Sinne AUERS 2004 getrennt sind) entlang des nördlichen Lechs zusätzlich an. Faktor 5 gliedert klar den sogenannten Lechrain (vgl. WÖLZMÜLLER 1987) aus seinem Umland aus, während Faktor 6 den nordbairischen Raum zusammenfasst. Besonderes Augenmerk verdient die Tatsache, dass praktisch alle Faktoren räumlich kohärent sind. Eine interessante Ausnahme bildet Faktor 10; er entspricht den bevölkerungsreichen Gebieten der „Städtelandschaft“ Ostschwabens, die in den Worten KIEBLINGS (2009: 47) „in Deutschland nur wenige Parallelen“ findet. Die dominanten Kernbereiche dieses Faktors sind Augsburg (und das 1913 eingemeindete Lechhausen) sowie Landsberg am Lech. Schwächer, aber dennoch deutlich, treten – neben den Stadtteilen und Vororten Augsburgs – Kaufbeuren, Kempten, Memmingen, Neu-Ulm, Dillingen, Nördlin-

gen und das Gebiet um Donauwörth hervor.<sup>6</sup> Faktor 11 wiederum verweist auf den Übergang zum fränkischen Dialektraum im Norden des USG.

### 3.2 Synopse der Lautung

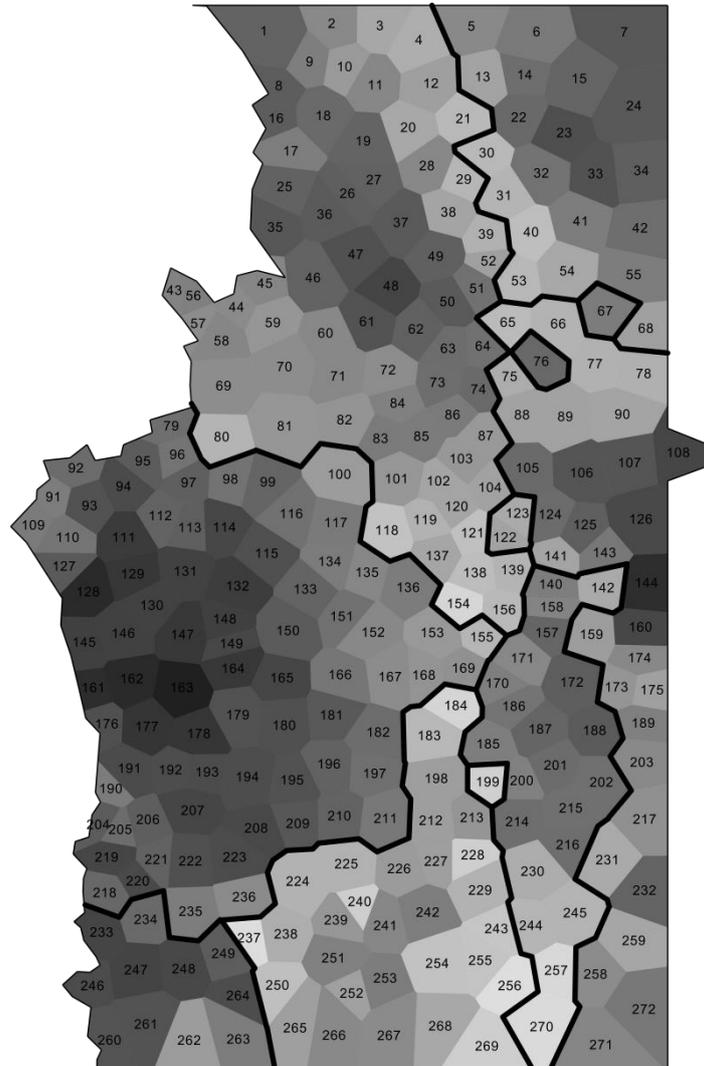


Abbildung 2: Dominante phonologische Faktoren im SBS.

Die kombinierte Darstellung aller dieser Faktoren in einer Karte erzeugt eine Art „Dialekteinteilung“ auf Basis der phonetisch-phonologischen Daten des SBS (siehe Abbildung 2; im Original farbig). Visualisiert wird pro Ort der stärkste Faktor.

6 Wie zu erwarten war, tragen zu diesem Faktor vor allem standardnähere Varianten im Vokalismus wie z. B. monophthongische *o*-Realisierungen (die im Gegensatz zu den ländlicheren Formen stehen) oder Kürze in Einsilblern wie *Frosch*, *Stall* oder *Darm* besonders stark bei.

Betont sei aber, dass für uns nicht die Grenze zwischen zwei Gebieten zentral ist (vgl. ausführlicher PICKL / RUMPF 2012), sondern die (dunkleren) Zentren der Faktoren, die deren Kerngebiete darstellen; Grenzen entstehen in der Darstellung als ‚Nebenprodukt‘, wenn ein Faktor stärker wird als die anderen.

### 3.3 Vergleich der lexikalischen, phonologischen und morphematischen Ebene

In Abbildung 3 sind die kombinierten Faktorenkarten für Lexik, Lautung und Formen (basierend auf 736, 601 bzw. 336 ursprünglichen SBS-Karten) nebeneinandergestellt.

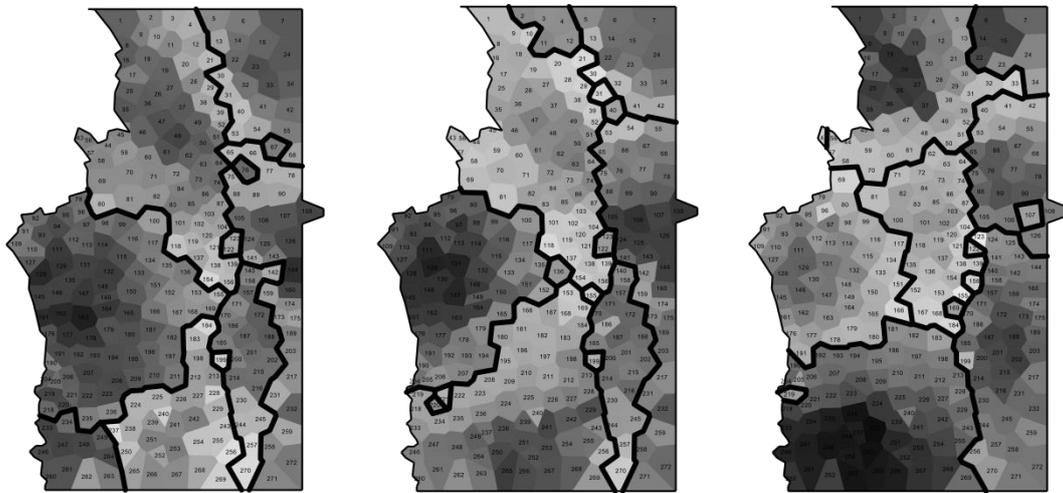


Abbildung 3: Dominante Faktoren, Teilsysteme (von links nach rechts) Phonologie (vgl. Abb. 2), Morphologie und Lexik.

Auf den ersten Blick zeigt sich, dass das Gebiet des Lechrains bei Phonologie und Morphologie fast völlig deckungsgleich ist, in der Lexik bildet er jedoch einen Teil des Mittelbairischen. Die Grenze zum nordbairischen Raum liegt im Bereich der Lexik und Morphologie weiter nördlich als in der Phonologie und ist insgesamt ein Gebiet mit transitorischem Charakter. In der Morphologie stellt sich im Norden auch ein wohl als „Fränkisch“ zu sehender Faktor ein (besonders deutlich tragen hierzu unter anderem die kurzen Wurzelvokale in schwachen Verbformen bei). Das Zentrum des USG (westlich von Augsburg) ist tendenziell eine Region, in der die verschiedenen Faktoren ineinander übergehen, indem sie sich gegenseitig überlagern; in der Lexik führt das zur Dominanz eines ansonsten eher schwachen eigenständigen Faktors.

In jedem Fall verdeutlicht die Gegenüberstellung, dass die den Einzelvariablen zugrundeliegenden Tendenzen sich nicht über die Grenzen des Teilsystems (Wortschatz, Lautung, Formen) hinaus decken. Abhängig von der Wahl der Datenbasis ergeben sich zum Teil deutlich abweichende latente Grundstrukturen der Variation.

### 3.4 Synopse aller Ebenen

Nimmt man nun alle 1673 Varianten zusammen, kann man auf Basis der Lexik, Phonologie und Morphologie ein sprachgeographisches Gesamtsystem innerhalb des bayerisch-schwäbischen Raums konstruieren. In Abbildung 4 sind wiederum die jeweils an einem Ort stärksten Faktoren kartiert. Das dabei entstehende Raumbild ist mit keiner der Einzelanalysen exakt deckungsgleich, keine sprachliche Ebene kann demnach als repräsentativ für die bayerisch-schwäbische Gesamtvariation (zumindest in ihrer durch den SBS abgedeckten Form) angesehen werden. Das Gesamtsystem ist aber auch nicht einfach eine Mischung der Teilsysteme; die Faktorenanalyse kann auf Grundlage aller Variablen Korrelationen globaler Natur finden, die über das einzelne Teilsystem hinausgehen und daher dort nicht offenbar werden.

Dabei bleiben die groben Zentren der Regionen zumindest ähnlich: Stärkster Faktor (13,4 % erklärte Varianz) ist in der Gesamtschau der Allgäuer Raum im Südwesten. „Mittelostschwaben“ scheint – wie schon in den Analysen der Teilebenen augenscheinlich war – nicht vom angrenzenden Großraum Ulm auszustrahlen, sondern hat sein Zentrum weiter östlich. Augsburg (mit Lechhausen) sticht auch in der Gesamtschau als städtisches Zentrum klar heraus. (Der entsprechende Faktor deckt sich räumlich im Wesentlichen mit Faktor 10 in Abbildung 2). Deutlich tritt auch der nördliche Verlauf des Lechs als eine der traditionell stärksten Binnengrenzen des deutschen Sprachraums zu Tage: Der Faktor, der den „mittelbairischen“ Raum im Osten abdeckt, ist im Südosten relativ schwach dominant, nordöstlich von Augsburg allerdings schon direkt am Lechufer überaus stark. Die gegenüberliegende alemannische Seite zeigt sich dagegen eher als Mischgebiet der Ausläufer des Mittel- und Nordostschwäbischen. Das zentrale Gebiet südlich davon kann überspitzt als ‚Niemandland‘ bezeichnet werden: Die stärkeren miteinander konkurrierenden Faktoren sind hier jeweils so schwach, dass ein anderer, ansonsten relativ undifferenzierter Faktor dominant werden kann. Die Region ist klar als Übergangsgebiet der umgebenden Zentren zu werten.

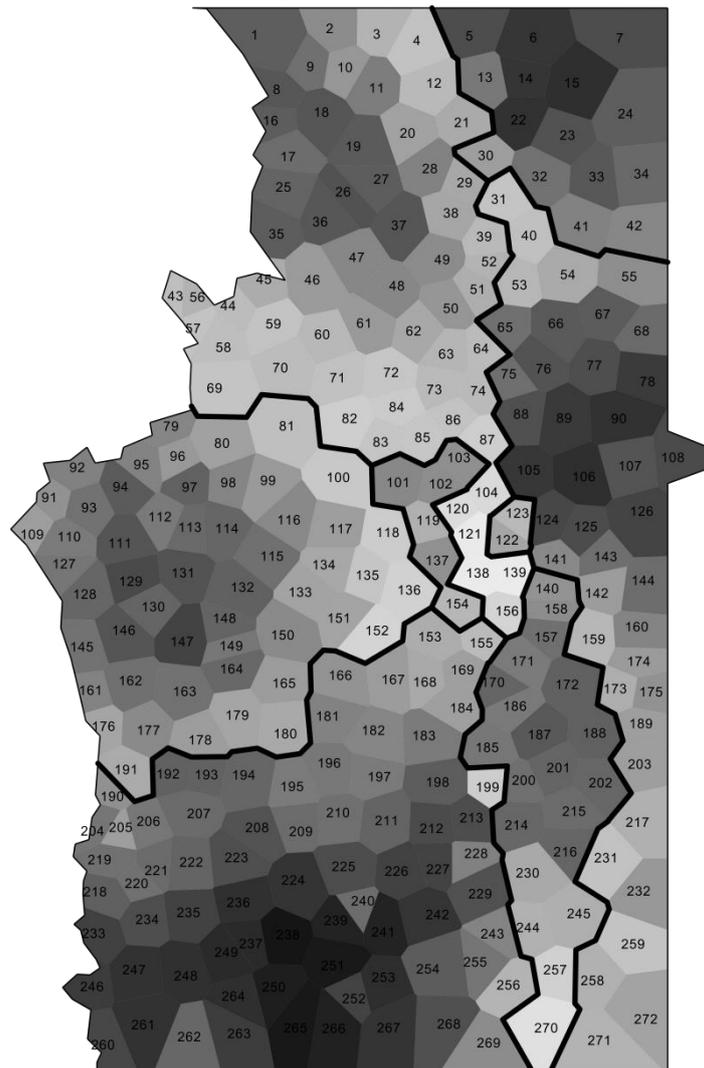


Abbildung 4: Dominante Faktoren des Gesamtsystems.

## FAZIT

Ziel dieser Fallstudie war es zum einen, die Faktorenanalyse als Instrument zur effektiven Auswertung und Interpretation geolinguistischer Daten zu propagieren. Es zeigt sich, dass die Faktorenanalyse tiefgreifende, latente Strömungen in der Gesamtvariation aufdecken kann, die sowohl beim Blick auf Einzelvarianten als auch in der bislang üblichen quantitativen Dialektologie verborgen bleiben. Dabei bleibt auch in großen Kartenkorpora der Zugriff auf die Rolle der Einzelvariante unverstellt. Zum anderen konnte dargestellt werden, wie die unterschiedlichen Ebenen des Sprachsystems auch unterschiedliche geografische Konfigurationen zeigen. Das führt zur Einsicht, dass a) auf die einzelnen sprachlichen Systemebenen jeweils individuelle Faktoren unterschiedlich stark einwirken sowie b) das Ergebnis von Dialekteinteilungen durch die Wahl des zugrundegelegten Materials

klar vordeterminiert wird – und im Umkehrschluss, dass Einteilungen, die nur auf Ausschnitten der Daten oder einzelnen Systemebenen beruhen, nicht oder nur sehr eingeschränkt für andere Systemebenen sprechen können.

## LITERATUR

- AUER, PETER (2004): Sprache, Grenze, Raum. In: Zeitschrift für Sprachwissenschaft 23/2, 149–180.
- BACKHAUS, KLAUS / BERND ERICHSON / WULFF PLINKE / ROLF WEIBER (2011): Multivariate Analysemethoden. 13., überarbeitete Auflage. Berlin: Springer.
- BIBER, DOUGLAS (1985): Investigating macroscopic textual variation through multifeature/multidimensional analyses. In: Linguistics 23/2, 337–360.
- BIBER, DOUGLAS (1991): Variation across speech and writing. Cambridge, New York [u. a.]: Cambridge University Press.
- BORTZ, JÜRGEN (1993): Statistik für Sozialwissenschaftler. 4. Auflage. Berlin, Heidelberg: Springer.
- GOEBL, HANS (1984): Dialektometrische Studien. Anhand italo-romanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF. 3 Bände. Tübingen: Niemeyer.
- GOEBL, HANS (2005): Dialektometrie. In: KÖHLER, REINHARD / GABRIEL ALTMANN / RAJMUND G. PIOTROWSKI (Hg.): Quantitative Linguistik. Ein internationales Handbuch. Berlin, New York: de Gruyter, 498–531.
- HEERINGA, WILBERT JAN (2004): Measuring dialect pronunciation differences using levenshtein distance. Dissertation, Rijksuniversiteit Groningen.
- KIEBLING, ROLF (2009): Kleine Geschichte Schwabens. Regensburg: Pustet.
- KÖNIG, WERNER (2010): Großmundarten und Dialektgrenzen. In: FREI, HANS / PANKRAZ FRIED / FRANZ SCHAFFER (Hg.): Historischer Atlas von Bayerisch-Schwaben. 2. Auflage. Augsburg: Verlag der schwäbischen Forschungsgemeinschaft. 5. Lieferung, XIII, 1.
- MESCHENMOSE, DANIEL / SIMON PRÖLL (2012): Using fuzzy clustering to reveal recurring spatial patterns in corpora of dialect maps. In: International Journal of Corpus Linguistics 17/2, 176–197.
- NERBONNE, JOHN (2006): Identifying Linguistic Structure in Aggregate Comparison. In: Literary and Linguistic Computing 21/4, 463–476.
- PICKL, SIMON (i.E.): Probabilistische Geolinguistik. Geostatistische Analysen lexikalischer Distribution in Bayerisch-Schwaben. Stuttgart: Steiner.
- PICKL, SIMON / JONAS RUMPF (2012): Dialectometric Concepts of Space: Towards a Variant-Based Dialectometry. In: HANSEN, SANDRA / CHRISTIAN SCHWARZ / PHILIPP STOECKLE / TOBIAS STRECK (Hg.): Dialectological and folk dialectological concepts of space. Berlin, New York: de Gruyter, 199–214.
- RUMPF, JONAS / SIMON PICKL / STEPHAN ELSPAß / WERNER KÖNIG / VOLKER SCHMIDT (2009): Structural analysis of dialect maps using methods from spatial statistics. In: Zeitschrift für Dialektologie und Linguistik 76/3, 280–308.
- SBS = KÖNIG, WERNER (Hg.) (1996–2009): Sprachatlas von Bayerisch-Schwaben. Band 1–14. Heidelberg: Winter.
- TABACHNICK, BARBARA G. / LINDA S. FIDELL (2007): Using Multivariate Statistics. Boston: Allyn & Bacon.
- WIESINGER, PETER (1983): Die Einteilung der deutschen Dialekte. In: BESCH, WERNER / ULRICH KNOOP / WOLFGANG PUTSCHKE / HERBERT ERNST WIEGAND (Hg.): Dialektologie. Ein

Handbuch zur deutschen und allgemeinen Dialektforschung. Band 1.2. Berlin, New York: de Gruyter. 807–900.

WÖLZMÜLLER, MARTIN (1987): Der Lechrainer und seine Sprache. Landschaft, Brauchtum, Mundart. Landsberg am Lech: Landsberger Verlagsanstalt Martin Neumeyer.

WOTTAWA, HEINRICH (1996): Multivariate Verfahren der Eigenschaftsanalyse. In: PAWLIK, Kurt (Hg.): Grundlagen und Methoden der Differentiellen Psychologie. Göttingen, Bern, Toronto, Seattle: Hogrefe. 807–831.