

SIMON PICKL

## VERDICHTUNGEN IM SPRACHGEOGRAFISCHEN KONTINUUM\*

## 1. Konzeptualisierungen des Sprachraums

Wie ist Sprache im Raum verteilt? Für diese allgemein formulierte Frage hält die Dialektologie zwei verschiedene Antworten bereit, die sich auf den ersten Blick gegenseitig zu widersprechen scheinen.

Traditionell wird der Sprachraum als bestehend aus Dialektgebieten konzeptualisiert, innerhalb derer relativ gleich gesprochen wird, und zwischen denen relativ große sprachliche Unterschiede existieren. Die Vorstellung solcher Dialektgebiete bestimmt seit JACOB GRIMM die Sprachgeografieforschung (vgl. CHRISTEN 2010, 269); man hat ihnen methodisch auf vielerlei Weise näherzukommen versucht, zum Beispiel indem man empirisch feststellen wollte, in welchen Regionen man ungefähr gleich spricht, und wo Bereiche unterschiedlichen Sprechens aufeinandertreffen.<sup>1</sup> Die damit postulierte Struktur könnte man als kristallin bezeichnen: Es geht um klare Ordnungsmuster, die aus in sich gleichförmigen Bereichen und harten Kanten zwischen ihnen bestehen. Die Idee von „einheitlichen Mundartgebieten“ und „scharfen Dialektgrenzen“ war jedoch wohl stets eine Idealisierung und wurde früh zurückgewiesen (vgl. unter anderem FISCHER 1895, 80–81; BACH 1950, 58–62; WIESINGER 1983, 809); man kam zu der Ansicht, dass auch sprachlich relativ einheitliche Gebiete interne Variation aufweisen, und dass die Grenzen zwischen ihnen keineswegs scharf, sondern fließend sind. Dennoch ist das Konzept des in Gebiete gegliederten Sprachraums bis heute ein vorherrschendes Paradigma in der Dialektgeografie.

\* Dieser Beitrag stellt eine veränderte und erweiterte Fassung der Teile 2.3 und 5.2 der Dissertation des Autors (PICKL 2013) dar, die sich mit variablenübergreifenden Raumstrukturen beschäftigen. Erweiterungen bestehen im Wesentlichen in der Diskussion der Konzepte des Kontinuums und der Areale und im Vergleich des hier vorgeschlagenen Verfahrens und seiner Ergebnisse mit herkömmlichen Verfahren und traditionellen Einteilungen des Dialektraums Bayerisch-Schwaben.

<sup>1</sup> Angefangen mit der Isoglossenmethode, bei der die Außengrenzen der Verbreitungsgebiete einzelner sprachlicher Erscheinungen übereinandergelegt werden, um von Bündeln solcher Linien auf Dialektgrenzen zu schließen, bis zur modernen Clusteranalyse, die auf der Grundlage umfangreicher Datenmatrizen Ortsdialekte zu immer größeren Gruppen – und damit zu Dialektgebieten – zusammenfasst, haben alle diese Verfahren die Vorstellung des in Dialektgebiete gegliederten Sprachraums gemein.

Dem steht eine Vorstellung des Sprachraums gegenüber, die nicht die areale Einteilung als primär ansieht, sondern den Sprachraum als eher amorph konzeptualisiert: „Die Dialekte bilden ein Kontinuum im Raum“ (WIESINGER 1983, 807). Das Modell des Dialektkontinuums geht davon aus, dass die Verbreitungen einzelner sprachlicher Varianten meist nicht miteinander übereinstimmen und deshalb auch keine Dialektgrenzen bilden, sondern dass sie eher unsystematisch im Raum verteilt sind, weshalb keine Gebiete abgesteckt werden können. Diese Vorstellung wird häufig durch das Bild eines Wanderers illustriert, der die Landschaft entlang einer Linie durchquert und dabei von Ort zu Ort kleine, oft kaum merkbare sprachliche Unterschiede feststellt. Diese Unterschiede akkumulieren sich jedoch mit dem zurückgelegten Weg, bis der Wanderer am Ende seiner Reise eine völlig andere Redeweise vernimmt als zu Beginn (CHAMBERS / TRUDGILL 1998, 5–7; vgl. auch BERRUTO 2010, 235–236; KÖNIG 2010b).

WILBERT HEERINGA und JOHN NERBONNE (2001) setzen sich empirisch mit dem Problem Kontinuum versus Areale auseinander. Für ihre Untersuchung werten sie Aussprachedaten quantitativ aus. Konkret betrachten HEERINGA / NERBONNE 27 Orte, die entlang einer Linie quer durch den niederländischen Sprachraum aufgereiht sind (in Anlehnung an das Bild des Wanderers bei CHAMBERS / TRUDGILL). Sie stellen dabei fest, dass geografische und phonetische Distanzen zwischen je zwei Orten stark korreliert sind, was auf ein Kontinuum hindeutet. Zwischen einzelnen Nachbarortepaaren bestehen jedoch signifikant höhere phonetische Abstände als durchschnittlich, was auf einen Bruch beziehungsweise eine Grenze hindeutet und damit die Vorstellung von Gebieten rechtfertigt.<sup>2</sup> Ausgehend von dieser Beobachtung führen die Autoren eine Clusteranalyse durch, um Gruppen von Orten zu ermitteln, die aufgrund ihrer phonetischen Ähnlichkeit als zu einem Gebiet gehörend interpretiert werden können. Es ergeben sich drei Hauptgruppen, deren Grenzen ungefähr mit den oben erwähnten signifikanten Abständen zwischen Nachbarorten zusammenfallen – interessanterweise aber nur ungefähr, das heißt um ein bis zwei Orte versetzt.<sup>3</sup> In diesem Zusammenhang sprechen HEERINGA und NERBONNE von parallelen, das heißt gestaffelten Isoglossen, die eine Übergangszone konstituieren.

HEERINGA / NERBONNE (2001, 399) gelangen in ihrer Studie zu der Ansicht, „that both the area view and the continuum view are useful for gaining insight in the nature of the dialect landscape, which may be described as a continuum with varying slope or, alternatively, as a continuum with unsharp borders between

<sup>2</sup> WERNER KÖNIG (2010b) spricht von „[k]leine[n] Stufen in einer schiefen Ebene des allmählichen Übergangs [...], wo sich Sprachgrenzen bündeln und die Dialekte nebeneinander liegender Orte sich stärker unterscheiden als im sonstigen Durchschnitt“.

<sup>3</sup> Dies zeigt zum einen, dass mittels der Clusteranalyse ermittelte Grenzen nicht unbedingt auffällige Brüche im Kontinuum darstellen, und zum anderen, dass eine so erzeugte Einteilung in klare Gruppen kein Beleg für eine solche Struktur in den Daten ist (denn die Clusteranalyse liefert immer Gruppen).

dialect areas“ (ähnlich GOEBL 1983). Die beiden Herangehensweisen – die Einteilung in Gebiete und die Darstellung von Kontinua – greifen unterschiedliche Aspekte der zugrundeliegenden sprachlichen Landschaft heraus. Beide Sichtweisen sind je nach Forschungsinteresse berechtigt und nützlich, dennoch handelt es sich zunächst um alternative, getrennte Sichtweisen. Die Betrachtungsweise als Kontinuum trägt der graduell veränderlichen, „amorphen“ Raumanordnung verschiedener Sprechweisen Rechnung, während die Einteilung des Sprachraums in diskrete Dialektgebiete letztlich Ausdruck des menschlichen Bedürfnisses nach Kategorien ist: Indem die sprachliche Vielfalt in geografisch kohärente Abschnitte eingeteilt wird, wird sie überschaubar und dadurch handhabbar (vgl. auch WIESINGER 1983, 807).

Die Grundvoraussetzung für eine Kompatibilität von Dialektarealen mit einem Dialektkontinuum ist die Annahme von Übergangszonen zwischen sprachlichen Raumeinheiten, wenngleich die Vorstellung des Dialektareals dabei teilweise „aufgeweicht“ wird. Während in der Untersuchung von HEERINGA / NERBONNE jede Übergangszone mit einem Bruch im Kontinuum (oder zumindest mit „varying slope“ desselben, siehe oben) einhergeht, so ist durchaus auch vorstellbar, dass ein Übergang zwischen zwei Gebieten so „sanft“ und graduell verläuft, dass keine Auffälligkeiten im Kontinuum bemerkt werden können.

Für den deutschen Sprachraum hat PETER WIESINGER (1983) mit seiner Einteilungskarte der deutschen Dialekte, die „anschaulich macht, dass man neben mehr oder weniger festumgrenzten dialektalen Kerngebieten in beträchtlichem Maße ‚Übergangsgebiete‘ anzunehmen hat“ (NIEBAUM / MACHA 2006, 217), erstmals in nennenswerter Weise den graduellen Charakter von Dialektgrenzen berücksichtigt. Die von WIESINGER angesetzten Dialektregionen und Dialektverbände beruhen auf ausgewählten Variablen, die „systemhafte und damit typologische Struktureigenschaften“ (WIESINGER 1983, 810–811) darstellen sollen. Wo die Verbreitungsgebiete mehrerer solcher struktureller Eigenschaften dergestalt zusammenfallen, dass ein Bereich relativer Homogenität entsteht, veranschlagt WIESINGER ein Dialektareal, wobei er unterschiedliche Hierarchieebenen („Dialektregionen, Dialektverbände, großräumige Dialektgruppen, kleinräumige Dialektgruppen, Ortsdialekte“; WIESINGER 1983, 814) unterscheidet. Dabei ist die Interpretation eines relativen Homogenitätsbereichs nicht immer eindeutig, etwa wenn ein von WIESINGER angesetztes Dialektgebiet bei näherer Betrachtung selbst den Charakter eines Übergangsgebiets zu haben scheint, wie etwa im Fall des Thüringischen, dem „bloß ihm eigene verbindende Merkmale gegenüber den Nachbardialekten“ (WIESINGER 1983, 859) fehlen. Insgesamt identifiziert WIESINGER neben Gebieten mit scharfen Grenzen, die aus geografisch genau koinzidierenden Isoglossen bestehen, auch solche mit unscharfen Übergängen, bei denen mehrere Isoglossen nur mehr oder weniger dasselbe Kerngebiet umfassen.<sup>4</sup>

<sup>4</sup> Siehe auch die Diskussion von WIESINGERS Einteilung in STRECK / AUER (2012, 157–162).

Die moderne Dialektometrie hält Analysemethoden bereit, die hauptsächlich für je einen der beiden disparaten theoretischen Ansätze geeignet sind, jedoch nicht für beide gleichzeitig: Mit der Clusteranalyse wird der Raum auf der Grundlage der Daten in areale Strukturen eingeteilt (auch dann, wenn solche zur Beschreibung der Daten eigentlich nicht geeignet sind), während mit Ähnlichkeitskarten geografische Sprachkontinua dargestellt werden (aber immer nur für einen einzelnen gewählten Bezugspunkt). Mithilfe von multidimensionaler Skalierung können kontinuierliche Karten erstellt werden, die ohne Bezugspunkt auskommen, aber keine kategoriale Einteilung bieten. Erst seit kurzem halten in der quantitativ orientierten Methodik auch Verfahren Einzug, die in der Lage sind, Dialektgebiete mit Übergangsbereichen zu erfassen und damit beide Aspekte zu verbinden. Dabei handelt es sich zum einen um Varianten der Clusteranalyse, die durch statistische Verfahren graduelle Zugehörigkeiten von Orten zu Clustern wiedergeben und dadurch unscharfe Dialektgebiete produzieren können (bei diesem sogenannten *fuzzy clustering* gibt es zwei methodisch verschiedene Verfahren, das *bootstrap clustering* und das *noisy clustering*, die aber beide zu sehr ähnlichen Ergebnissen führen; vgl. NERBONNE et al. 2011, 83). Auch die statistischen, dimensionsreduzierenden Verfahren der Faktorenanalyse und der eng verwandten Hauptkomponentenanalyse sind geeignet, „unscharfe“ sprachräumliche Strukturen zu identifizieren und abzubilden (siehe Abschnitt 3). Sowohl die Faktorenbeziehungsweise Hauptkomponentenanalyse als auch das *fuzzy clustering* (mit den beiden Implementierungen) stellen also auf methodischer Seite Kandidaten für die kontinuierlich-areale Analyse des Sprachraums dar.

Theoretische Grundlagen zu einer entsprechenden Modellierung des Sprachraums finden sich insbesondere bei GAETANO BERRUTO: „The arrangement of varieties in the language space constituting a language takes the form of continua“ (BERRUTO 2010, 235; für eine eingehendere Diskussion seines Ansatzes siehe Abschnitt 2). Auf sein Varietätenmodell aufbauend wird in diesem Beitrag der Versuch vorgestellt,

- 1) eine theoretische Konzeptualisierung der Struktur des Sprachraums vorzunehmen, aus der sowohl die Vorstellung von Dialektkontinua als auch die von unterscheidbaren diatopischen Varietäten hervorgeht (Abschnitt 2), sowie
- 2) eine Methodik einzuführen, die eine Operationalisierung dieser theoretischen Grundlage darstellt und mit ihr kompatibel ist (Abschnitt 3).

Bei der exemplarischen Anwendung der Methode (Abschnitt 4) auf Daten aus dem Sprachatlas von Bayerisch-Schwaben (SBS) wird zudem gezeigt, dass mit dieser Methodik Einsichten in latente, bedeutungsvolle sprachgeografische Strukturen ermöglicht werden, die dem Auge der Dialektologin / des Dialektologen mit den anderen Verfahren verschlossen bleiben.

## 2. Sprachgeografische Verdichtungen als Dialekttypen

Der Sprachraum besteht zunächst aus einer Vielzahl einzelner geografischer Variantenverteilungen, die sich teilweise decken, teilweise überlappen, und teilweise keine unmittelbar erkennbaren Ähnlichkeiten untereinander aufweisen. Diese Vielfalt scheint *prima facie* in der Tat amorph zu sein, was zunächst für eine Konzeptualisierung als Sprachkontinuum zu sprechen scheint; es sind jedoch durchaus immer wieder wiederkehrende Strukturen (d. h. ähnlich im Raum verteilte Varianten) erkennbar, die die amorphe Gesamtmasse durchbrechen. Insofern könnte man von einer teilkristallinen Gesamtstruktur sprechen; es handelt sich um ein „Kontinuum mit Verdichtungen“ (BERRUTO 2010, 236). Wie lassen sich nun in einem solchen prinzipiell ungeordneten, aber doch teilweise strukturhafte Züge aufweisenden Medium Kategorien bilden?

Der Ansatz GAETANO BERRUTOS beruht auf einer basalen Annahme: Wenn mehrere Varianten insgesamt in nennenswerter Weise dazu tendieren, in denselben Kontexten verwendet zu werden, so konstituieren diese Variantenkombinationen Varietäten. Entsprechend definiert er eine Varietät als Ansammlung ähnlich verteilter, das heißt tendenziell kookkurrenzer Varianten:

The tendential co-occurrence of variants gives rise to linguistic varieties. Therefore, a linguistic variety is conceivable as a set of co-occurring variants; it is identified simultaneously by both such a co-occurrence of variants, from the linguistic viewpoint, and the co-occurrence of these variants with extralinguistic, social features, from the external, societal viewpoint. (BERRUTO 2010, 229)

Varietäten sind demnach konstituiert durch das gehäufte gemeinsame Auftreten von mehreren Varianten, das heißt durch signifikante Variantenkombinationen.<sup>5</sup> BERRUTO bezieht sich dabei nicht primär auf Variantenverteilungen im geografischen Raum, sondern allgemeiner auf deren Distribution im „language space as a multidimensional continuum“ (BERRUTO 2010, 237), der durch die drei Dimensionen Diatopie, Diastratie und Diaphasie gebildet wird. Dieser Sprachraumbegriff lässt sich für unser Vorhaben jedoch ohne Probleme auf die diatopische Dimension reduzieren. Für das Feststellen diatopischer Varietäten ist demnach anzunehmen, dass sich Varianten anhand ihrer räumlichen Verbreitungen zusammenfassen lassen. Hierbei steht die Annahme im Vordergrund, eine auffällige „Ballung“ von kookkurrenten Varianten sei nichts Zufälliges, sondern durch sprachliche und nicht-sprachliche Faktoren bedingt, die in ähnlicher Weise auf mehrere Varianten gleichzeitig einwirken und so die genannten Verdichtungsgebiete erzeugen.

Varietäten sind demnach „Verdichtungen im Kontinuum“ (BERRUTO 1987; 2010), die sich durch das gehäufte Auftreten von Kookkurrenzen unter Varianten

<sup>5</sup> Dadurch unterscheidet sich dieser quantitativ-statistische Ansatz von kognitiv-perzeptiven und/oder strukturalistischen Varietätenbegriffen, wie sie etwa bei SCHMIDT / HERRGEN (2011, 50–53) zu finden sind.

in einem insgesamt eher strukturlosen Variantenmix äußern. Da diese Verdichtungsbereiche aus Kombinationen von miteinander auftretenden Varianten bestehen, weisen sie eine relative innere Homogenität auf (was eine Interpretation als Dialektgebiete nahelegt); sie sind jedoch nicht klar gegeneinander abgrenzbar, sondern „unscharf“, und können ineinanderfließen (und bilden somit ein Kontinuum):

[...] varieties in this continuum represent concentration areas, where a variety, though not clearly-cut separated from other varieties, is identified by a particular frequency of certain variants, by the co-occurrence of several features and possibly by some diagnostic traits, which appear in that variety only. (BERRUTO 2010, 236)

Dieses Konzept von unscharfen, aber dennoch erkennbaren Verdichtungsbereichen wurde mittlerweile in verschiedenen variationslinguistischen Untersuchungen zur Diaphasie angewandt, darunter LENZ (2003) und MÖLLER (2013), wobei diese beiden Arbeiten für das Ansetzen von Varietäten weitere Kriterien ins Feld führen.

ALEXANDRA LENZ (2003, 389–390) folgt dabei AUER (1986, 99), der zumindest für die diaphasische Ebene neben der inneren Homogenität und äußeren Abgegrenztheit noch weitere Kriterien für Varietäten anführt: das der eindeutig formulierbaren außersprachlichen Verwendungsbedingungen und das des emischen Status im Bewusstsein der Sprecher. Für den Nachweis einer Varietät wäre damit die Übereinstimmung einer 1. linguistisch homogenen und abgrenzbaren, 2. extralinguistisch genau definierbaren und 3. kognitiv repräsentierten Einheit zu zeigen. Mit diesen drei Zugängen – dem innersprachlichen, dem außersprachlichen und dem kognitiven – lassen sich im Prinzip drei verschiedene Arten von Varietäten beschreiben: Solche, die aufgrund von sprachlicher Homogenität und Abgegrenztheit (nur) linguistisch definiert sind; solche, die aufgrund von außersprachlichen Verwendungsbedingungen (nur) extralinguistisch definiert sind; und solche, die aufgrund von Sprecherurteilen (und anderen Indikatoren) (nur) laienlinguistisch beziehungsweise kognitiv definiert sind. Dabei ist klar, dass zwischen den so jeweils ermittelten Einheiten durchaus wechselseitige, nicht sofort durchschaubare Beziehungen bestehen können. Sie müssen daher meines Erachtens auch zunächst getrennt ermittelt werden; die Bedingungen der Deckungsgleichheit dieser Verfasstheiten sind dagegen ein eigener Untersuchungsgegenstand, etwa: Wann deckt sich ein sprachlich homogener Bereich mit einer im Bewusstsein der Sprecher so empfundenen sprachvariativen Einheit? Oder: Aufgrund welcher außersprachlichen Gegebenheiten lässt sich ein Bereich definieren, der nachweislich relative sprachliche Homogenität nach innen und Abgegrenztheit nach außen besitzt?

Aus diesem Grund werde ich das perzeptiv-kognitive Kriterium von AUER und LENZ in diesem Beitrag nicht ansetzen, da es epistemisch zu einer anderen Ebene gehört. Eine Varietät kann entweder wissenschaftlich-linguistisch, das heißt etisch, definiert werden, oder kognitiv-laienlinguistisch, das heißt emisch.

Zu versuchen, beide Blickwinkel gleichzeitig einzunehmen, wäre der Klarheit der Sicht auf die Dinge abträglich. In dieser Studie soll es um rein etisch verfasste Varietäten gehen, die mit emischen Konzepten nicht vermengt werden sollen.

Hier liegt es nahe, eine genaue Arbeitsdefinition des Begriffs der Varietät und verwandter Konzepte vorzunehmen. Insbesondere erscheint es hilfreich, eine Abgrenzung der beiden Bezeichnungen „Varietät“ und „Lekt“ vorzunehmen. Beide werden oft weitgehend synonym verwendet (etwa BUSSMANN 2008, 396, 772). Dabei werden sie entweder primär innersprachlich (zum Beispiel BERRUTO 2010) oder primär außersprachlich (etwa VEITH 2005, 24; BUSSMANN 2008, 396, 772) definiert, wobei eine gewisse innere Homogenität stets vorausgesetzt wird; bisweilen wird explizit einer „doppelten“ Definition der Vorzug gegeben (etwa AUER 1986, 99; LENZ 2003, 389–390; ROELCKE 2010, 16).<sup>6</sup>

Eine terminologische Unschärfe möchte ich vermeiden, indem ich außersprachlich definierte Ausschnitte aus dem Sprachraum „Lekte“ nenne, innersprachlich definierte Ausschnitte „Lekttypen“. Ein Lekttyp ist durch bestimmte sprachliche Varianten (und deren statistisch fassbare Kookkurrenz) bestimmt, betrifft deshalb nur ein Teilinventar und hat eine unscharfe Ausdehnung, da die Verteilungen der Merkmale nie ganz übereinstimmen; ein Lekt ist dagegen über außersprachliche Variationsparameter (z. B. Region, Schicht) festgelegt und hat deshalb eine zumindest potentiell scharfe Ausdehnung (z. B. ein bestimmter Ortsdialekt oder ein Regionaldialekt). Ein Lekttyp ist somit ein intensional bestimmter Ausschnitt aus dem Sprachraum, ein Lekt ein extensional bestimmter Ausschnitt. Von einer Varietät kann mit BERRUTO dann gesprochen werden, wenn der Geltungsbereich eines Lekttyps (d. h. einer signifikanten Kombination kookkurrenter Varianten) wesentlich mit einer oder mehreren außersprachlichen Gebrauchsbedingungen in Zusammenhang steht:

Wenn eine Menge von gewissen kongruierenden Werten bestimmter sprachlicher Variablen [...] zusammen mit einer gewissen Menge von Merkmalen auftreten, die Sprecher und/oder Gebrauchssituationen kennzeichnen, dann können wir von einer sprachlichen Varietät sprechen. (BERRUTO 2004, 189)

Eine konkrete Anwendung von BERRUTOS Varietätenmodell auf die Diatopie fehlt meines Wissens bislang. Es wäre zu diskutieren, ob auf diatopischer Ebene schon die Bestimmbarkeit der geografischen Verbreitung eines Lekttyps die Anforderungen an das außerlinguistische Kriterium des Varietätenstatus erfüllt (dann wäre jeder Lekttyp eine Varietät), oder ob dafür die Identifikation kongruenter außersprachlicher Merkmale erforderlich ist. Der Einfachheit halber werde ich im Folgenden nicht zwischen (Dia-)Lekttyp und (diatopischer) Varietät unterscheiden; in diesem Sinne ist eine „Varietät“ dann ganz allgemein ein sprachlich

<sup>6</sup> „Unter einer Varietät wird [...] ein sprachliches System verstanden, das [...] durch eine Zuordnung bestimmter innersprachlicher Merkmale einerseits und bestimmter außersprachlicher Merkmale andererseits gegenüber weiteren Varietäten abgegrenzt wird“ (ROELCKE 2010, 16).

definierter Ausschnitt aus dem sprachlichen Kontinuum, dem ein außersprachlicher Geltungsbereich zugewiesen werden kann.

Eine diatopische Varietät lässt sich nun entweder als (unscharfe) Menge von Orten (also als Dialektgebiet) oder als (ebenfalls unscharfe) Menge von Varianten (also als Inventar) beschreiben.<sup>7</sup> Kookkurrente und damit varietätskonstituierende Varianten können als die „Merkmale“ der jeweiligen Varietät bezeichnet werden, auch wenn sie zusammen nur einen Teil des sprachlichen Gesamtinventars abdecken und – einzeln betrachtet – durchaus abweichende Verbreitungen haben können. Somit ist eine Varietät durch die Menge ihrer Merkmale definiert, nicht durch ihren Anwendungsbereich.

Die Unschärfe einer Varietät würde sich zum Beispiel folgendermaßen äußern: Weist ein bestimmter Ortsdialekt etwa nur die Hälfte von ansonsten miteinander assoziierten Varianten auf, so könnte man sagen, dass die Varietät, die diese Varianten konstituieren, zu 50 Prozent vorliegt. Umgekehrt bedeutet dies, dass Varianten, die zwar vereinzelt zusammen mit den varietätskonstituierenden Varianten auftreten, insgesamt aber nicht in auffälliger Weise mit ihnen kookkurrieren, nicht als Bestandteile der Varietät anzusehen sind. Um es mit ADOLF BACH zu sagen:

Ein Dialekt ist also nicht zu bestimmen durch so und so viele Züge, die anderswo nicht vorkommen, er besteht vielmehr aus einer bestimmten Kombination von Erscheinungen, die jede für sich eine viel größere Verbreitung haben können, was nicht ausschließt, daß einzelne in der Tat nur in der betr. Kernlandschaft zu finden sind. (BACH 1950, 62)

Ein Variantenaufkommen, das exakt der Zusammenstellung der Merkmale einer Varietät entspricht, würde diese Varietät in prototypischer Reinform darstellen, auch wenn dies in der Realität nirgends ganz erreicht wird. Eine Varietät in der diatopischen Dimension entspricht demnach einem Dialekttyp, der durch ihre Merkmale charakterisiert ist und in Einzel(dia)lekten (z. B. Ortsdialekten) unterschiedlich stark repräsentiert ist. Einzelne Dialekte sind unterschiedlich typische Vertreter eines Dialekttyps; dadurch kommt den jeweiligen Orten eine unterschiedlich starke Zugehörigkeit zum entsprechenden Areal zu. Die Merkmale des Dialekttyps sind für ihn charakteristisch, da einzelne von ihnen stellvertretend für alle anderen, ähnlich verteilten Merkmale stehen können, die mit ihnen in unterschiedlichem Ausmaß kookkurrieren. Sie kommen dabei nicht nur im Kernbereich des betreffenden Dialektareals vor, sondern in abnehmenden Graden auch in seinen Rand- und umliegenden Bereichen. Dort stehen sie zunehmend neben Varianten, mit denen sie insgesamt gesehen eher nicht kookkurrieren. Vergleiche hierzu KÖNIG (2010a):

Die Übergänge zwischen den gesetzten, definierten Sprachgebieten gestalten sich in der Regel fließend, nicht abrupt. Sie gehen durch die Zu- oder Abnahme von für sie als cha-

<sup>7</sup> In unscharfen Mengen können die einzelnen Elemente unterschiedliche Grade an „Mitgliedschaft“ aufweisen. Zur Theorie der unscharfen Mengen (*fuzzy set theory*) vgl. ZADEH (1965); für die Dialektgeografie GIRARD / LARMOUTH (1993).



rakteristisch angenommenen Merkmalen ineinander über, wobei diese Merkmale meist nur in Teilen des Gebiets gelten oder noch weitere Gebiete einschließen können. So ist das als schwäbisch geltende „sch“ im Inlaut in Wörtern wie „Gascht“ oder „Fescht“ (für „Gast“ und „Fest“) auch im westlichen Bairischen (bis nach Pasing) daheim und umfasst den ganzen Südwesten bis nach Luxemburg [...]. Das oft als bairisches „Kennwort“ bezeichnete Fasching beginnt in den alten Ortsdialekten erst ein gutes Stück östlich von München, von da an aber gilt es bis zur ungarischen Grenze [...]. Das Wort ist also nicht in allen Ortsdialekten dessen, was man gewöhnlich „bairisch“ nennt, vorhanden.

Ein Ortsdialekt kann durch die Anteile beschrieben werden, zu denen er durch verschiedene Dialekttypen (bzw. durch deren Merkmale) geprägt ist. Die diatopische Schichtung von Varietäten lässt sich somit durch die Anteile ihrer Merkmale am lokalen Gesamtinventar beschreiben:

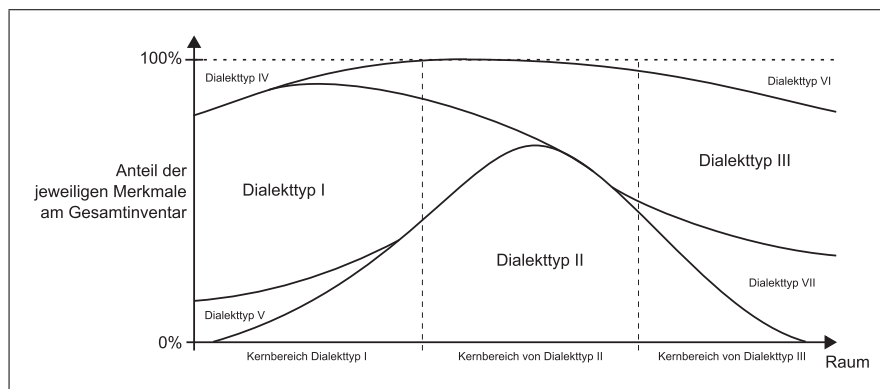


Abb. 1: Modell der Schichtung von Dialekttypen im Raum

Die Kernbereiche eines Dialekttyps sind die Regionen, in denen seine Merkmale den größten Anteil am sprachlichen Gesamtinventar ausmachen; er kann dort als „dominant“ bezeichnet werden. Doch selbst in seinem Kernbereich werden auch die Merkmale anderer Dialekttypen verwendet; diese sind dort lediglich nicht-dominant vorhanden. Umgekehrt tritt er als nicht-dominanter Dialekttyp im Kernbereich anderer Varietäten auf. Es ist auch vorstellbar, dass es Dialekttypen gibt, die ausschließlich nicht-dominant auftreten; solange sie sich durch eine eindeutige Verdichtungstendenz auszeichnen, ist dies ohne weiteres möglich.

Die Terminologie der auf ELEANOR ROSCH zurückgehenden Prototypentheorie, auf der auch die durch GEORGE LAKOFF (1987) begründete Prototypensemantik beruht, ist besonders gut geeignet, um Varietäten zu beschreiben. Dass laienlinguistische Dialektkonzeptualisierungen kognitive Prototypeneffekte aufweisen, wurde bereits verschiedentlich gezeigt (vgl. CHRISTEN 1998, 51–57; 2010; BERTHELE 2006, 164–167; KRISTIANSEN 2008; ANDERS 2010, 107–110). Solche Konzepte sind unscharfe Kategorien, in die sich einzelne Lekte (mit variierender Stringenz) einteilen lassen. Einzellekte können also mehr oder weniger typische Vertreter

einer Varietät – eines Lekttyps – sein. Dabei wäre zu unterscheiden zwischen emischen (d. h. kognitiv-laienlinguistischen) und etischen (d. h. wissenschaftlich-dialektologischen) Lekttypen. Die Frage, inwieweit sich beide decken und welche Beziehung zwischen ihnen besteht, stellt einen eigenen Forschungsgegenstand dar. Die wissenschaftlich-dialektologische Identifikation von Dialekttypen lässt sich jedenfalls empirisch über die Untersuchung von Kookkurrenzen unter Varianten erreichen.

Ein erster Schritt dorthin liegt in der Operationalisierung des etischen Dialekttyp-Begriffs, das heißt in der Bereitstellung einer Methodik für die Identifikation von Kookkurrenzen unter diatopischen Varianten. Eine solche Methodik wird im nächsten Abschnitt vorgestellt.

### 3. Faktorenanalyse zur Identifikation von Dialekttypen

Die sogenannte Faktorenanalyse ist ein Standard-Analyseverfahren aus der multivariaten Statistik, das auf Daten- beziehungsweise Dimensionsreduktion beruht. Details zur Funktionsweise der Faktorenanalyse sowie Informationen über das benötigte Datenformat und anderes finden sich zum Beispiel in TABACHNICK / FIDELL (2007, 607–627), VOGELBACHER (2011, 21–43) und PICKL (2013, 158–167). DOUGLAS BIBER, der die Faktorenanalyse in der Variationslinguistik erstmals verwendet hat, fasst ihre Arbeitsweise wie folgt zusammen:

In a factor analysis, a large number of original variables, in this case the frequencies of linguistic features, are reduced to a small set of derived variables, the ‘factors’. Each factor represents some area in the original data that can be summarized or generalized. That is, each factor represents an area of high shared variance in the data, a grouping of linguistic features that co-occur with a high frequency. (BIBER 1991, 79)

Die grundlegende Arbeitsweise der Faktorenanalyse ist es also, die Varianz einer großen Menge an Variablen (im mathematischen Sinne) auf wenige zugrundeliegende „Faktoren“ zurückzuführen. Angewandt auf unser Problem bedeutet dies: Wenn mehrere sprachliche Varianten dazu tendieren, gemeinsam aufzutreten (d. h. zu kookkurrieren), dann kann man diese zusammenfassen (verdichten), indem man ihre räumlichen Verteilungen durch einen einzelnen, repräsentativen Faktor ausdrückt. Dadurch lassen sich die räumlichen Verteilungen einer Vielzahl von Varianten durch einige wenige Faktoren beschreiben. Diese Faktoren stellen dann Variantenbündel oder -verdichtungen dar.

In der Sprachgeografie wird die Faktorenanalyse – und die damit eng verwandte Hauptkomponentenanalyse – meines Wissens erst seit ein paar Jahren verwendet;<sup>8</sup> Anwendungen in anderen humangeografischen Bereichen gibt es

<sup>8</sup> Für Vorzüge der Faktorenanalyse gegenüber der Hauptkomponentenanalyse bei dialektgeografischen Anwendungen siehe z. B. LEINO / HYVÖNEN (2009).

bereits seit den 1960er-Jahren (z. B. STEINER 1965). ROBERT SHACKLETON (2005) verwendet die Hauptkomponentenanalyse zur Untersuchung britischer und amerikanischer Dialekte, wobei er die ersten beiden Hauptkomponenten betrachtet, ohne sie kartografisch darzustellen; in SHACKLETON (2007), einer phonetischen Untersuchung traditioneller englischer Dialekte, betrachtet er bis zu vier Komponenten, die er einzeln kartiert. Der Fokus liegt hierbei auf der Identifikation von sprachraumbildenden phonetischen Varianten (wie auch in NERBONNE 2006, einer Untersuchung amerikanischer Aussprachedaten mithilfe der Faktorenanalyse). HYVÖNEN / LEINO / SALMENKIVI (2007) vergleichen die Ergebnisse der Anwendung von Clusteranalyse, Hauptkomponentenanalyse und anderen Verfahren auf lexikalische Daten aus Finnland und kommen im direkten Vergleich zum Schluss, dass „clustering might not be the best way to view dialect variation. Looking at the results of a component analysis might in fact be more informative. After all, borders between different dialects are usually not steep, but vocabulary changes gradually“ (HYVÖNEN / LEINO / SALMENKIVI 2007, 288). THERESE LEINONEN (2009; 2010) ermittelt bei der Untersuchung akustischer Daten schwedischer Dialekte mit der Faktorenanalyse räumliche Strukturen; dadurch, dass mit diesen Strukturen bestimmte Varianten assoziiert sind, werden gleichzeitig systematische Relationen zwischen sprachlichen Varianten sichtbar. So beschreibt etwa der erste Faktor in erster Linie den Öffnungsgrad der geschlossenen Vokale [i:], [ɪ], [y:], [ʏ], [e:] und [e], wobei die offeneren Realisierungen eine zentral-ostschwedische Verdichtung bilden.

Diese Beispiele aufschlussreicher Anwendungen von Faktoren- und Hauptkomponentenanalyse zeigen, dass diese Methoden zum einen zum Auffinden (auch unscharfer) sprachräumlicher Strukturen geeignet sind, die auf Verdichtungen von Varianten beruhen, zum anderen aber auch diese räumlich aneinander gekoppelten Varianten adressierbar machen und dadurch auf systematische linguistische Zusammenhänge hinweisen.<sup>9</sup> Mit der Faktorenanalyse wird in diesem Beitrag ein quantitatives Verfahren gewählt, bei dem nicht wie in strukturell orientierten Verfahren (wie etwa bei WIESINGER 1983) *a priori* entschieden werden muss, welche sprachlichen Phänomene „systemhafte und damit typologische Struktureigenschaften“ (siehe oben) besitzen, durch die sie aufgrund ihrer Stellung im Sprachsystem stellvertretend für viele weitere Phänomene stehen können. Bei der Faktorenanalyse als quantitativem Verfahren ist es nicht wie bei strukturell orientierten erforderlich, die „Wichtigkeit“ sprachlicher Phänomene im Vorhinein zu bewerten. Systemhafte Züge zeichnen sich gerade dadurch aus, dass Teile des Sprachinventars aufeinander bezogen und dadurch aneinander gekoppelt sind, wodurch eine größere Menge von Varianten ähnliche Verteilungen haben müssen, was sich wiederum quantitativ unmittelbar bemerkbar macht. So schlagen

<sup>9</sup> Weitere Beispiele sind u. a. CLOPPER / PAOLILLO (2006), LEINO / HYVÖNEN (2009), GRIEVE (2009), GRIEVE / SPEELMAN / GEERAERTS (2011) und SZMRECSANYI / WOLK (2011).

sich sprachliche Eigenschaften mit Systemcharakter durch die Gleich- beziehungsweise Ähnlichverteilung einer gewissen Menge von Varianten quasi „von selbst“ nieder. Dabei ist die Feststellung solcher strukturellen Assoziationen unter Varianten neben der Identifikation von Raumeinheiten ein weiteres, genuin linguistisches Ergebnis der Methode.

Die Datengrundlage für die Faktorenanalyse sind in unserem Fall die relativen Frequenzen sprachlicher Varianten in ihrer räumlichen Verteilung, die zwischen null und hundert Prozent liegen.<sup>10</sup> Sie geben an, wie häufig diese im Vergleich mit anderen Varianten derselben Variablen – zum Beispiel an einem bestimmten Ort – aufgezeichnet wurden (und lassen daher nur indirekt Schlüsse auf ihre tatsächliche Verwendungshäufigkeit zu). Neigen verschiedene Varianten zur Kookkurrenz, so korrelieren ihre relativen Auftretenshäufigkeiten, und die Faktorenanalyse fasst sie zu einem Faktor zusammen. Bei der Faktorenanalyse handelt es sich also um ein Verfahren, mit dem man multiple (d. h. zwischen mehreren Varianten auftretende) Korrelationen in den Daten ermitteln kann (d. h. in unserem Fall: Bündel von kookkurrenten Varianten). Das bedeutet, man klassifiziert wiederkehrende räumliche Verbreitungsmuster, ohne Vorannahmen darüber treffen zu müssen, a) dass und b) welche Strukturen in den Daten bestehen. Insofern ist der Ansatz *data-driven*; die Faktorenanalyse ist eine Data-Mining-Methode, die es erlaubt, einen großen Bestand an Daten explorativ nach latenten Strukturen zu durchleuchten. Diese latenten Strukturen stellen nichts anderes dar als Verdichtungsbereiche im sprachgeografischen Kontinuum. Dies erlaubt einerseits ihre Interpretation als Varietäten im Sinne BERRUTOS beziehungsweise als Dialekttypen, andererseits Rückschlüsse auf sprachdynamische Prozesse, die solche Strukturen erzeugen.

Die Resultate der Faktorenanalyse bestehen zunächst aus einer Anzahl Faktoren, die je einen Teilbereich der sprachgeografischen Variation im Untersuchungskorpus kompakt und kondensiert beschreiben. Interessant und aussagekräftig sind meist nur die „stärksten“ unter ihnen, das heißt diejenigen, die die größten Anteile der Gesamtvariation enthalten. Für die Entscheidung, welche Faktoren in die Interpretation einbezogen werden, gibt es verschiedene Kriterien (vgl. PICKL 2013, 160–161), die hier nicht näher erläutert werden. Die individuellen Faktoren haben die Form je zweier Datenmatrizen, eine mit den sogenannten Faktorladungen und eine mit den sogenannten Faktorwerten. Die Faktorladungen geben für jeden Ort an, wie stark der dortige Ortsdialekt durch das Vorkommen der Varianten der verschiedenen Faktoren geprägt ist; mit anderen Worten, wie viele Merkmale des entsprechenden Dialekttyps dort vorkommen oder wie typisch der Ortsdialekt als Vertreter dieses Dialekttyps ist. Die Faktorladungen lassen sich kartieren und visualisieren so die räumliche Ausdehnung einer Varietät. Die Faktorwerte geben für jede sprachliche Variante in der Datensammlung an,

<sup>10</sup> Die relativen Frequenzen wurden im vorliegenden Fall durch einen einfachen Vorverarbeitungsschritt aus den Belegdaten errechnet (vgl. PICKL 2013, 83–87, 159–160).

wie stark ihre individuelle räumliche Verteilung durch die einzelnen Faktoren geprägt ist, das heißt in welchem Ausmaß sie mit den Ausdehnungen der jeweiligen Varietäten übereinstimmt; mit anderen Worten, wie charakteristisch sie als Merkmal der einzelnen Dialekttypen ist. Ein Faktor verfügt daher über die unscharfe räumliche Ausdehnung einer Varietät und enthält den Teil des sprachlichen Inventars, der diese Varietät konstituiert.

In Abschnitt 1 wurde neben der Faktorenanalyse ein weiterer Kandidat für eine Operationalisierung der Ermittlung „unscharfer“ Areale erwähnt: das *fuzzy clustering*. An dieser Stelle soll kurz erläutert werden, warum der Faktorenanalyse der Vorzug gegeben wurde.

Ein Grund liegt in Unterschieden in der Arbeitsweise der beiden Verfahren. Das *fuzzy clustering* arbeitet vom Grundprinzip her wie eine herkömmliche Clusteranalyse: Einzelne Ortspunkte werden sukzessive zu immer größeren Gruppen zusammengefasst, ausgehend von ihrer dialektalen Ähnlichkeit. Da die Clusteranalyse als sensibel gegenüber kleinen Änderungen im Datenmaterial gilt, leiden ihre Ergebnisse an einer gewissen Unsicherheit. So klar die Darstellung der ermittelten Dialektareale in den einschlägigen Abbildungen ist, so unsicher ist die Zuordnung einzelner Ortspunkte zum einen oder zum anderen Areal. Diese Schwäche versucht man zu überwinden, indem die Clusteranalyse mehrfach mit jeweils leicht veränderter Datengrundlage durchgeführt wird, um die Zuverlässigkeit der Zuordnung einschätzen zu können. Wird ein Ort immer demselben Areal zugeordnet, so ist diese Einteilung sehr schlüssig und der Ort liegt wohl in einem Kerngebiet; wird ein Ort mal diesem, mal jenem Gebiet zugeordnet, so sind diese Zuordnungen nicht eindeutig und der Ort liegt wahrscheinlich in einem Übergangsbereich. Während so beim *fuzzy clustering* durchaus Zonen mit unterschiedlicher Gebietszugehörigkeit zu ermitteln sind, ist nicht ohne weiteres klar, ob die Cluster Verdichtungen im Sinne BERRUTOS, das heißt Kookkurrenzbündel darstellen. So ist es zum Beispiel denkbar, dass in einem sanften Übergangsbereich zwischen zwei Verdichtungsbereichen die sprachlichen Ähnlichkeiten unter den Orten so groß sind, dass dort ein eigenständiger Cluster angesetzt wird, der aber nicht von wesentlichen Kookkurrenzen geprägt ist. Bei der Faktorenanalyse ist das immer gewährleistet.

Auf ein anderes Problem weist WOLFGANG PUTSCHKE (1993) hin: Da es keine einheitlichen Richtlinien für die zentrale Frage der Clusteranzahl gibt, unterliegen die Ergebnisse der Clusteranalyse einem gewissen Verdacht der Subjektivität;<sup>11</sup> ein Problem, das bei der Faktorenanalyse zwar auch besteht, aber bei weitem nicht dieselbe Tragweite besitzt, da dort ein Mehr oder Weniger an Faktoren ab einer gewissen kritischen Masse nur zu geringfügigen Änderungen führt.

<sup>11</sup> „Die Problematik dieses Verfahrens liegt in der Kappung der Dendrogrammstränge, die jeweils auf unterschiedlichem Wertenniveau vorgenommen wurde. Da diese Verclustering nicht völlig automatisch durchzuführen ist, geht das Vorwissen des Dialektologen in seine Entscheidungen ein und steuert somit auch den gesamten Klassifizierungsprozeß.“ (PUTSCHKE 1993, 429)

Ein weiterer Punkt betrifft den Bezug der arealen Einheiten zur sprachlichen Grundlage. Da die Clusteranalyse immer auf einer aggregierten Ähnlichkeits- beziehungsweise Distanzmatrix arbeitet, werden Unterschiede zwischen einzelnen Variantenverteilungen eingeebnet. Ein Zurückführen der Cluster auf die sprachlichen Phänomene, die für sie verantwortlich sind, ist deshalb nur über Umwege möglich.<sup>12</sup> Bei der Faktorenanalyse ist hingegen jeder Faktor durch seine Faktorwerte direkt auf die durch ihn repräsentierten (bzw. die für ihn verantwortlichen) Varianten zurückführbar. Dies allein ist ein nicht zu unterschätzender Vorteil. Je mehr Varianten dies sind, um so „stärker“ ist der Faktor. Ein Faktor, der nur auf relativ wenige Varianten zurückgeht, ist schwächer als einer, der die Verteilungen sehr vieler Varianten repräsentiert. So können mit der Faktorenanalyse auch nicht-dominante Verdichtungen ermittelt werden, bei der Clusteranalyse ist dies nicht der Fall; dort treten nur dominante Strukturen in Erscheinung.

#### 4. Wortgeografische Verdichtungen in Bayerisch-Schwaben

Die folgende Anwendung der Faktorenanalyse wurde mit Daten des Sprachatlas von Bayerisch-Schwaben (SBS) durchgeführt, der in den Jahren 1984 bis 2009 unter der Leitung von WERNER KÖNIG an der Universität Augsburg erstellt wurde.<sup>13</sup> Der SBS dokumentiert die dialektale Variation im bayerischen Regierungsbezirk Schwaben (ohne jene Teile im Süden, die bereits durch den Vorarlberger Sprachatlas [VALTS] abgedeckt werden) und angrenzenden Gebieten in Oberbayern und Mittelfranken. Primäres Ziel dieses Projekts war es, den Basisdialekt beziehungsweise die älteste noch erreichbare Mundart in ihrer geografischen Variation festzuhalten. Hierfür wurden zwischen 1984 und 1989 ältere Personen befragt, vor allem aus dem bäuerlichen Milieu. Mit circa 2.700 Karten in 14 Bänden ist der SBS der umfangreichste Dialektatlas im deutschen Sprachraum. Das Untersuchungsgebiet erstreckt sich circa 90 km von West nach

<sup>12</sup> „Da diese Kritik genuin mit der Methode selbst verbunden ist, sind Verbesserungen nur über verfahrensmäßige Änderungen zu erzielen; die Anforderungen hierzu sind weitgehend klar und bestehen in einem Mitführen des sprachlichen Belegs. Dies läuft im Kern auf andere Verfahren hinaus, die nicht die Identität der Sprachdaten[,] sondern die Identität der Verbreitungsflächen benutzen, um so über die Metrisierung von Arealitäten zu klassifikatorischen Aussagen zu gelangen.“ (PUTSCHKE 1993, 428)

<sup>13</sup> Die vorgestellte Anwendung der Faktorenanalyse auf Daten des Sprachatlas von Bayerisch-Schwaben wurde im Rahmen eines interdisziplinären Projekts zwischen dem Lehrstuhl für Deutsche Sprachwissenschaft (Universität Augsburg) und dem Institut für Stochastik (Universität Ulm) realisiert, dem DFG-Projekt „Neue Dialektometrie mit Methoden der stochastischen Bildanalyse“ (unter der Leitung von STEPHAN ELSPASS, WERNER KÖNIG, VOLKER SCHMIDT und EVGENY SPODAREV). Die technische Umsetzung und Durchführung der Faktorenanalyse folgt in weiten Teilen der Diplomarbeit von JULIUS VOGELBACHER (2011, 21–41), dem ich herzlich für die Implementierung und Anwendung der Faktorenanalyse danke. Technische Details können dort und in PICKL (2013, 158–167) nachgelesen werden.

Ost und etwa 150 km von Nord nach Süd und umfasst 272 Erhebungsorte. Für jedes sprachliche Merkmal liegen im Normalfall für jeden Ort Antworten von einer Gewährsperson vor. Zusätzliche Informanten wurden jeweils nur dann hinzugezogen, wenn der erste Informant eine Frage nur unzureichend beantworten konnte. Details zur Informantenauswahl, Datenerhebung et cetera lassen sich in Band 1 des SBS nachlesen; sie sind für die vorliegende Demonstration des in Abschnitt 3 vorgestellten Verfahrens nicht weiter relevant. Weitere Informationen zur Datenvorverarbeitung finden sich in PICKL (2013, 72–78, 83–87, 159–160).

Für die vorliegende Studie wurde der wortgeografische Teil des SBS ausgewählt, da er von seiner Datenstruktur her leichter zu handhaben war als etwa die Laut- oder die Formengeografie. Zudem lagen die Daten der anderen sprachlichen Teilbereiche zum Zeitpunkt des Verfassens dieses Artikels noch nicht vollständig in ihrer benötigten Form vor. Dass es sich dabei um einen sprachlichen Teilbereich handelt, der als relativ „unsystematisch“ gilt (vgl. unter anderen GOOSSENS 1969, 53, 69; VIERECK 1986, 725; LABOV / ASH / BOBERG 2006, 41, 119), ist nicht von Nachteil, im Gegenteil: So wird demonstriert, dass das Vorgehen und der dahinterstehende Varietätenbegriff nicht strukturell orientiert ist (wie etwa bei WIESINGER 1983 oder SCHMIDT / HERRGEN 2011, 50–53; vgl. auch LÖFFLER 2003, 128–130), sondern primär quantitativ. Alle folgenden Ausführungen beschränken sich daher auf die lexikalische Ebene (beziehungsweise auf einen Ausschnitt davon); quantitative Dialektstudien sind immer von der Auswahl des Kartenkorpus abhängig, und ihre Aussagen lassen sich nur teilweise von Teilbereichen auf die gesamte sprachliche Variation verallgemeinern. Einer Einbeziehung weiteren, nicht-lexikalischen Materials für zukünftige Studien steht jedenfalls prinzipiell nichts im Wege (siehe PRÖLL / PICKL / SPETTL i. E.). In seiner digitalisierten Form umfasst das verwendete Kartenkorpus die Daten von 735 Wortkarten (also sprachliche Variablen), die die räumliche Verteilung von insgesamt 12.341 Ausdrücken (d. h. Varianten) beschreiben.

Der Dateninput besteht somit zunächst aus dem quantifizierten Vorkommen von 12.341 Varianten an 272 Orten. Die Faktorenanalyse ermittelt nun, welche dieser 12.341 Varianten kookkurrieren, und fasst diese zu Faktoren zusammen. Im vorliegenden Fall wurden 20 Faktoren ermittelt,<sup>14</sup> die zusammen 59,32 Prozent der Gesamtvariation in der Datensammlung umfassen. Die restlichen zwei Fünftel der Variation lassen sich nicht so zu Faktoren zusammenfassen, dass sie sinnvoll interpretiert werden können.

<sup>14</sup> Diese Zahl wurde im vorliegenden Fall durch eine Kombination aus dem sogenannten Kaiser-Kriterium mit einem von DANIEL WARTENBERG (1985) vorgeschlagenen Verfahren bestimmt (vgl. PICKL 2013, 160–161).

## 4.1. Einzelfaktoren

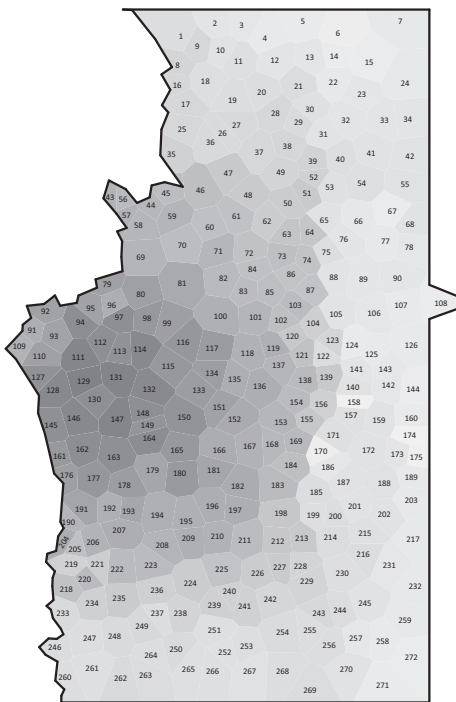


Abb. 2: Faktor 4

Abbildung 2 zeigt einen der 20 Faktoren (Faktor 4)<sup>15</sup> und damit einen empirisch ermittelten lexikalischen Verdichtungsbereich innerhalb des Untersuchungsgebiets. Er beschreibt 11,80 Prozent der Gesamtvariation und ist damit insgesamt der prominenteste; lokal (an einem einzelnen Ort) beschreibt er maximal 50 Prozent der dortigen Variation (Oberreichenbach [129]). Die Grautöne visualisieren die Faktorladungen der einzelnen Ortsdialekte. Je dunkler ein Ort dargestellt ist, umso mehr Varianten wurden dort aufgezeichnet, die auch an den anderen grauen Orten belegt sind. Insgesamt stellt diese Abbildung also einen nicht scharf umgrenzten Bereich dar, in dem bestimmte Varianten kookkurrieren. Welche Varianten dies sind, lässt sich auf wenigen Seiten nicht beantworten, da jede der 12.341 Varianten einen eigenen Faktorwert für Faktor 4 (F4) besitzt, der Auskunft darüber gibt, wie stark ihre Verteilung der des Faktors ähnelt (und damit aller anderen durch ihn repräsentierten Varianten). Die in Abbildung 3 abgebildeten Variantenverteilungen etwa haben verschiedene Faktorwerte für F4.

<sup>15</sup> Die Nummerierung richtet sich nach den lokalen Maxima, d.h. nach den höchsten Faktorladungen eines Faktors im Untersuchungsgebiet.



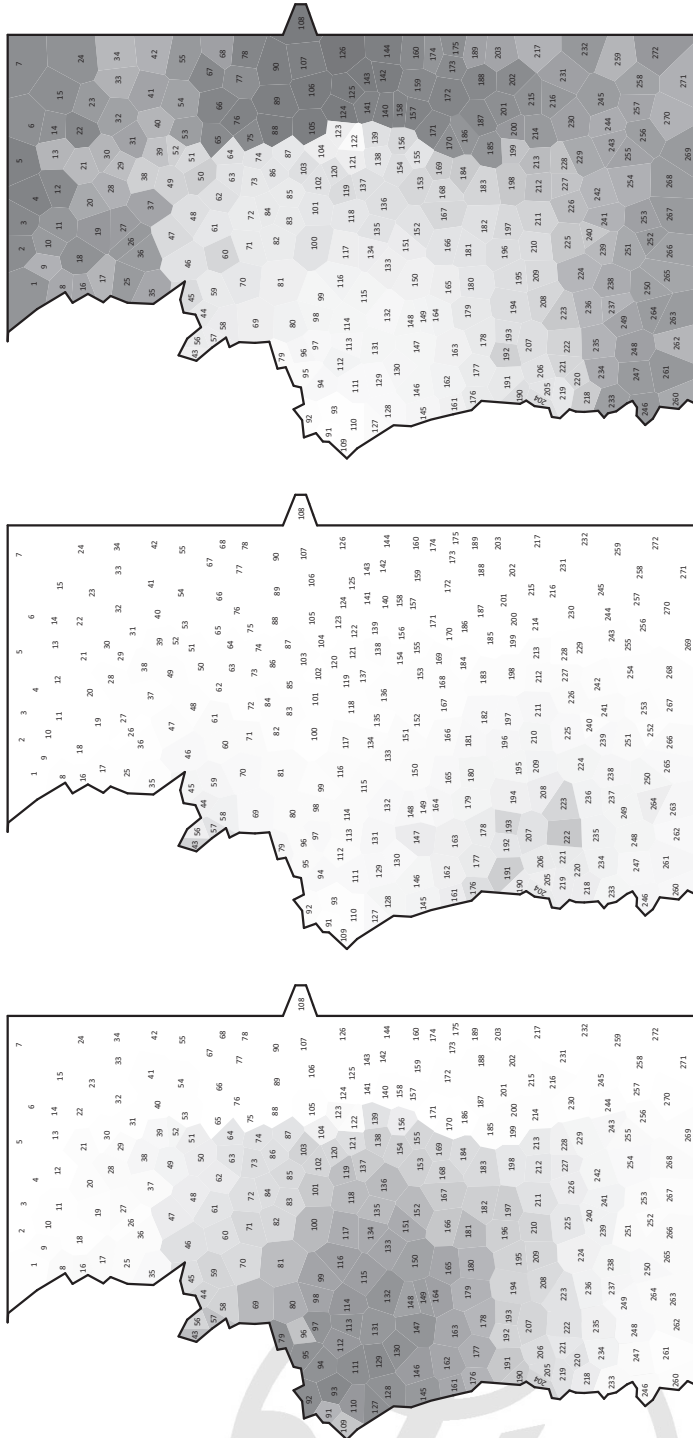


Abb. 3a: *däuen* 'wiederkäuen' (SBS, Bd. 11, K. 49)

Abb. 3b: (-) *süßern* 'die länger andauernde Reimung nach dem Kalben' (SBS, Bd. 11, K. 31)

Abb. 3c: *Erbse* 'Erbsen' (SBS, Bd. 8, K. 89)

Abb. 3: Die geografischen Verteilungen verschiedener lexikalischer Varianten aus dem SBS. Die Verteilungen wurden mittels sogenannter Intensitätsschätzung für die Darstellung optimiert (vgl. RUMPF et al. 2009, 285–288; PICKL / RUMPF 2011, 272–274; PICKL 2013, 86–93, 102–105)

Die Variante *däuen* (zur Variablen ‘wiederkäuen’) in Abbildung 3a hat mit 9,13 einen sehr hohen Faktorwert; entsprechend ist ihre Verteilung der von F4 sehr ähnlich. Die Variante (-) *säubern* (‘die länger anhaltende Reinigung nach dem Kalben’) hat einen Faktorwert nahe 0 für F4, das heißt ihre Verteilung hat wenig mit der seiner Faktorladungen zu tun. Die Variante *Erbes* (‘Erbse’) schließlich hat mit  $-3,72$  einen sehr niedrigen, negativen Faktorwert; ihre Distribution ist daher praktisch invers zu der von F4, das heißt diese Variante tendiert dazu, gerade da aufzutauchen, wo die Varianten von F4 nicht verwendet werden.

Wie Abbildung 4 zeigt, liegen die allermeisten der Faktorwerte nahe 0. Positive Werte haben aber immerhin noch 2.209 Varianten. Um einen Eindruck zu bekommen, durch welche Varianten F4 besonders geprägt ist, liegt es daher nahe, nur die extremen Werte zu betrachten. In Tabelle 1 sind die Varianten mit den 20 höchsten Faktorwerten für F4 abgebildet, deren räumliche Verbreitung am stärksten mit der von F4 übereinstimmt, und die deshalb als Merkmale des entsprechenden Dialekttyps gelten können.

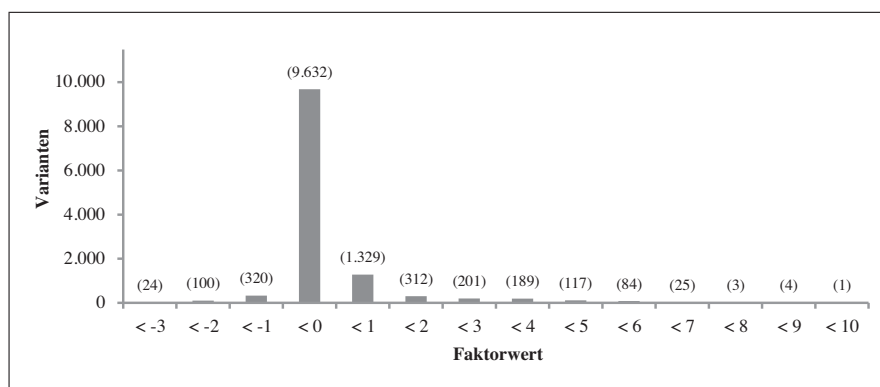


Abb. 4: Histogramm der Faktorwerte für F4

Abgesehen von einigen landwirtschaftlichen Fachtermini, die wohl keinen hohen Wiedererkennungswert haben, lassen sich wohl vor allem die Varianten *guck*, *Nähne(in)*, *Kirchweih* und *schwätzen* für Einheimische innerhalb des Untersuchungsgebiets durchaus mit dem durch F4 repräsentierten Verbreitungstyp in Verbindung bringen (wobei *schwätzen* auch bis an den Südrand des Untersuchungsgebiets vorkommt). Die Ausdehnung des Faktors (und die der mit ihm assoziierten Varianten) findet sich in ähnlicher Form in klassischen Dialekteinteilungskarten unter der Bezeichnung „Ostmittelschwäbisch“ oder „mittleres Ostschwäbisch“ wieder (vgl. z. B. NÜBLING 1988, 118; SBS, Bd. 1, K. 8).

Ein weiteres Beispiel ist F1 (Abbildung 5), der global (d. h. im ganzen Untersuchungsgebiet) 11,52 Prozent und lokal maximal (in Dietmannsried [248]) 60 Prozent der Variation beschreibt. Seine Ausdehnung, die ihren Kernbereich im Südwesten des Untersuchungsgebiets hat und sich nach Norden und Osten

Tab. 1: Die Varianten mit den 20 höchsten Faktorwerten für F4

FW	Variante	Variable	Karte
9,13	<i>eindäuen</i>	‘wiederkäuen’	11.49
8,57	<i>rosseln/rösseln</i>	‘brünstig sein (von der Stute)’	11.112
8,24	<i>Werre</i>	‘Gerstenkorn’	2.10
8,12	<i>(-)Poppel</i>	‘Wollknäuel’	10.45
8,03	<i>Wargelholz</i>	‘Nudelholz’	10.66
7,50	<i>schmatzgen</i>	‘gurgeln (in nassen Schuhen)’	8.50
7,41	<i>Kornhaus</i>	‘Dachboden (über dem Wohnhaus)’	8.4
7,39	<i>Lachefass</i>	‘Jauchefass auf dem Wagen’	12.51
6,98	<i>räu(n)sen/räusig</i>	‘brünstig sein (von der Sau)’	11.73
6,92	<i>schillen/schielen</i>	‘schielen’	2.69
6,90	<i>guck</i>	‘schau!’	2.68
6,82	<i>Nähle(in)</i>	‘Großvater’	2.99
6,80	<i>Liecher</i>	‘Heurupfer’	12.96
6,77	<i>einschläfen/anschläfen</i>	‘(Kleidung) anziehen’	2.148
6,74	<i>Schullehr(e)r</i>	‘Schullehrer’	2.111
6,69	<i>Buschel</i>	‘Bündel ausgedroschenen Strohs’	12.132
6,68	<i>Kirchweih</i>	‘Kirchweihfest’	2.135
6,60	<i>mens gehen/stehen/ sein</i>	‘die Kuh ein Jahr lang mit dem Kalben aussetzen lassen’	11.35
6,59	<i>schwätzen</i>	‘reden’	2.59
6,54	<i>raig(g)el(en)</i>	‘spannen (von Seil oder Kette beim Holztransport)’	13.58

„ausdünn“, entspricht in etwa der der Landschaft „Allgäu“. Ihr kann ebenfalls keine scharf umrissene Ausbreitung zugesprochen werden, sondern sie zeichnet sich durch gerade nach Norden hin eher fließende Grenzen aus (vgl. KÖNIG 1976; CRÄMER 1954, v. a. 9–11), ebenso der damit in Verbindung gebrachte Sprachtyp.

Hier sind die Varianten *lug(e)*, *heiren*, *jucken* und *dornäglen/dur(ch)nägeln* (vgl. Tabelle 2) als auch perzeptiv charakteristisch „allgäuerische“ Varianten und damit als Merkmale eines allgäuerischen (wenngleich vorerst nur auf wortgeografischer Basis ermittelten) Dialektyps einzuschätzen. Empirisch gesehen handelt es sich auch bei den anderen Varianten um Merkmale dieses Dialektyps, auch wenn sie kognitiv wohl nicht als solche anzusehen sind.

Wie auch bei F4 ansatzweise erkennbar, ist in der Südosthälfte des Untersuchungsgebiets eine erkennbare Linie oder Kante festzustellen, bei der die Faktorladungen schlagartig nach Osten hin abfallen, was gegenüber ihrem ansonsten eher graduellen Zurückgehen bemerkenswert ist. Der Verlauf dieser Linie fällt mit dem Lech zusammen, der als starke Dialektscheide bekannt ist und als Grenze zwischen den alemannischen und den bairischen Dialekten gesehen wird. Die Barrierewirkung des Lechs manifestiert sich in Form einer Art „Bruch“ im

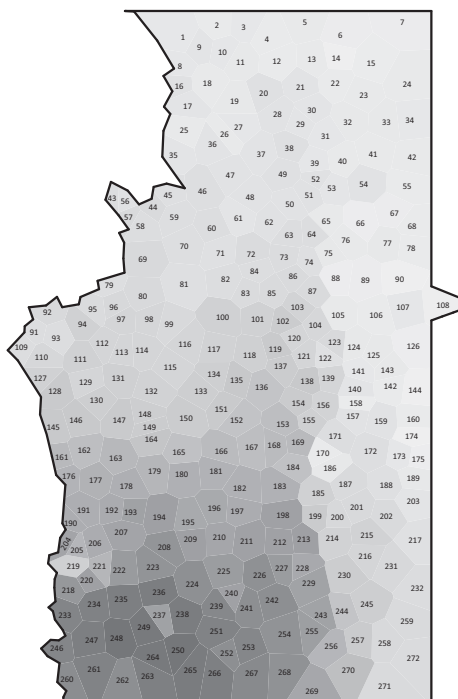


Abb. 5: Faktor 1

Tab. 2: Die Varianten mit den 20 höchsten Faktorwerten für F1

FW	Variante	Variable	Karte
7,48	<i>eindrucken</i>	‘wiederkäuen’	11.49
7,34	<i>birlingen</i>	‘Heuhaufen machen’	12.72
7,31	<i>lug(e)</i>	‘schau!’	2.68
7,28	<i>heiren</i>	‘heiraten’	2.115
7,25	<i>Schlot(t)er</i>	‘ungekochte saure Milch’	11.68
7,20	<i>steinen</i>	‘Wiese von Steinen säubern’	12.41
7,20	<i>Lucke</i>	‘Durchlass auf der Viehweide (vorwiegend mit verschiebbaren Stangen)’	13.69
7,16	<i>jucken</i>	‘hüpfen’	2.73
7,04	<i>(Wäsch<sup>e</sup>-)Kluppe</i>	‘Wäscheklammer’	10.41
6,99	<i>Überladet</i>	‘Schicht auf dem Heuwagen’	12.84
6,97	<i>Hahl(en)</i>	‘Deichselarme/Hahl (zur Vorderachse)’	13.83
6,96	<i>B<sup>e</sup>schütte</i>	‘Jauche’	12.47
6,93	<i>Halde</i>	‘steiler Grashang’	12.4
6,93	<i>b<sup>e</sup>schütten</i>	‘Jauche ausbringen’	12.48
6,84	<i>Aufbietgabel</i>	‘Aufladgabel’	12.83

FW	Variante	Variable	Karte
6,83	<i>futtern</i>	‘das Vieh füttern’	11.44
6,74	<i>Bu(r)zenstengel</i>	‘Purzelbaum’	10.13
6,64	<i>dornäglen/ dur(ch)nägeln</i>	‘Schmerzen der erfrorenen Fingerspitzen’	2.24
6,62	<i>Loreie</i>	‘kleine Reihen von halbdürrem Heu’	12.69
6,57	<i>Bollenkarren</i>	‘einrädrieger Karren (mit Holzkiste)’	13.106

Kontinuum; an dieser Schwelle treten sprachliche Unterschiede nicht graduell, sondern plötzlich auf. Man könnte von einer kristallinen Makrostruktur innerhalb des Sprachraums sprechen.

Ähnlich wie die Grenzwirkung eines Flusses können auch kleinere sprachgeografische Auffälligkeiten, die sich dem Auge der Dialektologin / des Dialektologen beim Betrachten eines Sprachatlas und sogar beim Durchführen gängiger dialektometrischer Verfahren verschließen, mithilfe der Faktorenanalyse zum Vorschein gebracht werden. F2 (Abbildung 6; global/lokal erklärte Varianz: 9,22 Prozent / 56 Prozent) bildet einen wortgeografischen Verdichtungsbereich im Nordwesten des Untersuchungsgebiets ab, dessen Ausdehnung ungefähr mit einem in der Literatur als „Nordostschwäbisch“ bezeichneten Dialektgebiet identifiziert werden kann (vgl. unter anderen NÜBLING 1988, 118, SBS, Bd. 1, K. 8). Auf den Abdruck der 20 charakteristischsten Merkmale wird aus Platzgründen verzichtet (siehe aber PICKL 2013, 173). Der Kernbereich dieses Dialektyps fällt zum einen ungefähr mit der topografischen Formation des Nördlinger Rieses zusammen, einem prähistorischen Meteoritenkrater, der eine deutlich von der Umgebung verschiedene Landschaft ausgebildet hat. Einen Zusammenhang dieser geologischen Formation mit dem sprachlichen Verdichtungsbereich F2 anzunehmen, wäre aber reine Spekulation. Plausibler (und von der geografischen Ausdehnung her besser passend) ist es, eine Beziehung zur ehemaligen Grafschaft Oettingen herzustellen, einem über Jahrhunderte relativ stabilen Territorium, das sich nicht nur durch politische, sondern auch vor allem durch konfessionelle Isolation auszeichnet: Das Bekenntnis in der Grafschaft war über lange Zeit mehrheitlich evangelisch, was sie vom katholisch geprägten Umland unterscheidet. Diese These wurde in PICKL (2013, 153–154) mit anderer Methodik bestätigt.

Die eigentliche Auffälligkeit bei F2 liegt jedoch bei Ort 156 (Königsbrunn), der sich durch einen relativ hohen Faktorwert deutlich von seiner Umgebung abhebt. Hier scheinen Varianten aufzutreten, die für die Region relativ untypisch sind und die eher mit dem Nordosten des Untersuchungsgebiets assoziiert sind. Wie ist dies zu erklären? Königsbrunn ist eine relativ junge Ortsgründung. In der seit Anfang des 19. Jahrhunderts besiedelten Kolonie ließen sich neben Zuzüglern aus dem Nachbarort Bobingen (154) und der Pfalz vor allem Leute aus dem Ries und dem Donaumoos nieder, also aus Kern- und Saumzonen von F2, die ihren Dialekt mitbrachten und in Königsbrunn mit dem Dialekt der Region vermengten.

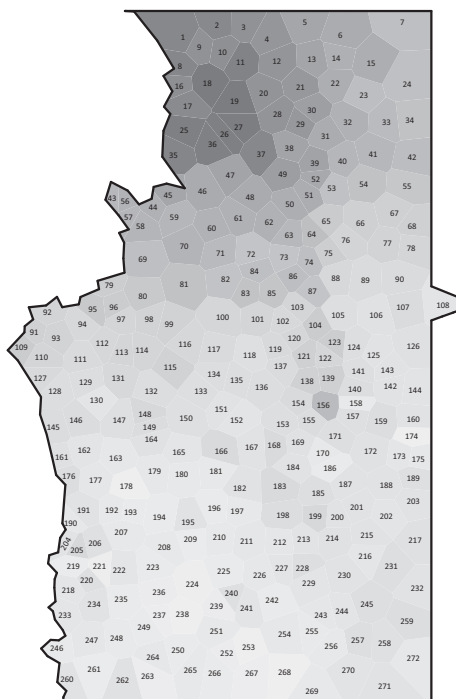


Abb. 6: Faktor 2

Hier handelt es sich also um eine Sprachinsel, eine kristalline Mikrostruktur im sprachgeografischen Kontinuum.

Abbildung 7 (siehe Farbabbildungsteil) zeigt mit F15 einen wesentlich „schwächeren“ Verdichtungsbereich als die vorhergehenden Faktoren. Hier sind nun einige Orte rötlich eingefärbt, was bedeutet, dass sie eine negative F15-Faktorladung haben: Diese Orte neigen für einen bestimmten Teil des Lexikons dazu, nicht dieselben Varianten aufzuweisen wie die Orte mit den positiven Faktorladungen. Das lokale Maximum von F15 liegt bei lediglich 11 Prozent (d. h. selbst dort, an Ort 193 [Daxberg], sind andere Faktoren stärker vertreten als F15), global beschreibt F15 nur 0,61 Prozent der Gesamtvariation. Derart nachrangige Faktoren werden bei der Faktorenanalyse meist als nicht aussagekräftig aussortiert; nur die stärksten Faktoren werden interpretiert. Dennoch sind hier zwei Umstände zu berücksichtigen: 1. Den 0,61 Prozent der Gesamtvariation, die F15 beschreibt, entspricht immerhin das Äquivalent von ca. 75 Varianten-Einzelverteilungen (= 12.341 Variantenverteilungen  $\times$  0,61 Prozent); 2. der Verdichtungsbereich zeigt eine geografisch kohärente Verteilung, was auf das Vorhandensein eines zugrundeliegenden geolinguistischen Mechanismus hinweist, denn eine solche Verteilung macht statistische Fluktuationen in den Daten („Rauschen“) als Ursache unwahrscheinlich (vgl. WARTENBERG 1985, PICKL 2013, 160–161). Es besteht

also Grund zu der Annahme, dass mit F15 tatsächlich ein aussagekräftiger Faktor vorliegt, der bestehende Tendenzen in der Wortgeografie zusammenfasst.

Im konkreten Fall liegt ein Bezug zu Memmingen (205) als Markt- und Einkaufsort nahe. Band 1 des SBS enthält die beiden nicht-sprachlichen Karten zu den Themen „Wohin ging man früher auf den Markt?“ (Karte 12) und „Wohin fährt man heute zum Einkaufen?“ (Karte 13). Die jeweiligen Einzugsgebiete von Memmingen sind in Abbildung 8 dargestellt. Der Zusammenhang zu F15 ist unmittelbar erkennbar, so dass hier davon ausgegangen werden kann, dass Memmingen für bestimmte Varianten eine ausgleichende Wirkung innerhalb seines Einzugsgebiets entfaltet hat. Unter den 20 Faktoren finden sich weitere Beispiele, die einen Zusammenhang mit Marktorten nahelegen (unter anderen Mindelheim, Weilheim, Lauingen), was diese These zusätzlich stützt.

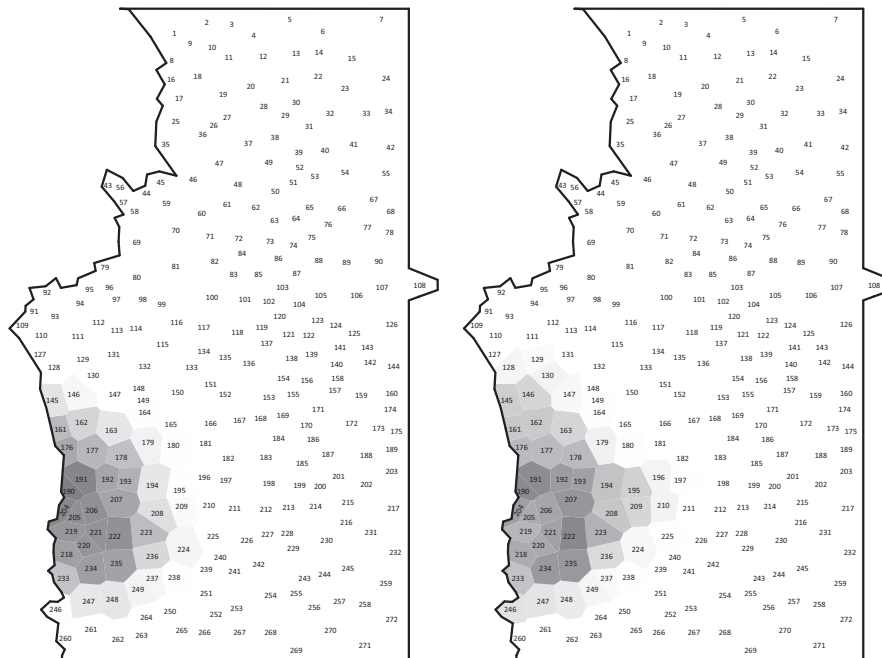


Abb. 8a: Antwort „Memmingen“ (Bd. 1, K. 12, „Marktorte [früher]“)      Abb. 8b: Antwort „Memmingen“ (Bd. 1, K. 13, „Einkaufsorte [heute]“)

Abb. 8: Einzugsgebiet der Stadt Memmingen als früherer Marktort und heutiger Einkaufsort. Die Verteilungen wurden wiederum mittels sogenannter Intensitätsschätzung für die Darstellung optimiert (vgl. Abbildung 3)

Als letztes Beispiel sei kurz F13 vorgestellt (Abbildung 9, im Farbabbildungsteil), der eine ungewöhnliche geografische Verteilung zeigt. Er gehört mit global 0,63 Prozent und lokal maximal 13 Prozent (am Ort 122: Augsburg) an beschriebener Variation ebenfalls zu den schwächeren Faktoren. Er zeigt insgesamt keine geografisch kohärente Verteilung; stattdessen sind Orte mit höheren Faktorladungen

ungleichmäßig über das gesamte Untersuchungsgebiet verteilt. Lediglich um das Maximum, Ort 122 (Augsburg), zeigt sich eine auffällige Ballung. Die 20 Orte mit den höchsten Ladungen sind:

Augsburg (122), Lechhausen (123), Stadtbergen (121), Günzburg (96), Gersthofen (104), Kaufbeuren (240), Memmingen (205), Haunstetten (139), Inningen (138), Weilheim (259), Landsberg (199), Königsbrunn (156), Ottmarshausen (120), Kaisheim (38), Neu-Ulm (109), Wallerstein (16), Oettingen (10), Nördlingen (17), Herrsching (217) und Kempten (262)

Dies sind mit Ausnahme von Kaisheim, Wallerstein und Herrsching alles Städte oder Stadtteile. Es scheint sich hier also um einen Faktor von besonders „urbaner“ Ausprägung zu handeln. Eine Regressionsanalyse zeigt einen starken Zusammenhang zwischen den Faktorladungen von F13 und den Einwohnerzahlen der einzelnen Orte (siehe Tabelle 3).<sup>16</sup> Es scheint sich dabei eher um einen logarithmischen denn um einen linearen Zusammenhang zu handeln; zudem ist er bei den aktuellsten vorliegenden Einwohnerzahlen (vor der Datenerhebung für den SBS) am ausgeprägtesten. Dies könnte bedeuten, dass hier selbst jüngere Bevölkerungsentwicklungen noch merklichen Einfluss bei der in den Jahren 1984 bis 1989 erfolgten Befragung hatten.

Wie sieht nun ein solcher geografisch diskontinuierlicher, urbaner Verdichtungsbereich sprachlich aus? Die Vermutung liegt nahe, dass es in den Dialekten der Städte eine Tendenz zu standardnäherer Lexik gibt. Und in der Tat sind die Varianten mit den 20 höchsten Faktorwerten für F13 (Tabelle 4) allesamt die jeweilige (oder eine) Standardvariante.

Tab. 3: Statistische Zusammenhänge zwischen den Einwohnerzahlen der 272 Untersuchungsorte in verschiedenen Jahren und den Faktorladungen von F13.  $R^2$  ist der Anteil der durch den linearen bzw. logarithmischen Zusammenhang erklärbaren Varianz

	$R^2$	
	linearer Zusammenhang	logarithmischer Zusammenhang
Einwohnerzahlen 1871	0,20	0,27
Einwohnerzahlen 1925	0,14	0,28
Einwohnerzahlen 1939	0,16	0,33
Einwohnerzahlen 1951	0,19	0,35
Einwohnerzahlen 1971	0,23	0,36

<sup>16</sup> Die Einwohnerzahlen der Untersuchungsorte für die einzelnen Stichjahre stammen neben Band 1 des SBS aus folgenden Quellen: BAYERISCHES STATISTISCHES LANDESAMT 1953; 1972; BAYERISCHES LANDESAMT FÜR STATISTIK UND DATENVERARBEITUNG 1991.



Tab. 4: Die Varianten mit den 20 höchsten Faktorwerten für F13

FW	Variante	Variable	Karte
10,29	<i>Spinne</i>	‘Spinne’	8.60
10,18	<i>Henkel</i>	‘Griff am Kochtopf’	10.63
9,81	<i>Gießkanne</i>	‘Gießkanne’	8.103
9,75	<i>Mittwoch</i>	‘Mittwoch’	10.191
9,66	<i>Schuppen</i>	‘Kopfschuppen’	2.8
9,32	<i>Wespe</i>	‘Wespe’	8.58
9,25	<i>Dienstag</i>	‘Dienstag’	10.190
9,14	<i>Braut</i>	‘Braut’	2.116
8,58	<i>Lehrer</i>	‘Schullehrer’	2.111
8,36	<i>Schirm</i>	‘Regenschirm’	10.46
8,32	<i>horch</i>	‘horch!’	2.67
8,26	<i>Viertel nach sechse</i>	‘6 Uhr 15’	10.204
8,17	<i>Westwind</i>	‘Westwind’	8.41
7,98	<i>Kaffee trinken</i>	‘Frühstück’	10.69
7,91	<i>auslöschen</i>	‘Kerze löschen’	10.23
7,90	<i>Sarg</i>	‘Sarg’	2.123
7,89	<i>Bräutigam</i>	‘Bräutigam’	2.117
7,78	<i>(zu) Abend essen</i> <i>(das) Abendessen /</i>	‘Nachtessen’	10.74
7,70	<i>Schaum</i>	‘Schaum (z. B. auf dem Bier)’	10.163
7,70	<i>Großmutter</i>	‘Großmutter’	2.100

#### 4.2. Synopse

Die individuellen Faktoren bilden die jeweiligen Verdichtungsbereiche im wortgeografischen Kontinuum – und damit (lexikalische) Dialekttypen beziehungsweise diatopische Varietäten – einzeln ab. Im Sinne der Übersichtlichkeit kann man alle Faktoren in einer einzigen kartografischen Darstellung zusammenfassen. Dabei werden zwangsläufig bestimmte Informationen ausgeblendet. Konkret wird an jedem Ort derjenige Faktor verzeichnet, der dort den größten Anteil an lokaler Variation beschreibt (siehe Abbildung 10 im Farbabbildungsteil).<sup>17</sup>

Die unterschiedlichen Farbschattierungen geben an, wie stark ein Ort durch den dort dominanten Faktor geprägt ist. Hier ist zu beachten, dass die braun markierten

<sup>17</sup> Einer der 20 Faktoren (F9) wurde dabei ignoriert, da er lediglich Informationen über die Zahl an fehlenden Belegen pro Ort abbildet, also keine linguistische Information enthält (vgl. PICKL 2013, 180–182); ein Effekt, der bei der Faktoren- und Hauptkomponentenanalyse häufig auftritt (vgl. unter anderen HYVÖNEN / LEINO / SALMENKIVI 2007, 275–276).

Linien zwischen den Gebieten keine scharfen Grenzen darstellen, sondern die Zentren von Übergangsbereichen, in denen ein Faktor einflussreicher wird als der andere.<sup>18</sup> Durch diese Linien wird aber auch illustriert, dass die Faktorenanalyse bei Bedarf auch als Instrument zur Konstruktion von „scharfen“ Dialektarealen geeignet ist (den Geltungsbereichen von Dialekten, nicht von Varietäten; vgl. die Unterscheidung in Abschnitt 2). Als solches ist die Faktorenanalyse besser geeignet als etwa die Clusteranalyse, da sie wesentlich robuster (d. h. unempfindlicher gegenüber kleinen Schwankungen in den Daten) ist.<sup>19</sup> Der eigentliche Vorteil der Dialekteinteilung mithilfe der Faktorenanalyse besteht aber gerade darin, der Tatsache Rechnung zu tragen, dass die Zuordnung von Ortsdialekten zu Dialektgebieten meist nicht eindeutig entscheidbar ist.

Es ist in Abbildung 10 jedoch nicht zu sehen, welche Faktoren unter der Oberfläche zusätzlich zu den dominanten für die an einem Ort verwendeten Varianten verantwortlich sind. Faktoren wie F15, die nirgends „stärkster“ Faktor sind, sind in der Gesamtschau überhaupt nicht zu sehen. Um Einblick in die faktorielle Zusammensetzung eines Ortsdialekts zu bekommen, kann man deshalb für einzelne Orte sogenannte Ortsprofile erstellen, die sich etwa in Form eines Tortendiagramms visualisieren lassen. Abbildung 11 (im Farbbildungsteil) zeigt ein solches Ortsprofil für Türkheim (197).

Auf den ersten Blick wird klar, dass Türkheim zu ungefähr gleichen Teilen durch den Allgäuer und den mittelostschwäbischen Dialekttyp geprägt ist. Elf Prozent der Varianten gehen auf die Wirkung des naheliegenden Markt- beziehungsweise Einkaufsortes Mindelheim zurück; zu vier Prozent weist Türkheim Varianten auf, die eigentlich eher für die andere Lechseite charakteristisch sind, und zu drei Prozent orientiert sich Türkheim lexikalisch an den Stauden (ein Gebiet westlich von Augsburg). Gut ein Drittel der Varianten ist nicht durch Tendenzen erklärbar („n. e.“), die die Faktorenanalyse identifizieren konnte. Dies bedeutet nicht, dass 37 Prozent der Varianten nur in Türkheim vorkommen, sondern nur, dass diese Varianten in einer Weise im Raum verteilt sind, die idiosynkratisch oder regellos ist und deshalb von der Faktorenanalyse nicht erfasst werden konnte.

Die Informationen über das Zusammenspiel der Dialekttypen, die durch die Darstellungsweise in Abbildung 10 verborgen bleiben, lassen sich durch die Ortsprofil-Darstellung detailgenau, aber jeweils nur punktweise nachvollziehen.

<sup>18</sup> „The boundary between two dialect areas would then be defined as the point where the [influence] functions take on equal values. Given two influence functions, it is a straightforward task to construct a corresponding membership function where the break-point corresponds to a value of 0.5 for the membership function.“ (GIRARD / LARMOUTH 1993, 112–113)

<sup>19</sup> „Recent research has shown that cluster analysis should be applied with caution to dialect data [NERBONNE et al. 2008; PROKIĆ / NERBONNE 2008]. Small differences in the input data can lead to substantially different clustering results. Because of this, the results of different clustering algorithms should be compared and the results should be carefully evaluated. In data that is truly continuous clustering algorithms are unlikely to find meaningful clusters.“ (LEINONEN 2010, 37)

Diese Methode lässt sich weiter ausbauen, indem man statt einzelner Orte einen kompletten Querschnitt durch das Untersuchungsgebiet entlang eines Pfades vornimmt und die dort vertretenen Faktoranteile grafisch umsetzt, zum Beispiel von Ort 176 (Kellmünz) im Südwesten bis Ort 126 (Tödtenried) im Osten (siehe Abbildung 12 im Farbabbildungsteil). Dies erlaubt eine (pseudo-)kontinuierliche Darstellung (ähnlich der schematischen Darstellung in Abbildung 1), da die betrachteten Erhebungsorte in der Regel benachbart sind (siehe Abbildung 12 im Farbabbildungsteil).

In dieser Darstellungsweise wird ein graduelles, kontinuierliches Wechselspiel der Dialekttypen suggeriert. Dargestellt sind dennoch nur die Werte von 16 diskreten Orten. Hier wird noch einmal klar, dass die Faktoren (meist) räumlich kontingente Zusammenfassungen von im Raum verdichteten Varianten darstellen. Im konkreten Fall lässt sich das graduelle Überlappen der „großen“ Dialektgebiete Allgäuerisch, Mittellostschwäbisch, Lechrainisch und Mittelbairisch beobachten, wobei ein Bruch auf Höhe des Lechs (zwischen Haunstetten [139] und Ottmaringen [141]) festzustellen ist. Außerdem wird noch einmal der relativ geringe, aber belegbare Anteil des angenommenen Einflusses der Markt- und Einkaufsorte deutlich: An den Orten 176, 177 und 163 beziehungsweise 165, 150 und 151 spielen Memmingen beziehungsweise Mindelheim eine deutlich feststellbare, wenngleich vergleichsweise kleine Rolle in der Zusammensetzung der lokalen Dialekte. Zudem wird mit den Orten 154, 138 und 139 das urbane Umfeld Augsburgs gestreift, was sich in einer Zunahme von F13 an diesen Orten manifestiert. Auch hier ist der Anteil zwar gering, aber eindeutig.

Die „Querschnitt“-Darstellung der Faktoren zeigt die Grundstruktur des in Abbildung 1 vorgestellten Dialektschichtenmodells: Es wird deutlich, wie manche Faktoren beziehungsweise Varietäten dominante Kerngebiete ausbilden, andere im Hintergrund und damit nicht-dominant bleiben. Dennoch sind auch sie aufgrund ihrer spezifischen räumlichen Verbreitungscharakteristik als selbständige Dialekttypen identifizierbar. Einzelne Ortsdialekte lassen sich in Abhängigkeit der Dialekttypen beschreiben; manche (wie etwa Ort 163) sind stark durch einen bestimmten Typ charakterisiert; andere (wie etwa Ort 139) lassen sich keinem einzelnen Typ zuordnen, sondern stehen zwischen verschiedenen Varietäten.

Im Vergleich mit herkömmlichen Dialekteinteilungsverfahren – seien sie mittels struktureller oder quantitativ-aggregativer Methoden erzielt – sind zunächst die Faktorladungen, die die Zugehörigkeit eines Ortsdialekts zu einem Dialekttyp quantifizieren, als Alleinstellungsmerkmal der Faktorenanalyse zu nennen. In diesem Punkt vermag sie wesentlich detailliertere (und nebenbei auch robustere) Ergebnisse zu liefern als die reguläre Clusteranalyse, die festgefügte Gebiete liefert. Auch im Vergleich mit dem statistisch stabilisierten *fuzzy clustering*, das auch Übergangsbereiche abzubilden in der Lage ist, vermag die Faktorenanalyse feinere sprachgeografische Strukturen zu erfassen; hiermit ist in erster Linie die Identifikation von nicht-dominanten Strukturen gemeint, die die aggregativen

Verfahren durchweg „übersehen“. Dies ist auch bei der multidimensionalen Skalierung (vgl. Abschnitt 1) der Fall, die die Kontinuität sprachgeografischer Verhältnisse abbildet, aber keinerlei Muster unterhalb der Dominanzschwelle sichtbar machen kann. Ein weiterer Punkt, der die Faktorenanalyse von aggregativen Verfahren abhebt, ist der der Beziehbarkeit der Resultate auf die Einzelvariablen: Mit den Faktorwerten lässt sich für jeden Faktor angeben, welche Varianten in welchem Ausmaß mit ihm assoziiert sind. Mithilfe von Faktorwerten und -ladungen lassen sich sogar die Einzelverteilungen der Varianten aus den Gesamtergebnissen rekonstruieren (siehe PICKL 2013, 200–203).

Im Hinblick auf die konkreten Ergebnisse ist eine unmittelbare Vergleichbarkeit der Resultate dieser Studie mit bisherigen Einteilungen nicht gegeben, da solche im Normalfall nicht auf wortgeografischem Material beruhen und man so Aussagen über unterschiedliche sprachliche Ebenen gegenüberstellen würde. Dennoch sei hier angemerkt, dass die ermittelten dominanten wortgeografischen Strukturen im Wesentlichen zu bekannten Dialekteinteilungen passen, so zum Beispiel mit der dialektalen Einteilung des Raums Bayerisch-Schwaben durch EDUARD NÜBLING (1988, 118–119),<sup>20</sup> die auf einer Reihe ausgewählter lautlicher Phänomene beruht. NÜBLINGS Dialektregionen Nordostschwäbisch, Ostmittelschwäbisch, Allgäuisch, Nordbairisch, Mittelbairisch und Lechrainisch scheinen in ihrer prinzipiellen Gestalt auch auf der Faktorenkarte auf, wenngleich die genaue Lage der Grenzen (abgesehen von der Grenze am Lech) variiert, was abermals den Übergangscharakter der meisten Grenzen illustriert. Das nördlich-zentrale, türkis eingefärbte Gebiet der Faktorenanalyse, das mit dem sogenannten Stauden-Gebiet zusammenfällt, ist dagegen in bisherigen Einteilungsversuchen nicht aufgeschienen und stellt sozusagen einen „Neufund“ dar. Das Lechrainische wertet NÜBLING (1988, 113) als Übergangsmundart, da es keine eigenständigen lautlichen Merkmale aufweist; im Bereich der Lexik ist das nach Ausweis der Faktorenanalyse offenbar durchaus der Fall, es handelt sich demnach um einen eigenen Dialekttyp. Das Ostfränkische, das bei NÜBLING in einem kleinen Bereich im äußersten Norden in Bayerisch-Schwaben hineinragt, wird von der Faktorenanalyse ebenfalls identifiziert (als Faktor 19; siehe Abbildung 13 im Farbbildungsteil). Dieser Faktor ist aber sehr schwach und bildet lediglich einen nicht-dominanten Dialekttyp ab; in dieser Region stellen das Nordostschwäbische und das Nordbairische die prominenteren Größen dar.

Durch die farblichen Abstufungen in der Faktorenkarte werden die Übergangszonen unmittelbar sichtbar, und mit Orts- oder Querschnittprofilen ist auch ein Blick „unter die Oberfläche“ möglich, der das Zusammenspiel verschiedener abstrakter dialektaler Typen und ihrer Realisationen in konkreten Dialekten veranschaulicht.

<sup>20</sup> Ähnlich dazu beziehungsweise darauf aufbauend SBS (Bd. 1, K. 8); KÖNIG (2010a; 2010b).

## 5. Fazit

Verdichtungen im sprachgeografischen Kontinuum, das heißt Konglomerate gemeinsam auftretender (= kookkurrierender) Varianten, lassen sich mit GAETANO BERRUTO (1987; 2010) als diatopische Varietäten interpretieren. Ein Verdichtungsbereich betrifft deshalb immer nur einen Teil des Sprachinventars; es gehört somit zum Wesen von diatopischen Varietäten, sich räumlich zu überlappen und sich lokal gegenseitig zu vollen Inventaren zu ergänzen. Da sie innersprachlich definiert sind, lassen sie sich als unscharf ausgedehnte räumliche Formationen beschreiben, die in einzelnen Ortsdialekten in unterschiedlichem Ausmaß vertreten sind. Die scheinbare Diskrepanz zwischen Dialektkontinuum und Dialektgebieten ist damit hinfällig. Aus prototypentheoretischer Perspektive (vgl. auch CHRISTEN 1998; 2010; BERTHELE 2006, ANDERS 2010) lassen sich solche Verdichtungen als Dialekttypen beschreiben, die in einzelnen Dialekten (z. B. Ortsdialekten) mit unterschiedlichen Typikalitätsgraden realisiert sind, und für die einzelne Varianten (Merkmale) unterschiedlich charakteristisch sein können.

Die Faktorenanalyse ist ein geeignetes Instrument, um solche Verdichtungsbereiche, das heißt Bündel von gemeinsam auftretenden Varianten, in der sprachgeografischen Gesamtvariation zu isolieren, wenn diese in Form eines digital aufbereiteten Kartenkorpus vorliegt. Bei der händischen Auseinandersetzung mit einem Dialektatlas könnten solche Strukturen meist nicht erfasst werden. Der Grad der Merkmalshaftigkeit einzelner Varianten wird dabei durch die sogenannten Faktorwerte quantifiziert, der Grad der Typikalität einzelner Ortsdialekte durch die sogenannten Faktorladungen. Damit ist nicht nur einer differenzierteren Dialekteinteilung, als sie mit etablierten Methoden möglich ist, Vorschub geleistet; es können außerdem latente Strukturen identifiziert werden, die unter Umständen nur wenige Varianten betreffen, die aber dennoch Aufschluss über sprachgeografische Dynamik liefern. Gleichzeitig bleibt auch in großen Kartenkorpora der Blick auf die Einzelvariante unverstellt. Dies wurde am Beispiel der wortgeografischen Variation in Bayerisch-Schwaben demonstriert.

Konkret wurden aus dem lexikalischen Teil des Sprachatlas von Bayerisch-Schwaben 20 Faktoren isoliert, die jeweils einen Teil der wortgeografischen Variation beschreiben und kollektive Variantenkookkurrenzen abbilden. Dabei konnte nicht nur gezeigt werden, dass schon auf der Grundlage lexikalischer Daten Dialekträume differenzierter sichtbar werden, als dies in bisherigen Dialekteinteilungen der Fall war, sondern auch, dass dadurch Strukturen zutage treten, die mit bisheriger Methodik nicht erfasst werden konnten. Dies unterstreicht nicht nur ganz allgemein den Wert statistischer Verfahren für die Dialektologie, die von der Dialektometrie schon beträchtlich profitiert hat, sondern auch den Wert speziell dieses Verfahrens innerhalb der Dialektometrie. Nur durch einen *data-driven*-Ansatz, wie er hier vorliegt, ist es möglich, unvermutete Strukturen zu entdecken; und erst durch ein varianten- und ortsbasiertes Verfahren wie der Faktorenanalyse (im Gegensatz etwa zur Clusteranalyse, die nur ortsbasiert ar-

beitet) können auch Strukturen ermittelt werden, von denen nur ein kleiner Teil des Korpus betroffen ist.

## LITERATUR

- ANDERS, CHRISTINA ADA (2010): Wahrnehmungsdiakologie. Das Obersächsische im Alltagsverständnis von Laien. Berlin/New York: De Gruyter (Linguistik – Impulse & Tendenzen. 36).
- AUER, PETER (1986): Konversationelle Standard/Dialekt-Kontinua (Code-Shifting). In: Deutsche Sprache 14, 97–124.
- BACH, ADOLF (1950): Deutsche Mundartforschung. Ihre Wege, Ergebnisse und Aufgaben. Mit 58 Karten. Zweite Auflage. Heidelberg: Winter (Germanische Bibliothek. Dritte Reihe: Untersuchungen und Einzeldarstellungen).
- BAYERISCHES LANDESAMT FÜR STATISTIK UND DATENVERARBEITUNG (Hg.) (1991): Die Gemeinden Bayerns nach dem Gebietsstand 25. Mai 1987. Die Einwohnerzahlen der Gemeinden Bayerns und die Änderungen im Bestand und Gebiet von 1840 bis 1987. München (Beiträge zur Statistik Bayerns. 451).
- BAYERISCHES STATISTISCHES LANDESAMT (Hg.) (1953): Historisches Gemeindeverzeichnis. Die Einwohnerzahlen der Gemeinden Bayerns in der Zeit von 1840 bis 1952. München (Beiträge zur Statistik Bayerns. 192).
- BAYERISCHES STATISTISCHES LANDESAMT (Hg.) (1972): Einwohnerzahlen am 31. Dezember 1971. Jährliches Ergänzungsheft zum Amtlichen Gemeindeverzeichnis für Bayern. München (Beiträge zur Statistik Bayerns. 312).
- BERRUTO, GAETANO (1987): Sociolinguistica dell'italiano contemporaneo. Rom: La Nuova Italia Scientifica.
- BERRUTO, GAETANO (2004): Sprachvarietät – Sprache (Gesamtprache, historische Sprache). In: AMMON, ULRICH / NORBERT DITTMAR / KLAUS J. MATTHEIER / PETER TRUDGILL (Hg.): Soziolinguistik. Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft. 2. Aufl. Berlin/New York: de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft. 3.1), 188–195.
- BERRUTO, GAETANO (2010): Identifying dimensions of linguistic variation in a language space. In: AUER, PETER / JÜRGEN ERICH SCHMIDT (Hg.): Language and Space. An International Handbook of Linguistic Variation. Volume 1: Theories and Methods. Berlin/New York: De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft. 30.1), 226–241.
- BERTHELE, RAPHAEL (2006): Wie sieht das Berndeutsche so ungefähr aus? Über den Nutzen von Visualisierungen für die kognitive Laienlinguistik. In: KLAUSMANN, HUBERT (Hg.) (2006): Raumstrukturen im Alemannischen. Beiträge der 15. Arbeitstagung zur alemannischen Dialektologie. Schloss Hofen, Lochau (Vorarlberg) vom 19.–21.9.2005. Graz: Neugebauer (Schriften der Vorarlberger Landesbibliothek. 15), 163–175.
- BIBER, DOUGLAS (1991): Variation across speech and writing. Cambridge/New York/Melbourne: Cambridge University Press.
- BUSSMANN, HADUMOD (Hg.) (2008): Lexikon der Sprachwissenschaft. Stuttgart: Kröner.
- CHAMBERS, J. K. / PETER TRUDGILL (1998). Dialectology. Cambridge: Cambridge University Press (Cambridge Textbooks in Linguistics).
- CHRISTEN, HELEN (1998): Dialekt im Alltag. Eine empirische Untersuchung zur lokalen Komponente heutiger schweizerdeutscher Varietäten. Tübingen: Niemeyer (Germanistische Linguistik. 201).
- CHRISTEN, HELEN (2010): Was Dialektbezeichnungen und Dialektattributionen über alltagsweltliche Konzeptualisierungen sprachlicher Heterogenität verraten. In: ANDERS, CHRISTINA ADA / MARKUS HUNDT / ALEXANDER LASCH (Hg.): Perceptual Dialectology. Neue Wege der Dialektologie. Berlin/New York: De Gruyter (Linguistik – Impulse & Tendenzen. 38), 269–290.
- CLOPPER, CYNTHIA G. / JOHN C. PAOLILLO (2006): North American English Vowels: A Factor-analytic Perspective. In: Literary and Linguistic Computing 21/4, 445–462.
- CRÄMER, ULRICH (1954): Das Allgäu. Werden und Wesen eines Landschaftsbegriffs. Mit 18 Karten. Remagen: Bundesanstalt für Landeskunde (Forschungen zur deutschen Landeskunde. 84).

- FISCHER, HERMANN (1895): Geographie der Schwäbischen Mundart. Mit einem Atlas von achtundzwanzig Karten. Tübingen: Laupp.
- GIRARD, DENNIS / DONALD LARMOUTH (1993): Some Applications of Mathematical and Statistical Models in Dialect Geography. In: PRESTON, DENNIS R. (Hg.): *American Dialect Research*. Amsterdam/Philadelphia: Benjamins, 107–131.
- GOEBL, HANS (1983): ‚Stammbaum‘ und ‚Welle‘. Vergleichende Betrachtungen aus numerisch-taxonomischer Sicht. In: *Zeitschrift für Sprachwissenschaft* 2/1, 3–44.
- GOOSSENS, JAN (1969): Strukturelle Sprachgeographie. Eine Einführung in Methodik und Ergebnisse. Mit 30 Karten. Heidelberg: Winter (Sprachwissenschaftliche Studienbücher. Zweite Abteilung).
- GRIEVE, JACK (2009): A corpus-based regional dialect survey of grammatical variation in written Standard American English. Dissertation. Flagstaff: Northern Arizona University.
- GRIEVE, JACK / DIRK SPEELMAN / DIRK GEERAERTS (2011): A statistical method for the identification and aggregation of regional linguistic variation. In: *Language Variation and Change* 23, 193–221.
- HEERINGA, WILBERT / JOHN NERBONNE (2001): Dialect Areas and Dialect Continua. In: *Language Variation and Change* 13, 375–400.
- HYVÖNEN, SAARA / ANTTI LEINO / MARKO SALMENKIVI (2007): Multivariate Analysis of Finnish Dialect Data – An Overview of Lexical Variation. In: *Literary and Linguistic Computing* 22/3, 271–290.
- KÖNIG, WERNER (1976): Der Landschaftsname Allgäu. Zur Abhängigkeit seines Bedeutungsumfanges von regionalen, soziologischen und psychologischen Faktoren. In: *Alemannica. Landeskundliche Beiträge*. Festschrift für Bruno Boesch. Bühl (Baden): Konkordia (Alemannisches Jahrbuch. 1973/75), 186–200.
- KÖNIG, WERNER (2010a): Alemannisch-Schwäbische Dialekte in Bayern. In: *Historisches Lexikon Bayerns*. URL: <[http://www.historisches-lexikon-bayerns.de/artikel/artikel\\_45732](http://www.historisches-lexikon-bayerns.de/artikel/artikel_45732)>; Stand 25.01.2013.
- KÖNIG, WERNER (2010b): Großmundarten und Dialektgrenzen. In: FREI, HANS / PANKRAZ FRIED / ROLF KIESSLING (Hg.) (1982–): *Historischer Atlas von Bayerisch-Schwaben*. 2. Auflage. Augsburg: Schwäbische Forschungsgemeinschaft. 5. Lieferung, XIII,1.
- KRISTIANSEN, GITTE (2008): Style-shifting and shifting styles: A socio-cognitive approach to lectal variation. In: KRISTIANSEN, GITTE / RENÉ DIRVEN (Hg.): *Cognitive Sociolinguistics. Language Variation, Cultural Models, Social Systems*. Berlin/New York: De Gruyter Mouton (Cognitive Linguistics Research. 39), 45–88.
- LABOV, WILLIAM / SHARON ASH / CHARLES BOBERG (2006): *The Atlas of North American English. Phonetics, Phonology and Sound Change. A Multimedia Reference Tool*. Berlin: De Gruyter Mouton.
- LAKOFF, GEORGE (1987): *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- LEINO, ANTTI / SAARA HYVÖNEN (2009): Comparison of Component Models in Analysing the Distribution of Dialect Features. In: NERBONNE, JOHN / CHARLOTTE GOOSKENS / SEBASTIAN KÜRSCHNER / RENÉE VAN BEZOOIJEN (Hg.) (2009): *Computing and Language Variation*. Edinburgh: Edinburgh University Press (International Journal of Humanities and Arts Computing 2/1–2: Special Issue), 173–187.
- LEINONEN, THERESE (2009): Factor Analysis of Vowel Pronunciation in Swedish Dialects. In: NERBONNE, JOHN / CHARLOTTE GOOSKENS / SEBASTIAN KÜRSCHNER / RENÉE VAN BEZOOIJEN (Hg.) (2009): *Computing and Language Variation*. Edinburgh: Edinburgh University Press (International Journal of Humanities and Arts Computing 2/1–2: Special Issue), 189–204.
- LEINONEN, THERESE (2010): An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects. Groningen: Rijksuniversiteit Groningen (Groningen Dissertations in Linguistics. 83).
- LENZ, ALEXANDRA N. (2003): *Struktur und Dynamik des Substandards. Eine Studie zum Westmitteldeutschen (Wittlich/Eiffel)*. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik. Beihefte. 125).
- LÖFFLER, HEINRICH (2003): *Dialektologie. Eine Einführung*. Tübingen: Narr.
- MÖLLER, ROBERT (2013): *Erscheinungsformen rheinischer Alltagssprache. Untersuchungen zu Variation und Kookurenzregularitäten im „mittleren Bereich“ zwischen Dialekt und Standardsprache*. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik. Beihefte. 153).

- NERBONNE, JOHN (2006): Identifying linguistic structure in aggregate comparison. In: *Literary and Linguistic Computing* 21/4, 463–475.
- NERBONNE, JOHN / PETER KLEIWEIG / FRANZ MANNI / WILBERT HEERINGA (2008): Projecting Dialect Distances to Geography: Bootstrap Clustering vs. Noisy Clustering. In: PREISACH, CHRISTINE / LARS SCHMIDT-THIEME / HANS BURKHARDT / REINHOLD DECKER (Hg.): *Data Analysis, Machine Learning, and Applications. Proceedings of the 31st Annual Meeting of the German Classification Society*. Berlin: Springer, 647–654.
- NERBONNE, JOHN / RINKE COLEN / CHARLOTTE GOOSKENS / PETER KLEIWEIG / THERESE LEINONEN (2011): Gabmap – A Web Application for Dialectology. In: *Dialectologia: Special Issue II*, 65–89.
- NIEBAUM, HERMANN / JÜRGEN MACHA (2006): *Einführung in die Dialektologie des Deutschen*. Tübingen: Niemeyer (Germanistische Arbeitshefte. 37).
- NÜBLING, EDUARD (1988): *Studien und Berichte zur Geschichts-, Mundart- und Namenforschung Bayerisch-Schwabens*. Festgabe zum 80. Geburtstag des Verfassers. Augsburg: Schwäbische Forschungsgemeinschaft. (Veröffentlichungen der Schwäbischen Forschungsgemeinschaft bei der Kommission für bayerische Landesgeschichte. Reihe 1: Studien zur Geschichte des bayerischen Schwabens. 16).
- PICKL, SIMON (2013): Probabilistische Geolinguistik. Geostatistische Analysen lexikalischer Variation in Bayerisch-Schwaben. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik. Beihefte. 154).
- PICKL, SIMON / JONAS RUMPF (2011): Automatische Strukturanalyse von Sprachkarten. Ein neues statistisches Verfahren. In: GLASER, ELVIRA / JÜRGEN ERICH SCHMIDT / NATASCHA FREY (Hg.): *Dynamik des Dialekts – Wandel und Variation*. Akten des 3. Kongresses der Internationalen Gesellschaft für Dialektologie des Deutschen (IGDD). Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik. Beihefte. 144), 267–285.
- PROKIĆ, JELENA / JOHN NERBONNE (2008): Recognizing groups among dialects. In: *International Journal of Humanities and Arts Computing* 2, 153–171.
- PRÖLL, SIMON / SIMON PICKL / AARON SPETTL (i.E.): Latente Strukturen in geolinguistischen Korpora. In: ELMENTALER, MICHAEL / MARKUS HUNDT / JÜRGEN ERICH SCHMIDT (Hg.): *Deutsche Dialekte. Konzepte, Probleme, Handlungsfelder*. Stuttgart: Steiner.
- PUTSCHKE, WOLFGANG (1993): Zur Kritik dialektologischer Einteilungskarten. In: VIERECK, WOLFGANG (Hg.): *Verhandlungen des Internationalen Dialektologenkongresses, Bamberg 29.7.–4.8.1990*. Plenarvorträge, computative Datenverarbeitung, Dialektgliederung und Dialektklassifikation. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik. Beihefte. 74), 421–443.
- ROELCKE, THORSTEN (2010): *Fachsprachen*. Berlin: Erich Schmidt (Grundlagen der Germanistik. 37).
- 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*. 14 Bände. Heidelberg: Winter.
- SCHMIDT, JÜRGEN ERICH / JOACHIM HERRGEN (2011): *Sprachdynamik. Eine Einführung in die moderne Regionalsprachenforschung*. Berlin: Erich Schmidt (Grundlagen der Germanistik. 49).
- SHACKLETON, ROBERT (2005): English-American Speech Relationships. A Quantitative Approach. In: *Journal of English Linguistics* 33/2, 99–160.
- SHACKLETON, ROBERT (2007): Phonetic Variation in the Traditional English Dialects: A Computational Analysis. In: *Journal of English Linguistics* 35/1, 30–102.
- STEINER, DIETER (1965). Die Faktorenanalyse – ein modernes statistisches Hilfsmittel des Geographen für die objektive Raumlagerung und Typenbildung. In: *Geographica Helvetica* 20/1, 20–34.
- STRECK, TOBIAS / PETER AUER (2012): Das raumbildende Signal in der Spontansprache. Dialektometrische Untersuchungen zum Alemannische in Deutschland. In: *Zeitschrift für Dialektologie und Linguistik* 79/2, 149–188.
- SZMRECSANYI, BENEDIKT / CHRISTOPH WOLK (2011): Holistic corpus-based dialectology. In: *Revista Brasileira de Linguística Aplicada* 11/2, 561–592.
- TABACHNICK, BARBARA G. / LINDA S. FIDELL (2007): *Using Multivariate Statistics*. Fifth Edition. Boston/München: Pearson.



- VALTS = GABRIEL, EUGEN (Hg.) (1985–2006): Vorarlberger Sprachatlas mit Einschluß des Fürstentums Liechtenstein, Westtirols und des Allgäus. 5 Bände. Bregenz: Vorarlberger Landesbibliothek.
- VEITH, WERNER H. (2005): Soziolinguistik. Ein Arbeitsbuch. 2. Auflage. Tübingen: Narr.
- VIERECK, WOLFGANG (1986): Dialectal Speech areas in England: Orton's lexical evidence. In: KASTOVSKY, DIETER / ALEKSANDER SZWEDEK (Hg.): Linguistics across Historical and Geographical Boundaries. In Honour of Jacek Fisiak on the Occasion of His Fiftieth Birthday. Volume 1: Linguistic Theory and Historical Linguistics. Berlin/New York/Amsterdam: De Gruyter Mouton (Trends in Linguistics. Studies and Monographs. 32), 725–740.
- VOGELBACHER, JULIUS (2011): Statistische Analyse von Wortvarianten des Sprachatlas von Bayerisch-Schwaben. Diplomarbeit in Mathematik. Universität Ulm.
- WARTENBERG, DANIEL (1985): Spatial Autocorrelation as a Criterion for Retaining Factors in Ordinations of Geographic Data. In: Mathematical Geology 17/7, 665–682.
- 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. Berlin/New York: De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft. 1.2), 807–900.
- ZADEH, LOTFI A. (1965): Fuzzy Sets. In: Information and Control 8, 338–353.

## SUMMARY

In dialectology, there are two competing conceptualizations of linguistic space: one based upon dialect areas/regions and the other on the idea of a dialect continuum. To date these two have been neither theoretically nor methodologically unified. In this article an attempt is made to furnish a theoretical basis from which both concepts can be derived, and a compatible method for the evaluation and interpretation of empirical data is introduced. GAETANO BERRUTO's suggestion that varieties be seen as concentrations of co-occurrent variants serves as a starting point; when applied to linguistic geography this implies that diatopic varieties can be conceived of as regions of concentration within a geolinguistic continuum. As such, they are fuzzy and can overlap one another. A prototype-theoretical dimension is added to the model, allowing varieties to be understood as abstract dialect types that are manifested to varying degrees in concrete individual dialects. To operationalize this concept, factor analysis, a statistical technique for dimensionality reduction is employed as a category-building instrument in linguistic geography; its mode of operation is eminently suited to identifying and quantifying regions of concentration, i. e., plausible structures within a geolinguistic corpus. At the same time, linguistic co-occurrences are employed to identify systematic relations between variants. Applying the method to lexical data from the *Sprachatlas von Bayerisch-Schwaben* reveals the usefulness of the procedure, not just for the categorization of linguistic space, but also in the search for previously unidentified spatial structures.

Adresse des Autors: Simon Pickl  
 Universität Salzburg  
 Fachbereich Germanistik  
 Erzabt-Klotz-Straße 1  
 A-5020 Salzburg  
 E-Mail: <simon.pickl@sbg.ac.at>

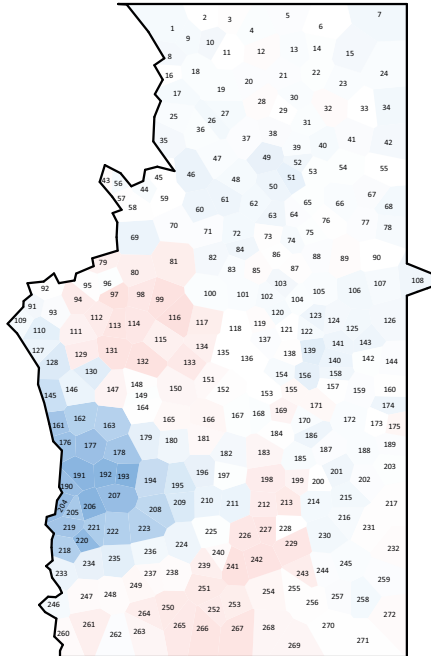


Abb. 7: Faktor 15

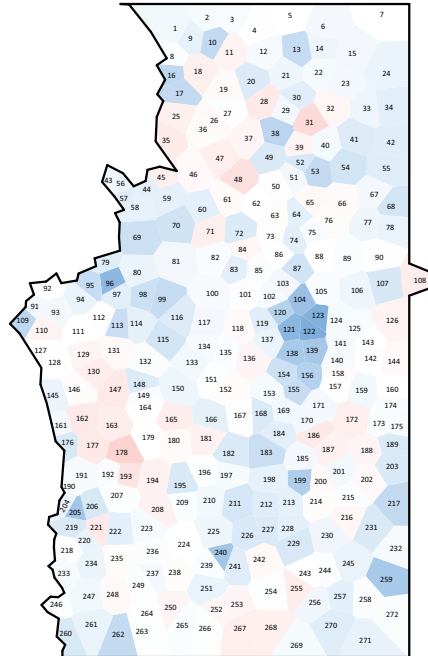


Abb. 9: Faktor 13

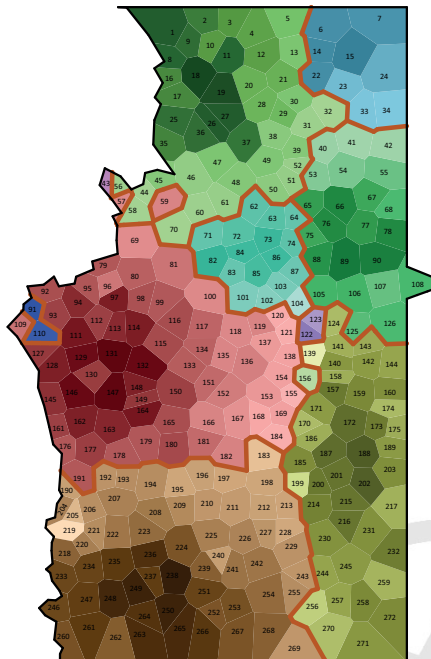


Abb. 10: Gradierte Faktoren-Flächenkarte

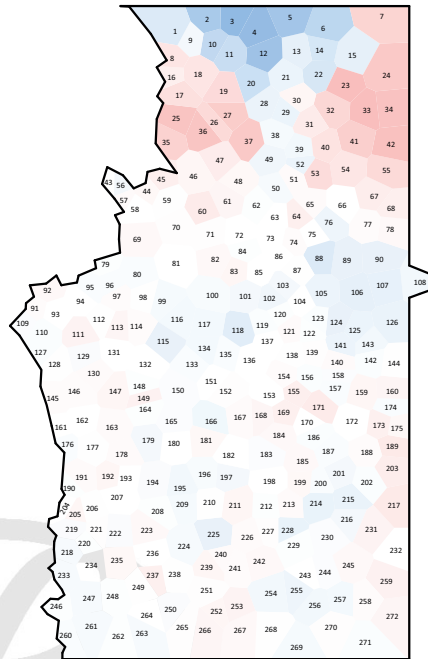


Abb. 13: Faktor 19

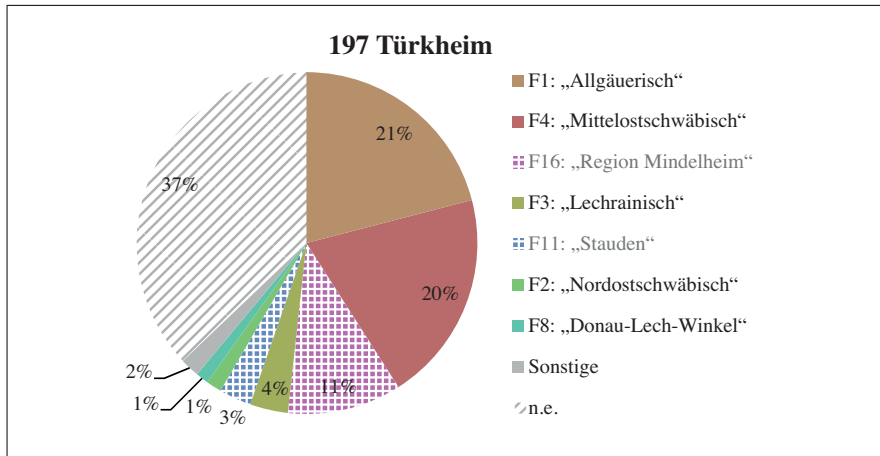


Abb. 11: Faktoren-Ortsprofil von Türkheim (197). Global nicht-dominante Faktoren sind gekachelt dargestellt. Die Faktoren sind in der hier angegebenen Reihenfolge im Uhrzeigersinn abgebildet; Faktoren unterhalb einer Schwelle von einem Prozent wurden zu „Sonstige“ zusammengefasst

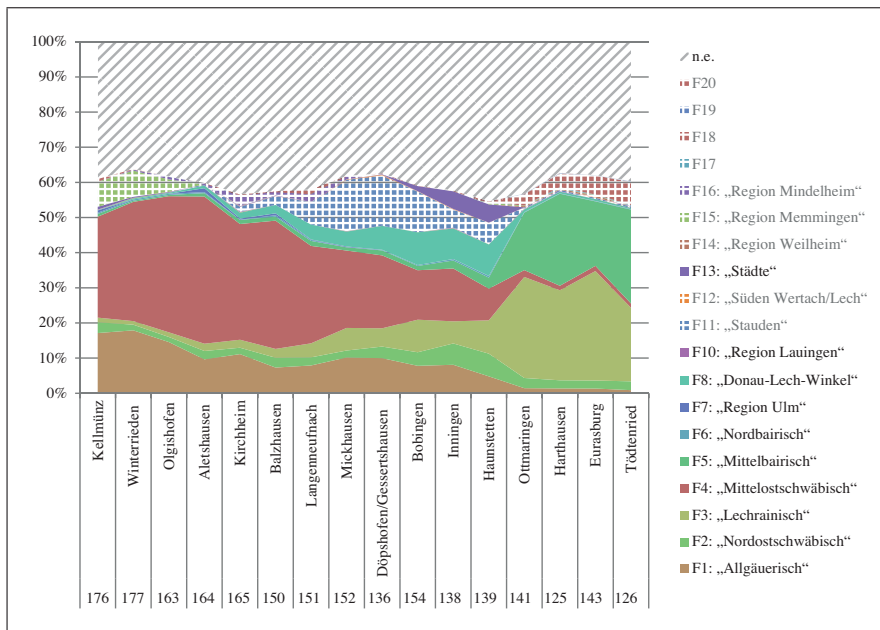


Abb. 12: Faktoren-Querschnittsprofil entlang einer geraden Linie von Kellmünz (176) bis Tödenried (126). Global nicht-dominante Faktoren sind gekachelt dargestellt