

John Nerbonne

Quantitative strukturelle Dialektologie:

Wie zwei Forschungsrichtungen zusammenpassen

1. Einleitung

Zu oft werden quantitative und strukturelle Einsätze in der Dialektologie – und übrigens auch in anderen sprachwissenschaftlichen Subdisziplinen – als ausschließliche, „entweder-oder“ Alternative statt als sich ergänzende wissenschaftliche Perspektiven betrachtet. Wenn beide Perspektiven legitim und sogar verständlich sind, ist es reizvoller, über ihr Verhältnis nachzudenken und über mögliche synthetisierende Herangehensweisen zu spekulieren, als lediglich die eine Perspektive gegen die andere auszuspielen zu versuchen. Weinreichs erste Aufruf zu einer strukturalistischen Dialektologie kann als Vorbild dafür dienen, wie sich verschiedene wissenschaftliche Forschungslinien bereichern lassen können, indem man sich bemüht, sie als gleichermaßen einsichtsreiche Ausprägungen eines einzelnen Phänomens – der Sprache – zu verstehen.

Ziel dieses Aufsatzes ist es, das heutige Verhältnis zwischen strukturalistischer und quantitativer Dialektologie zu skizzieren und Anregungen zu einer tiefer gehenden Synthese zu geben..

2. Strukturelle Dialektologie

2.1 Die Anfänge

Mit seinem zurecht berühmten Aufsatz „Is a structural dialectology possible?“ stieß Weinreich (1954) die Entwicklung der strukturellen Dialektologie zum ersten Mal an. Weinreich anerkannte einerseits vorbehaltlos den Erfolg des Strukturalismus in der Sprachbeschreibung, fragte jedoch andererseits aus Neugierde, wie man denn dialektale Variation aus dieser Perspektive verstehen sollte. Vielleicht, erwog er, könnte Dialektologie interessanter werden, wenn

Sprachstruktur eine größere Rolle spielte. Dies passte auch zur Breite seines Interesses als allgemeiner Sprachwissenschaftler, der zu verschiedensten linguistischen Subdisziplinen, u.a. Sprachkontakt, Zweitspracherwerb und Semantik beigetragen hatte.¹

Der Aufsatz Weinreichs ist auch bewusst programmatisch, z.B. wo der Autor um mehr Aufmerksamkeit für die strukturellen Auswirkungen der Fusion von zwei Phonemen wirbt, wobei er nicht versäumt, sogar einige Konsequenzen für die Feldarbeit zu skizzieren. Man erkennt das programmatische Interesse auch dort, wo Weinreich vorschlägt „Diasysteme“ als theoretische Begrifflichkeit einzuführen. Letztere sollten es ermöglichen, verschiedene verwandte Dialekte doch als systematisch verbunden zu beschreiben und dadurch den Verbund verwandter Dialekte kompakter und einsichtiger zu erfassen. Schließlich erkennt man in Weinreichs Bemerkung, dass zwei Dialekte dem phonetisch gleichen Element [a] einen ganz anderen Status im Phonemsystem (als allophonische Realisierung bloß eines bzw. zweier Phoneme) zuweisen kann, sein programmatisches Interesse.

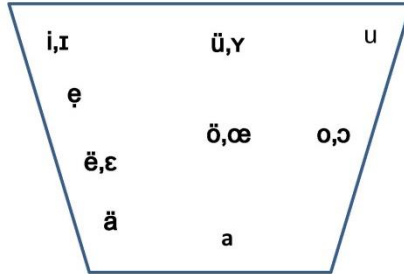
Dem Werk Weinreichs kann man noch heute viel Anregung abgewinnen, wenn er darauf besteht, eine „einheitliche Sprachtheorie“ (p.388) zu entwickeln, in der u.a. strukturelle Phonologie ebenso wie Dialektologie eine Grundlage erhalten.

Obwohl die Rezeption nicht unmittelbar erfolgte, haben doch viele sehr kompetente Forscher insbesondere im deutschsprachigen Raum auch an der Ausführung des Weinreichschen Programms gearbeitet. Moulton (1960, 1968) hat Phoneminventare in der Schweiz verglichen, und wir betrachten eines seiner Beispiele näher, um eine Idee dieser Forschung zu vermitteln. Die Diskussion unten folgt den Präsentationen von Goossens (1969: 18ff) sowie Niebaum & Macha (2006: 73).

Moulton (1960) vergleicht verschiedene Dialektpaare aus der nördlichen Schweiz bezüglich ihrer Kurzvokale. Weil die konkreten Vergleiche zwischen Basilekten schnell komplex werden, nehmen wir hier Genüge mit einem Vergleich zwischen den Kurzvokalsystemen im Mittelhochdeutschen und im Neuhochdeutschen (Goossens 1969:18). Beide Systeme schöpfen ihr Kurzvokalinventar aus der gleichen, nicht allzu großen Menge (siehe Figur 1). Aber wo

¹ Diese Bemerkung ist natürlich umso geeigneter in einem Sammelband zu Ehren Werner Abrahams, dessen Interessebreite und Produktivität nicht nur in seinen eigenen Schriften zu erkennen sind, sondern auch in seinem Werk als Herausgeber und als Mentor jüngerer Forscher. Es ist immer ein Vorrecht, sich mit Werner über Sprache zu unterhalten. Zu seinen eigenen Schriften zur Dialektologie vgl. u.a. Abraham 1972 bis 2013. Wie Nerbonne (2016) unterstreicht, waren wir in wesentlichen Punkten oft nicht einer Meinung, was das Gespräch auch bereichert hat.

das Neuhochdeutsche nur den Kurzvokal /ɛ/ hat, gab es dazu drei Entsprechungen im Mittelhochdeutschen. Goossens (1969) schlägt vor, das Diasystem wie in Figur 2 zu repräsentieren.



Figur 1 Die gesamten Kurzvokale des Mittelhochdeutschen (Mhd) und des Neuhochdeutschen (Nhd). Ich folge Goossens (1969) darin, dass ich sowohl die mhd graphischen Symbole als auch die (nhd) IPA Symbole gebrauche. Wo zwei zusammenstehen wie [i,ɪ] oben links, ist das erste mhd, das zweite nhd IPA. Man vermutet in diesen Fällen Ähnlichkeit.

$$\left\| \frac{i}{ɪ} \approx \frac{e \sim \ddot{e} \sim \ddot{a}}{\varepsilon} \approx \frac{a}{a} \approx \frac{\ddot{u}}{y} \approx \frac{\ddot{o}}{\text{œ}} \approx \frac{u}{u} \approx \frac{o}{ɔ} \right\|$$

Figur 2 Das mittelhochdeutsch-neuhochdeutsche Diasystem (Goossens 1969:18). Die Doppeltilde , $\ddot{}$ ‘ zeigt einen phonemischen Unterschied, der in beiden Subsystemen gilt, während die einfache Tilde , \sim ‘ einen Phonemunterschied in nur einem Teilsystem anzeigt.

Moulton (1960: 176-177) hat darauf hingewiesen, dass Weinreichs Programm durchaus zu Rätseln führen kann. Bei dem Vergleich zwischen Luzern und Appenzell stellte es sich heraus, dass sie genau die gleichen Kurzvokale haben, nämlich: / ɪ, e, ε, œ, a, ɔ, ɔ, o, u, ʏ, ø, ɜ /.² Abgesehen von der lexikalischen Verteilung der Vokale (Verteilung in verschiedenen Wörtern) sind die Systeme auch gleich – obwohl die Dialektsprecher sie ohne Zweifel als verschieden erfahren. Wenn man aber die lexikalische Verteilung mitberück-

² Bei den vorderen Vokalen gibt es fünf Höhenunterscheidungen, bei den zentralen drei und bei den hinteren Vokalen vier.

sichtigt, teilen die zwei Basisdialekte ein einziges Phonem, /a/, was als Schlussfolgerung aber zu radikal scheint.

Moulton (1960, 1968) war auch nicht allein in seinem Interesse an der strukturellen Dialektologie des Deutschen. Goossens (1969) hat weiter an der Theorie gearbeitet, z.B. über die Konsequenzen des strukturellen Herangehens zu Begriffen wie ‚Lautgrenze‘, ‚Dialektgebiet‘ oder ‚Übergangsgebiet‘. Wiesinger (1970) hat systematisch zwölf verschiedene deutsche Dialekräume bezüglich ihrer Gesamtvokalinventare untersucht und verglichen. Weil er stets vom mittelhochdeutschen Phoneminventar ausgeht, sieht man die strukturellen Konsequenzen der Entwicklung deutlich, aber seinen Ansatz könnte man auch als historisch-strukturell bezeichnen.

Trotz dieser vielversprechenden Entwicklungen hat sich die strukturalistische Dialektologie in dieser Form (in der Beschreibung von Diasystemen) nicht gut etabliert. Hierfür gab es zumindest die folgenden Gründe. Erstens konzentrierte man sich zu sehr auf den Vergleich von Vokalsystemen, vielleicht weil diese sich übersichtlich in zweidimensionalen Graphiken zusammenfassen ließen (siehe oben), womit sich der Vergleich visuell nachvollziehen ließ. Andere Elemente der Phonologie wie allophonische Regeln oder phonotaktische Einschränkungen sind weniger leicht zu visualisieren und entsprechend zu vergleichen.

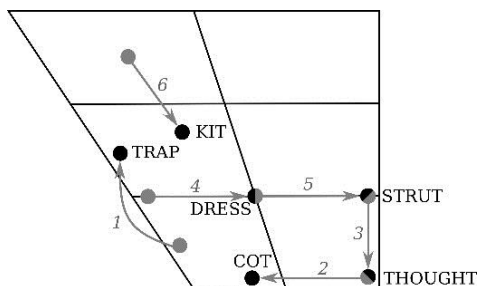
Zweitens wird in den strukturalistischen Ansätzen die subphonemische Variation nicht mitberücksichtigt. Jede Dialektlandschaft kennt Phänomene wie z.B. die Unterscheidung zwischen dem apikalen („Zungenspitzen“) [r], das man im Bairischen und im Plattdeutsch hört, und dem uvularen „(Zäpfchen“) [r], das anderswo vorkommt. Die Unterscheidung ist auch für Nichtlinguisten auffällig, also sozial bedeutsam, spielt aber in der strukturalistischen Analyse keine Rolle. Hierzu kommt, dass die Soziolinguistik oft und überzeugend gezeigt hat, dass Mitglieder der Sprachgemeinschaft nicht nur für absolute Unterschiede im Sprachgebrauch empfindlich sind, sondern auch für Häufigkeiten, wo Variation im Spiel ist. Dieses nennen Weinreich, Labov & Herzog (1968) „*orderly heterogeneity*“ (etwa ‚geordnete Heterogenität‘).

Drittens setzt das Programm voraus, dass man aufgrund verhältnismäßig vollständiger Phonemanalysen (zumindest von Teilbereichen) zu allen zu untersuchenden Varietäten Systemvergleiche erhalten kann. Aber Phonemanalyse setzt phonetische und semantische Daten voraus und ist zudem absolut untrivial, so dass sich verschiedene Forscher über Details öfter streiten können. Chambers und Trudgill (1998) machen die weitere Beobachtung, dass Dialekte manchmal wohl systemgleich sind, sich aber doch unterscheiden, weil die Elemente (Vokale) in verschiedenen Wörtern vorkommen (siehe auch Moulton

1960:176-177); sie ziehen letztendlich den Schluss (S.37), dass diese Forschungslinie versagt hat („failed“).

2.1 Labov und strukturelle Dialektologie

Zweifelsohne betrachtet Labov sein Werk (u.a., Labov, Ash & Boberg, 2005) zurecht als strukturalistisch (siehe z.B. S.48-56), wobei er aber der Demonstration eine innovative Wendung gibt. Anstatt bloß Phoneminventare zu bestimmen und zu vergleichen zeigen Labov et al., wie eine Änderung an einem Element im Phonemsystem einen Druck kreiert, der schließlich anderen Änderungen Vorschub leistet. Hier folgen sie André Martinet (1955), der von „Kettenverschiebungen“ (*reaction en chaines*) gesprochen hat,³ wo Phoneme sich als Glieder in einer Kette verschieben. Auch Labov et al. arbeiten an Vokalsystemen, wo die eine Änderung an einem Vokal es akustisch so in die Nähe eines zweiten bringt, dass Sprecher oft geneigt sind, deswegen auch den zweiten zu ändern. Einen überzeugenden Fall dafür stellen die Kettenverschiebungen der englischen Kurzvokale dar.⁴



Figur 3 Kettenverschiebung der englischen Kurzvokale (Labov et al. 2005). Weil der Vokal in ‚trap‘ höher ausgesprochen wird, entsteht Pstruktureller Raum dafür den Vokal in ‚thought‘ weiter vorne auszusprechen. Weitere Glieder in der Kette folgen dann.

Die Existenz von Kettenverschiebungen untermauert den strukturalistischen Blick, weil die zugrundeliegende Struktur eine Voraussetzung für die Verschiebung ist. Nicht irgendein Vokal ändert sich infolge einer Verschiebung, sondern jener Vokal, dessen Platz im System durch die Verschiebung „usurpiert“ wurde. Der Belang des Systems wird auch nicht durch die Tatsache geschmälert, dass das System in diesem Fall sehr gut mit dem akustischen

³ Martinet unterscheidete auch zwischen „Ziehketten“ (*chaines de traction*) und „Schubketten“ (*chaines de propulsion*), aber wir können hier ruhig dieses Niveau von Detail vernachlässigen.

⁴ <https://commons.wikimedia.org/w/index.php?curid=14867752>

Formantenraum übereinkommt, d.h. das System ist auch physikalisch anwendend. Auch ein akustischer Raum bildet ein System.⁵

Im Allgemeinen haben neuere Autoren sich darauf konzentriert zu zeigen, dass sich Sprachstruktur auf die Variation auswirken kann (Jongen 1982; Barbiers 2010).

3. Quantitative Dialektologie⁶

1.1 Anfänge der Dialektometrie

Jean Séguy (1971, 1973) gilt als Gründer der Dialektometrie, die eine Lösung des dialektologischen Problems der nichtzusammenfallenden geographischen Verteilungen einzelner Sprachphänomene in der Aggregation eingeleitet hat. Dialektologen vor Séguy hatten für jedes Phänomen eine geographische Verteilungskarte gemacht;⁷ wo viele Verteilungen zusammenfielen, konnte man eine Dialektgrenze vermuten; wo die Grenzen zusammenfielen, so dass sie geschlossene Areale umrissen, konnte man verschiedene Dialektgebiete ausmachen. Problematisch bei diesem Herangehen war aber, dass die Verteilungen selten zusammenfielen und dass genauere Aufzeichnungen (z.B. phonetisch oder morphologisch elaboriertere) das Problem nur verschärften (Nerbonne, 2009).

Nach Séguy's frühem Tod hat Hans Goebel (1982) sehr wichtige Arbeit geleistet, u.a. indem er Clusterverfahren zur Bestimmung von Dialektgebieten einführte. In Goebels Arbeiten wird die Perspektive jedes Dorfes berücksichtigt, indem er stets Gesamtabstände bezüglich jedes einzelnen Erhebungsortes berechnet.

Kessler (1995) und Nerbonne et al (1996) haben den Gebrauch des Levenstheinabstands auf phonetische Transkriptionen eingeführt. Die Methode vergleicht Aussprachetranskriptionen (des gleichen Wortes an mehreren Erhebungsorten) und liefert ein numerisches Maß des Unterschieds. Mittels dieser Methode konnte man die Aussprachedaten aus zahlreichen Atlanten relativ

⁵ Natürlich gibt es Debatten, wo dies wichtig ist, z.B. über die besondere kognitive Grundlage der Sprache, aber das ist nicht unser Thema in diesem Artikel.

⁶ Auch Soziolinguisten betrachten ihre Methode, die Frequenzunterschiede in einzelnen Phänomenen mittels logistischer Regression analysiert, als quantitativ (Bayley 2013). Dies war auch korrekt – vor allem als der Name erst gebraucht wurde (Labov, Yaeger & Steiner 1972). Aber die Dialektometrie analysiert öfter hunderte solche Phänomene gleichzeitig (Wieling, Heeringa & Nerbonne 2009) und verdient m.E. den Namen noch mehr. In diesem Artikel werden „dialektometrisch“ und „quantitativ“ synonym gebraucht.

⁷ Wobei Haag (1898) (!) es verdient als ehrenwerte Ausnahme erwähnt zu werden.

direkt analysieren, und man hat die Methode bezüglich der Urteile von Laien und auch von Sprachwissenschaftlern validiert.

Die Dialektometrie hat zahllose Beiträge zur Mutterwissenschaft Dialektologie geleistet. Nachdem Embleton (1993) den Einsatz multidimensionaler Skalierung (MDS) in die Dialektologie befürwortet hat, haben Heeringa und Kleiweg gezeigt, wie man MDS Resultate auf geographische Karten projizieren kann (Nerbonne et al. 1999). Heeringa et al. (2002) führen in die Diskussion das wichtige Nebenthema der Validierung der vorgeschlagenen Maßen für Dialektverwandschaft bzw. -abstand ein; das Thema wird u.a. in Wieling et al. (2014) weiter entwickelt, wobei aufgrund einer Stichprobe von über 800 Sprechern, gezeigt wird, dass der Levenshteinabstand stark mit menschlichen Urteilen über „fremd“ vs. „natürlich“ übereinstimmt ($r \approx 0.8$). Szmrecsanyi (2012) entwickelte korpusbasierte Techniken, um das Vorkommen von morpho-syntaktische Variablen aus Gesprächskorpora herauszulesen und sie in der Folge einer dialektometrischen Analyse zu unterziehen. Die Daten aus Gesprächskorpora spiegeln treu echten Sprachgebrauch wider statt sich auf die Urteile von Dialektsprechern zu verlassen, die öfter als Grundlage in Dialektatlanten dienen, und sie weisen dadurch ein erhöhtes Niveau von Authentizität auf. Grieve, Speelman & Geeraerts (2011) extrahieren variationslinguistische Variablen aus Briefen an die Redaktion regionaler Zeitungen in den USA, und sie wenden Techniken aus „Geographischen Informationssystemen“ (GIS) dazu an die Verteilungen zu analysieren. Prokić, Çöltekin & Nerbonne (2012) bieten quantitative Techniken an, um Merkmale zu bestimmen, die in einem vorgegebenen Gebiet regelmäßig vorkommen, außerhalb dagegen in sehr geringem Maße. Auf diese Weise will man Dialekteigenschaften identifizieren, die für ein in Frage kommendes Gebiet charakteristisch sind.

1.1 Dialektometrie heute

Dialektometrie genießt heute großes Interesse, wie der Übersichtsartikel von Wieling & Nerbonne (2015) belegt, wie auch die Tatsache, dass fast die Hälfte der Artikel in Côté, Knooihuizen & Nerbonne's (2016) Selektion der Beiträge der *Methods in Dialectology Conference* aus einer dialektometrischen Perspektive hervorgehen oder aber versuchen, die Dialektometrie *sec* weiterzuentwickeln. Wir beschränken die Diskussion auf drei neue Ansätze, die besonders vielversprechend zu sein scheinen.

Erstens gibt es mehrere Softwarepakete, mittels derer Analysen ausgeführt werden können, ohne erst zunächst viel in die Entwicklung computerlinguistischen Könnens zu investieren (Goebel 2004; Nerbonne et al. 2011; Aurrekoetxea et al. 2013). Wie Leinonen, Çöltekin & Nerbonne (2016)

dokumentieren, werden solche Pakete vor allem von jungen Forschern dafür entdeckt und auch eingesetzt.

Zweitens können neue Techniken aus der Statistik sehr ertragreich mit dialektometrischen Analysen kombiniert werden, insbesondere Gemischteffekten Regressionsmodelle und generalisierte additive Modelle (GAMs). Im dialektometrischen Aggregationsschritt (siehe oben) droht man die für die Sprachwissenschaft sehr interessanten Details aus dem Auge zu verlieren. Wieling, Nerbonne & Baayen (2011) empfehlen daher, die einzelnen sprachlichen Merkmale (wie etwa Wörter oder phonologische Merkmale) als *Zufallseffekte* zu analysieren (neben Haupteffekten wie geographische Lage, Alter, oder Bildungsniveau). So erhält man innerhalb einer aggregierten Analyse auch Aufschluss drüber, ob sich die Elemente in der Aggregation einheitlich verhalten. In §4.2 (unten) gehe ich weiter auf die GAMs ein.

Drittens entwickelt sich das Interesse für Soziolinguistik innerhalb der Computerlinguistik sehr rasch. Nguyen et al. (erscheint) kündigen die Entstehung einer neuen Subdisziplin, die „Computative Soziolinguistik“ (*Computational Sociolinguistics*) an, wobei sie auf dutzende computerlinguistische Artikel hinweisen, in denen vor allem die Sprache sozialer Media (Twitter, Blogs, Facebook u.ä.) unter die Lupe genommen wird. Sehr beeindruckend sind die Datenmengen, die die Computerlinguisten miteinbeziehen. Eisenstein et al. (2014) analysieren 107 Million „Tweets“ von 2,7 Million Gebrauchern, um u.a. sehr deutliche Auswirkungen der geographische Lage und Demographie (u.a. Rasse und Ethnizität) einordnen zu können. Sie entkräften hiermit auch die Vermutung, dass elektronische Kommunikation nicht ortsgebunden sei.

4. Zur Synthese

Obwohl automatisierte, dialektometrische Methoden die Verarbeitung größerer Datensätze ermöglichen, so dass sie umfassendere und zuverlässigere Ergebnisse über räumliche und soziale Verteilung von Sprachvarianten erzielen, bieten sie bis jetzt wenig explizite Unterstützung für die *strukturelle* Analyse von Sprachdaten. Verglichen mit Weinreich (1954) unterstützen diese Methoden damit zu wenig strukturalistische Ansätze, in denen sprachliche Variation aus der Perspektive des Sprachsystems untersucht wird. Und obwohl die Beweisführung Labovs et al. (2005) überzeugend ist, bleibt sie eine (brillante) Fallstudie ausgewählter Elemente innerhalb einer Sprache. Wenn wir der Zielsetzung Weinreichs umfassender entsprechen wollen, lohnt es sich, über alternative Wege nachzudenken.

1.1 Ein neuer Anlauf

Ein realistisches lohnendes Ziel wäre, den Beitrag der Sprachstruktur quantitativ an der Variation zu messen.⁸ Dies geht deutlich über die früheren Arbeiten der strukturalistischen Dialektologen hinaus, die Dialekte strukturell verglichen (Moulton 1968; Goossens 1969; Wiesinger 1970). Schließlich ist es das Ziel zu messen, welche Rolle die Struktur in der Variation spielt. Der Ansatz vermeidet auch die Voraussetzung, dass man die Strukturen der zu analysierenden Sprachen zuvor einigermaßen vollständig (jedenfalls bezüglich eines Teilgebiets) bestimmt haben muss. Wie wir gesehen haben, wählen Labov et al. (2005) einen anderen Ansatz, wobei sie zeigen, dass die Struktur des Vokalraums zu Kettenverschiebungen geleitet haben muss. Der vorliegende Vorschlag dagegen sieht vor, dass man mittels einer größeren Stichprobe den sprachlichen Abstand zwischen Varietäten (bezüglich eines Niveaus, z.B. der Phonologie) bestimmt, so dass man den Schein vermeidet, Struktureffekte seien nur in gewissen Bereichen mancher Sprachen zu konstatieren.

Nachdem man den aggregierten Abstand zwischen den Paaren einer hinreichenden Anzahl von Sprachvarietäten gemessen hat, müssen dann die in Frage kommenden Struktureigenschaften bestimmt werden. Diese dürfen verschiedenster Art sein, z.B. ‚Fortis-Verschlusslaute aspiriert‘, ‚Anzahl der Vokalhöhen‘ oder ‚Anzahl der Vokalphoneme‘. Manchmal sind diese schon in Projektdaten kodiert, manchmal wird man sie mittels einfacher Programme ableiten müssen, aber es ist jedenfalls leicht an Beispiele zu denken, wo auch manuelles Kodieren unentbehrlich wäre. Das strukturelle Übereinkommen zwischen den Struktureigenschaften kann dann in Regressionsentwürfen als Prädiktor (als unabhängige Variable) fungieren, wobei es das Ziel ist, den aggregierten Abstand vorherzusagen.

1.2 Ein Beispiel

⁸ Um keine Verwirrung zu stiften füge ich hier gerne hinzu, dass ich schon eher zu diesem allgemeinen Thema publiziert habe (Nerbonne 2006), wobei die Herangehensweise in dem 2006- Artikel aber wesentlich anders war. Die Stichprobe bestand aus Material aus 240 Erhebungsorten in LAMSAS (siehe Nerbonne 2006 zu Details). Zuerst habe ich jeden Vokal in jedem Wort (z.B. den ersten Vokal in *Alabama*) in einer Stichprobe von Aussprachen mit 200 Vokalwortkombinationen (VinW, ‚Vokal in Wort‘) extrahiert. Man konnte den Abstand zwischen den Vokalen mittels eines Merkmalsystems ableiten, wobei man 200 Matrizen (eine pro VinW) erhielt. Hieraus wurden Korrelationen zwischen den VinWs berechnet, wonach schließlich eine Faktoranalyse angewandt wurde. Es zeigte sich u.a. dass die verschiedenen Vorkommen eines Phonems größtenteils zu gleichen Faktoren gehörten, z.B. die reduzierten (zweiten) Vokalen in Wörtern wie *closet*, *kitchen*, *ashes*, oder *pallet*. Der heutige Versuch ist leichter zu implementieren und allgemeiner anzuwenden.

Wir geben hier ein Beispiel jener Analyse, die in dieses Programm passt, kehren aber zuvor zurück zur Erklärung der zweiten neuen Art von Statistik, der sich auch die zeitgenössische Dialektometrie bedient.

Generalisierte additive Modelle (GAMs) ermöglichen die Analyse der nichtlinearen Einflüsse wie z.B. den Einfluss der geographische Lage auf die Sprachvariation. Darüberhinaus ist es möglich, multiple unabhängige Kandidateneinflüsse in ein einziges Modell miteinzubeziehen einschließlich nicht-geographischer Einflüsse wie sprachlicher und sozialer Einflüsse.⁹ Um diese Stärken der GAMs zu illustrieren und um insbesondere ihre Brauchbarkeit bezüglich struktureller quantitativer Dialektologie zu demonstrieren, betrachten wir eine Analyse von Ko et al. (2014).

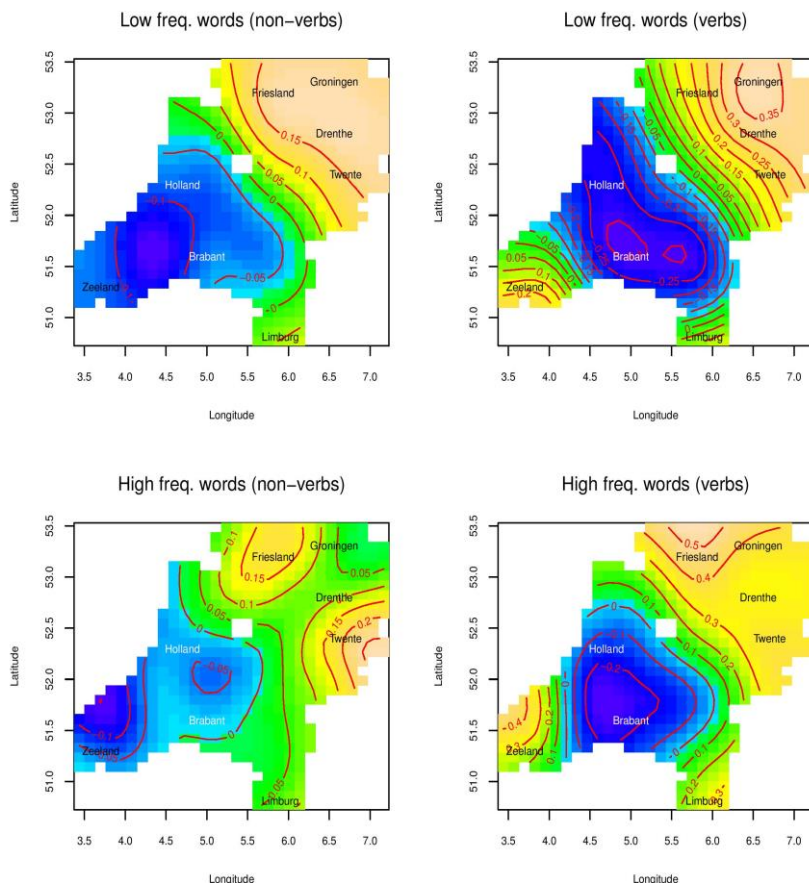
Bezüglich der Nichtlinearität der Analyse wollen wir kurz sein, denn dies ist hier nicht das Hauptthema. Die Kontourenlinien in Figur 4 zeigen die Nichtlinearität des geographischen Einflusses, wo Längengrad und Breitengrad als unabhängige Variablen eingesetzt wurden.¹⁰ Vielleicht lohnt sich aber doch die Bemerkung, dass Nichtlinearität denjenigen Forschern neue Möglichkeiten bieten könnte, die das Verhältnis zwischen Sprachähnlichkeit und geographischem Abstand untersuchen. Jedenfalls vermeidet GAMs zumindest die Notwendigkeit d geographische Lage als eindimensionale Distanz zu einzusetzen – gewiss ein Fortschritt.

Ko et al. (2014) betrachteten außer der geographische Lage zwei sprachinterne Variablen, die uns als Vorbilder für eine Analyse der strukturellen Variablen dienen können. Keine der beiden ist eine strukturelle Variable im Sinne der strukturellen Dialektologie, aber die Möglichkeit, solche Variablen in die Analyse zusammen mit sprachexternen Variablen miteinzubeziehen, ist die gleiche. Die eine Variable war die Wortart (syntaktische Kategorie) des aufgezeichneten Wortes – Verbum oder nicht Verbum, die in der explorativen Phase der Analyse aufgefallen war. Die zweite sprachinterne Variable in ihrer Analyse war das Verhältnis von Vokalen zu Konsonanten (V/C) in der Aussprache des Wortes. Dies ist zum Beispiel im Falle ‚Pa‘ [pa] 1, ‚Opa‘ [ˈo.pa] 2, und ‚schlank‘ [ʃlʌŋk] 0.25. Der Grund, dies in die Analyse mitzunehmen, war die weit verbreitete Meinung, dass Vokale mehr Variabilität aufweisen (Keating et al. 1994), so dass Wörter mit größerem V/C eher eine abweichendere Aussprache erwarten lassen. Ko et al.

⁹ Durch die Möglichkeit, soziale und geographische Einflüsse in ein Modell zu integrieren, trägt die Dialektometrie, wenn GAMs gebraucht werden, übrigens technisch zu der Realisierung des Chambers-Trudgill (1998) Programms bei, worin Dialektologie und Soziolinguistik als eine einzige Disziplin gesehen werden sollten.

¹⁰ Mehr in Ko et al. (2014) en Details in Wood, Goude & Shaw (2014).

analysierten auch eine dritte Variable, nämlich die Häufigkeit des ausgesprochenen Wortes, die die meisten Linguisten auch als sprachextern einordnen würden und wo die Vermutung stark ist, dass häufigere Wörter sich weniger schnell ändern (Bybee 2002; Pagel, Atkinson & Meade 2007).



Figur 4. Geographische Analyse niederländischer Dialektvariation zusammen mit der Auswirkung von Frequenz und Wortart. Je stärker die Blautöne, umso standardähnlicher ist die Aussprache. Es lagen keine Daten dort vor, wo das weiße Quadrat steht. Siehe Text zu weiterer Erklärung. Aus Ko et al. (2014).

Figur 4 zeigt wie standardnah die niederländischen Dialektausssprachen sind, abhängig nicht nur von der geographischen Lage, sondern auch von der Wortart (Verbum vs. andere Wortarten) und Worthäufigkeit (Frequenz). Gebiete mit standardnäherer Aussprache sind dunkelblau getönt; Gebiete mit weniger standardnaher Aussprache werden dann hell blau, grün, gelb en dann hellbrau gefärbt.. Im Vergleich zwischen den zwei Karten links und den anderen zwei (rechts) kann man feststellen, dass bezüglich der geographischen Lage Verben sich stärker als Nichtverben differenzieren lassen (d.h. es gibt mehr „Isolinien“) und dass sie insgesamt standardnäher ausgesprochen werden. Im Vergleich zwischen den zwei Karten oben und den anderen zwei (unten) kann man feststellen, dass häufiger auftretende Wörter sich leichter der Anziehungskraft der Standardsprache widersetzen können und dass insbesondere die häufigen Verben (unten rechts) extrem abweichende Aussprachen beibehalten können. Dies ist am Grad der Abweichung (die Zahlen, die bei den Isolinien abzulesen sind) erkennbar. Die Isolinien 0.4-0.5\$ erscheinen nur hier --- in Niedersachsen und Friesland.

Die numerischen Resultate der Analyse bestätigen den Eindruck der Karten. Es stellte sich heraus, dass der Beitrag der Wortart nicht sonderlich groß war (nur $p < 0.03$, trotz einer großen Datenmenge), aber es illustriert immerhin, dass eine sprachinterne Variable mit sprachexternen Variablen wie z.B. geographische Lage quantitativ verglichen werden kann. Häufigkeit war auch signifikant, was die Resultate von Wieling et al. (2011) bestätigte. Schließlich war das V/C Verhältnis extrem signifikant ($p < 0.001$), was natürlich bestätigt, dass Vokale sich leichter ändern, auch in Richtung einer Standardsprache.

Mit Nachdruck weisen wir hier auf die Möglichkeit hin, strukturelle Faktoren wie ‚Wortart‘ oder ‚Vokal/Konsonant-Verhältnis‘ als unabhängige Variablen in das Regressionsmodell aufzunehmen. Dies sind keine strukturellen Variablen im Sinne Weinreichs et alii, aber sie sind aufschlussreich. Man kann genauso eine Variable wie ‚Drei-niedrige-Vorderkurzvokale‘ definieren, damit sich die mhd. Varietäten mit /e, ε, æ/, die den Wert ‚1‘ zugewiesen bekommen, von den nhd. mit nur /ε/ (folglich mit dem Wert ‚0‘) unterscheiden.

In einem Punkt ist die Analyse von Ko et al. (2014) leicht irreführend, nämlich darin, dass dieser Artikel auf den Abstand der Basisdialekte zum Standard fokussiert war anstatt auf den Abstands aller Basisdialekte zu allen anderen. Es ist natürlich einfacher, sich auf das Verhältnis Standard-Basisdialekt zu konzentrieren statt das Verhältnis eines jeden Basisdialekts zu allen anderen zu analysieren. Die Prinzipien bleiben aber gleich.

Im Gegensatz zu den älteren Arbeiten in struktureller Dialektologie wird hier nicht vorgeschlagen, dass man von einer annähernd vollständigen Beschreibung eines Teilbereichs der Sprache ausgeht. Wir brauchen z.B. nicht von einem vollständigen Vokalinventar auszugehen, um den Beitrag einer Phonemspaltung zu dem aggregierten Abstand zwischen zwei Varietäten zu messen. Es besteht kein starker Gegensatz zu der Arbeit von Labov et al. (2005); wo diese Autoren den Nachweis geliefert haben, dass Struktur eine Rolle spielt. Hier, wird vorgeschlagen, dass man darüber hinaus den Beitrag zum aggregierten Abstand zwischen Varietäten auch einschätzt.

Es sollte vollkommen klar sein, dass der Vorschlag des vorliegenden Aufsatzes keine Realisierung des strukturalistischen Programms Weinreichs bedeuten kann. Einerseits verlangen wir keine vollständigen Strukturbeschreibungen der Dialekte – auch nicht von Teilbereichen der Sprache – obwohl Weinreich und seine Nachfolger immer hiervon ausgegangen waren. Ein strenger Strukturalist dürfte damit unzufrieden sein. Andererseits kann man Einschätzungen des Beitrags bestimmter Strukturelemente relativ frei ausführen, d.h. ohne die Voraussetzung einer vollständigen Beschreibung, dies mit allen Vorteilen partieller Analysen. Oft einmal kann man bezüglich der Teilanalysen relativ sicher sein, wo aber der Schritt zur Vollständigkeit dennoch problematisch bleibt.

5. Schlussbemerkungen

In diesem Aufsatz wurde die heutige Kluft zwischen einerseits strukturellen Analysen von Dialekten und Dialektunterschieden und andererseits quantitativen, dialektometrischen Untersuchungen beschrieben. Wo die erste Forschungslinie die quantitative Methodologie größtenteils vermeidet, um sich mit ungeteilter Energie den strukturellen Unterschieden zu widmen, misst die zweite Forschungslinie unermüdlich verschiedene Phänomene, um die aggregierten Abstände zwischen Varietäten zu bestimmen, zu validieren, und zu interpretieren. Dabei kann letztere die Sprachstruktur aus dem Auge verlieren, obwohl ja jede Sprache hochgradig und vielschichtig strukturiert ist. Es ist an der Zeit, diese Kluft zu überwinden, und dieser Aufsatz hat vorgeschlagen, dass man es zum Ziel machen sollte, den Beitrag der Struktur am aggregierten Abstand zu bestimmen. Man hätte dann auf jeden Fall eine Antwort auf die Frage, wie wichtig die Struktur in der Sprachvariation, und insbesondere in der Dialektologie, ist.

Weinreich (1954: 388) betrauerte die Kluft zwischen strukturalistischer und dialektologischer Sprachbeschreibung, und rief explizit dazu auf, sich nicht mit einem „Waffenstillstand“ zufrieden zu geben, sondern den wissenschaftlichen

Ehrgeiz zu entwickeln, die theoretischen Zusammenhänge dieser zwei Perspektiven besser zu begreifen. Im erweiterten wissenschaftlichen Rahmen der quantitativen Dialektologie wurde in diesem Aufsatz die Position eingenommen, dass es möglich und sinnvoll ist, den Beitrag von Sprachstruktur an der sprachlichen Variation zu messen.

Dankwort

Mein Dank gilt Werner Abraham für Diskussion über diese aber auch viele andere Arbeiten, und auch Vinnie Ko für Hilfe mit der Karte.

Literaturverzeichnis

- Abraham, Werner 1972. Das Vokalsystem der Mundart des Montafons. In: *Zeitschrift für Dialektologie und Linguistik* 38, 95 - 120.
- Abraham, Werner / Anko Wiegel (1993), „Reduktionsformen und Kasus-synkretismen bei deutschen und niederländischen Pronomina.“ In: Werner Abraham / Josef Bayer (Hrsg.) *Dialektsyntax*.. Westdeutscher Verlag: Op-laden. 12-49.
- Abraham, Werner / Elisabeth Leiss (2013), (Hrsg.). *Dialektologie in neuem Gewand. Zu Mikro-/Varietätenlinguistik, Sprachenvergleich und Universalgrammatik*. LB-Sonderband. Hamburg: Buske.
- Abraham, Werner (2014), The Upper German differential: Main Austrian-Bavarian vs. (High-)Alemannic differences. In: Günter Grewendorf / Helmut Weiß (eds.): *Bavarian syntax. Contributions to the theory of syntax*.. [Linguistik Aktuell / Linguistics Today 220]. Amsterdam: John Benjamins. 305-335.
- Aurrekoetxea, Gotzon / Karmele Fernandez-Aguirre / Jesus Rubio / Borja Ruiz / Jon Sánchez (2013), „‘DiaTech’: a new tool for dialectology“. In: *LLC: Journal of Digital Scholarship in the Humanities* 28, 23–30
- Barbiers, Sjef (2010) „Language and space: Structuralist and generative approaches“. In: Peter Auer / Jürgen Erich Schmidt (eds.) *Language and Space. An International Handbook of Linguistic Variation. Volume 1: Theories and Methods*, Berlin: Mouton De Gruyter. 125-142.
- Bayley, Robert (2013), „The quantitative paradigm“. In: J.K.Chambers / Natalie Schilling-Estes (eds.): *The handbook of language variation and change*. Boston: Wiley, 117-141.
- Bybee, Joan (2002), „Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change“. In: *Language variation and change* 14(3), 261-290.

- Chambers, J.K. / Peter Trudgill (1998), *Dialectology*. Cambridge: Cambridge University Press.
- Côté, Marie-Hélène / Remco Knooihuizen / John Nerbonne (eds.): *The Future of Dialects. Selected Papers from Methods in Dialectology XV*, Berlin: Language Science Press (*Language Variation* 1).
- Eisenstein, Jacob / Brendan O'Connor / Noah Smith / Eric Xing (2014), „Diffusion of lexical change in social media“. In: *PloS ONE*, 9(11), e113114. doi:10.1371/journal.pone.0113114
- Embleton Sheila (1993), „Multidimensional scaling as a dialectometric technique: Outline of a research project“. In: Reinhard Köhler / Burkhardt Rieger (eds.): *Contributions to quantitative linguistics*. Dordrecht: Kluwer, 267-76.
- Goebel, Hans (1982), *Dialektometrie: Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie*. Vienna: Österreichische Akademie der Wissenschaften.
- Goebel, Hans (2004), „VDM—visual dialectometry. Vorstellung eines dialektometrischen Software-Pakets auf CD-ROM“. In: Wolfgang Dahmen / Günter Holtus / Johannes Kramer / Michael Metzeltin / Wolfgang Schweickard / Otto Winkelmann (Hrsg.): *Romanistik und neue Medien*. Tübingen: Narr, 209–41.
- Goossens, Jan (1969), *Strukturelle Sprachgeographie*. Heidelberg: Winter.
- 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(2). 193-221.
- Haag, Karl (1898), *Die Mundarten des oberen Neckar- und Donaulandes (schwäbisch-alemannisches Grenzgebiet: Baarmundarten)*. Reutlingen: Buchdruckerei Egon Hutzler.
- Heeringa, Wilbert / John Nerbonne / Peter Kleiweg (2002), „Validating dialect comparison methods“. In: Wolfgang Gaul / Gerd Ritter (eds.): *Classification, automation, and new media. Proc. 24th Ann. Conf., Gesellschaft für Klassifikation*, Berlin: Springer, 445-452.
- Jongen, René (1982), „Theoriebildung der strukturellen Dialektologie“. In: Besch, Werner (Hrsg.) *Dialektologie: ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Walter de Gruyter, 248-277.
- Keating, Patricia / Björn Lindblom / James Lubker / Jody Kreiman (1994), „Variability in jaw height for segments in English and Swedish VCVs“. In: *Journal of Phonetics* 22(4), 407-422.
- Kessler, Brett (1995), „Computational dialectology in Irish Gaelic“. In: *Proceedings of the 7th Conference of the European Chapter, Association for Computational Linguistics*. Shroudsburg, PA: ACL, 60-66. Avail.

- www.aclweb.org/anthology/E/E95/E95-1009.pdf
- Labov, William / Sharon Ash / Charles Boberg (2005), *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: Walter de Gruyter.
- Labov, William / Malcah Yaeger / Richard Steiner (1972), *A quantitative study of sound change in progress*. Band 1. US Regional Survey.
- Leinonen, Therese / Çağrı Çöltekin / John Nerbonne (2016), „Using Gabmap“. In: *Lingua* 178, 71-83.
- Martinet, André (1955), *Économie des changements phonologiques*. Bern: Francke.
- Moulton, William G. (1960), „The short vowel systems of northern Switzerland: a study in structural dialectology“. In: *Word* 16(2), 155-182.
- Moulton, William G. (1968), „Structural dialectology“. In: *Language* 44(3), 451-466.
- Nerbonne, John (2006), „Identifying linguistic structure in aggregate comparison“. In: *Literary and Linguistic Computing* 21(4), 463-475.
- Nerbonne, John (2009), „Data-driven dialectology“. In: *Language and Linguistics Compass* 3(1), 175-198.
- Nerbonne, John (2016), Rezension von Werner Abraham / Elisabeth Leiss (Hrsg.) *Dialektologie im neuen Gewand*. 2013. *Linguistische Berichte Sonderheft 19. Zeitschrift für Dialektologie und Linguistik* 82., 88-92.
- Nerbonne, John / Rinke Colen / Charlotte Gooskens / Peter Kleiweg / Therese Leinonen (2011), „Gabmap – a web application for dialectology“. In: *Dialektologia: revista electrònica*, 65-89.
- Nerbonne, John / Wilbert Heeringa / Peter Kleiweg (1999), „Edit distance and dialect proximity“. In: David Sankoff / Joseph Kruskal (eds.): *Time warps, string edits and macromolecules: The theory and practice of sequence comparison, 2nd ed.*, Stanford: CSLI Press. i-iv.
- Nerbonne, John / Wilbert Heeringa / Erik van den Hout / Peter van der Kooij / Simone Otten / Willem van de Vis (1996), „Phonetic distance between Dutch dialects“. In: Gert Durieux / Walter Daelemans / Steven Gillis (eds.): *CLIN VI: Proceedings of the sixth CLIN meeting*. Antwerp, Centre Dutch Language & Speech (UIA). 185-202.
- Niebaum, Hermann / Jürgen Macha (²2006, ¹1983) *Einführung in die Dialektologie des Deutschen*. Tübingen: Niemeyer.
- Nguyen, Dong / A. Seza Doğruöz / Carolyn P. Rosé / Franciska de Jong (zu erscheinen), „Computational sociolinguistics: A survey“. In: *Computational Linguistics, arXiv preprint arXiv:1508.07544*.
- Pagel, Mark / Quentin D. Atkinson / Andrew Meade (2007), „Frequency of word-use predicts rates of lexical evolution throughout Indo-European his-

- tory“. In: *Nature*, 449.7163, 717-720.
- Prokić, Jelena / Çağrı Çöltekin / John Nerbonne (2012), „Detecting shibboleths“, *Proc. EACL 2012 Joint Workshop of LINGVIS & UNCLH*. Shroudsburg: Association for Computational Linguistics. 72-80.
- Séguy, Jean (1971), „La relation entre la distance spatiale et la distance lexicale“. In: *Revue de Linguistique Romane* 35, 335-357.
- Séguy, Jean (1973), „La dialectométrie dans l'Atlas linguistique de la Gascogne“. In: *Revue de Linguistique Romane* 37, 1-24.
- Szmrecsanyi, Benedikt (2012), *Grammatical variation in British English dialects: A study in corpus-based dialectometry*. Cambridge: Cambridge University Press.
- Weinreich, Uriel (1954), „Is a structural dialectology possible?“. In: *Word* 10(3), 388-400. DOI: 10.1080/00437956.1954.11659535
- Weinreich, Uriel / William Labov / Marvin I. Herzog (1968) „Empirical foundations for a theory of language change“. In: Winfred P. Lehmann / Yakov Malkiel (eds.): *Directions for historical linguistics*. Austin: University of Texas Press. 95–188.
- Wieling, Martijn / Jelke Bloem / Kaitlin Mignella / Mona Timmermeister / John Nerbonne (2014), „Automatically measuring the strength of foreign accents in English“. In: *Language Dynamics and Change* 4(2), 253-269.
- Wieling, Martijn / Wilbert Heeringa / John Nerbonne (2007), „An aggregate analysis of pronunciation in the Goeman-Taeldeman-Van Reenen-Project data“. In: *Taal en Tongval* 59(1), 84-116.
- Wieling, Martijn / John Nerbonne (2015), „Advances in dialectometry“. In: *Annual Review of Linguistics* 1, 243-264.
- Wiesinger, Peter (1970), *Phonetisch-phonologische Untersuchungen zur Vokalentwicklung in den deutschen Dialekten*. 2 Bände. Berlin: de Gruyter.
- Wood, Simon / Yannig Goude / Simon Shaw (2015), „Generalized additive models for large data sets“. In: *Journal of the Royal Statistical Society: Series C* 64(1), 139-155.