

Dialektologie
HSK 1.1



Handbücher zur Sprach- und Kommunikationswissenschaft

Herausgegeben von
Gerold Ungeheuer und
Herbert Ernst Wiegand

Band 1.1

Walter de Gruyter · Berlin · New York
1982

Dialektologie

Ein Handbuch zur deutschen
und allgemeinen Dialektforschung

Herausgegeben von
Werner Besch · Ulrich Knoop
Wolfgang Putschke · Herbert Ernst Wiegand

Erster Halbband

Walter de Gruyter · Berlin · New York
1989

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Handbücher zur Sprach- und Kommunikationswissenschaft / hrsg. von Gerold Ungeheuer u. Herbert Ernst Wiegand. — Berlin; New York: de Gruyter
NE: Ungeheuer, Gerold [Hrsg.]
Bd. 1. — Dialektologie

Dialektologie: e. Handbuch zur dt. u. allg. Dialektforschung / hrsg. von Werner Besch ... — Berlin; New York: de Gruyter
(Handbücher zur Sprach- und Kommunikationswissenschaft; Bd. 1)
NE: Besch, Werner [Hrsg.]
Halbbd. 1 (1982).
ISBN 3—11—005977—0

© Copyright 1982 by Walter de Gruyter & Co., vormals G. J. Göschen'sche Verlagshandlung — J. Guttenberg, Verlagsbuchhandlung — Georg Reimer — Karl J. Trübner — Veit & Comp., Berlin 30. Printed in Germany
Alle Rechte des Nachdrucks, der photomechanischen Wiedergabe, der Herstellung von Photokopien — auch auszugsweise — vorbehalten.
Satz und Druck: H. Heenemann GmbH & Co, Berlin
Bindearbeiten: Lüderitz & Bauer, Berlin

Tokugawa/Yamamoto 1967 = Munemasa Tokugawa/Takeshi Yamamoto: An attempt to draw a linguistic map with a computer. In: *Mathematical Linguistics* 40. 1967, 27—30.

Tollenaere 1969 = F. de Tollenaere: Automation und Mundartforschung. In: *Korrespondenzblatt des Vereins für niederdeutsche Sprachforschung* 76. 1969, 57—59.

Uskup/Al-Azzawi 1972 = Frances Land Uskup/Mary Lee Al-Azzawi: Editing and printing a dialect atlas by computer. In: *American Speech* 47. 1972, 203—210.

Vasilyev 1979 = A. S. Vasilyev: Algorithm for machine reading of normalized demographic maps. In: M. Rimmel (Hrsg.): *Symposium. Mathematical Processing of Cartographic Data* (Tallinn, December 18—19, 1979). *Summaries*. Tallinn 1979, 93—97.

Veith 1970 = Werner Veith: [— explikative, + applikative, + komputative] Dialektkartographie. Ihre wissenschaftlichen Voraussetzungen und Möglichkeiten auf der Grundlage der kontrastivtransformationellen Methode und der automatischen Datenverarbeitung. (= *Germanistische Linguistik* 4. 1970).

Veith 1977 = Werner H. Veith: Probleme der Datenauswahl für phonologische Dialektatlanten. Unter besonderer Berücksichtigung des Deutschen Sprachatlas. In: *Germanistische Linguistik* 3—4/77, 137—144.

Viereck 1980 = Wolfgang Viereck: Dialektometrie und englische Dialektologie. In: *Grazer Linguistische Studien* 11/12. 1980, 335—356.

Vorwerk 1977 = Erich Vorwerk: Untersuchungen über die Ähnlichkeit von Zeichenketten in großen Datenbeständen. In: *Bericht Nr. 114 der Gesellschaft für Mathematik und Datenverarbeitung*. 1977.

Weinreich 1964 = Uriel Weinreich: Machine aids in the compilation of linguistic atlases. In: *American Philosophical Society. Year Book* 1963. Philadelphia 1964, 622—625.

Westerhoff 1977 = Wilfried Westerhoff: Zur Behandlung von Mehrfachmeldungen in der Automatischen Wortkartographie. In: *Germanistische Linguistik* 3—4/77, 211—224.

Westerhoff 1979 = Wilfried Westerhoff: Ein Programmpaket zur automatischen Erzeugung sprachgeographischer Karten. Münster 1979. *Schriftenreihe des Rechenzentrums der Universität Münster* Nr. 39.

Wood 1969 = Gordon R. Wood: Dialectology by computer. *International Conference on Computational Linguistics*. Preprint No. 19. Stockholm 1969.

Wood 1972 = Gordon R. Wood: The computer in analysis and plotting. In: *American Speech* 47. 1972, 195—202.

*Wolfgang Putschke, Robert Neumann,
Marburg*

45. Ansätze zu einer computativen Dialektometrie

1. Problemstellung
2. Meßansatz
3. Q-analytische Meßmomente
4. R-analytische Meßmomente
5. Literatur (in Auswahl)

1. Problemstellung

Wort und Sache sind überaus rezent und gehen auf den Tolosaner Romanisten J. Séguéy (1973 a, 1) zurück. Demnach ist die *Dialektometrie* im vollsten Wortsinn in statu nascendi und vorderhand mehr Programm als etablierte Realität. Die in der Folge gezeigten Ansätze (Séguéy verwendet noch keine EDV) weisen die Dialektometrie als eine quantifizierend arbeitende Sprachgeographie aus (vgl. Art. 38), wobei es methodologisch zu starken Anlehnungen an die numerische Taxonomie (Taxometrie) (Sneath/Sokal 1973, Sodeur 1974, Vogel 1975), an allgemeine

Probleme der empirischen Sozialforschung (vgl. dazu etwa König 1974), an andere -metrien wie etwa die Biometrie (Cavalli-Sforza/Lorenz 1972), die Statistik im weitesten Sinn (Schaich 1977) und die quantitativ arbeitende Geographie (Bahrenberg/Giese 1975, Cole/King 1968, Haggett 1973, Hammond/McCullagh 1974) kommt. In diesem Sinn als 'dialektometrisch' anzusprechende Arbeiten liegen vor von Goebel (1976, 1977, 1978 a, b; 1979 a, 1980) und von Séguéy (1971, 1973 a, 1973 b).

Es darf wohl mit Recht behauptet werden, daß sämtliche bislang in allen Teilphilologien vorliegenden *Sprachatlanten*, die ausgezeichnete empirische Datenbasen darstellen, in bezug auf die ihnen innewohnende Datenstruktur noch unzulänglich bekannt sind ('Datenfriedhöfe'). Dies deshalb, weil es erst seit der verallgemeinerten Verwendung der EDV möglich geworden ist, die Sprach-

atlasdaten in kumulativer Synopse zu betrachten. Allerdings ist diese neu eröffnete Möglichkeit zur Synthese ein überaus arbeits- und methodenaufwendiger Forschungskomplex.

2. Meßansatz

Sprachatlanten lassen sich ihrem Aufbau nach als zweidimensionale Matrizen darstellen: Atlaspunkte mal Atlaskarten (vgl. Abb. 45.1 und 45.2). Üblicherweise müssen die zur Vermessung anstehenden empirischen Rohdaten in eine derartige Matrix 'eingefüllt' werden, um so in der für die Verrechnung notwendigen Ordnung zur Verfügung zu stehen. Zu diesem Vorgang einige terminologische und sachliche Klarstellungen.

2.1. Taxandum vs. Taxat

2.1.1. Taxandum

Die erste Überlegung des Dialektometers zu Arbeitsbeginn hat den zu klassifizierenden (data taxanda) empirischen Rohdaten zu gelten. Welchen Objektbereich, welchen Sprachatlas vermessen? Das gesamte Punktenetz über alle Atlaskarten hinweg oder nur über jeweils Teilbereiche davon? Notgedrungenmaßen werden derartige Überlegungen weniger objektbezogen (mit Blick auf die Qualität der Rohdaten) als vielmehr methoden- und arbeitsaufwandbezogen erfolgen. Dies muß man unbedingt bedenken, um bei der Interpretation der dialektometrischen Resultate nicht in zu weitgehende Generalisierungen zu verfallen. Man sollte einen praktikablen *Kompromiß* aus möglichst viel Atlaspunkten (mehr als 100) und Atlaskarten (mehr als 200) schließen, dabei aber nie vergessen, daß die Bedingungen des Zustandekommens dieses Kompromisses keineswegs als aleatorisch gesteuerte Auswahlverfahren interpretiert werden können, wie sie etwa aus der empirischen Sozialforschung bekannt sind (vgl. Scheuch 1974). Die Wahl des Taxandums befrachtet den zu vermessenden Objektbereich unweigerlich mit einer aus dem *Wissensstand* des Dialektometers einfließenden Information, die zudem selbst kaum meßbar ist (Problem der Reliabilität und der Repräsentativität; vgl. dazu Altmann/Lehfeldt 1973, 71 f.).

2.1.2. Taxat

Aus dem Taxandum muß unter Fortführung des Taxierungsprozesses vonseiten des Dia-

lektometers jene Datenmenge herausgeschält werden, die aus disjunkten Einheiten besteht, die ihrerseits sich direkt in die Zellen der Meßmatrix einfüllen lassen (datum taxatum > Taxat). Dieser vom Taxandum zum Taxat fortgeführte *Taxierungsprozeß* erfordert noch viel tiefgreifendere datenreduktionistische Eingriffe vonseiten des Dialektometers, als dies bei der Auswahl des Taxandums der Fall war. *Beispiel*: Gegeben sei eine 'Sprachatlaskarte' mit 3 Atlaspunkten zum Konzept 'Bruder' (frei nach AIS I 13): Punkt 1: *fradɛl*, Punkt 2: *frar*, Punkt 3: *fratɛllo*. Soweit das Taxandum. Wie daraus die für die Matrizenbeschickung geeigneten Taxate ableiten? Der Dialektometer kann dabei auf die grammatischen Kategorien Phonetik, Morphologie, Syntax und Lexikon zurückgreifen und die Taxanda wahlweise als Träger phonetischer, morphologischer, etc. Eigenschaften auffassen. *Fradeł*, *fratɛllo* (< lat. FRATĒLLU) und *frar* (< lat. FRĀTRE) zeigen als gemeinsame phonetische Eigenschaft die Erhaltung von lateinisch FRA-, stellen also insofern identische Taxate dar, wenn man dabei von der verschiedenen Akzentstruktur absieht. Aber derartige autoritativ gesetzte, also nach eigenem Dafürhalten vorzunehmende Analyseeinschränkungen sind einfach unvermeidbar. Ein entsprechender Matrizeneintrag (vgl. Abb. 45.1) würde eine identisch besetzte Zeile ergeben. Alle drei Atlaspunkte tragen das gleiche nominale Taxat, etwa a. Eine andere Möglichkeit bestünde darin, unter Ausnützung der allgemeinen Kenntnisse der diachronen Entwicklung der romanischen Sprachen die Taxanda *fradeł* und *fratɛllo* als Nachfolgeformen von FRATĒLLU einerseits und das Taxandum *frar* als Nachfolgeform von FRĀTRE andererseits als je ein Taxat zu betrachten und somit zu einem Matrizeneintrag mit zwei nominalen Taxaten (etwa a und b) pro Zeile zu gelangen. Der Dialektometer kann derartige Analysen kategoriell getrennt vornehmen (also etwa nur phonetische Analysen durchführen), aber auch seine Meßmatrix mit Taxaten verschiedener grammatischer Kategorien beschicken. — Die beiden Beispielsanalysen zeigen zudem deutlich, daß man aus einer Sprachatlaskarte durch entsprechende Taxierung mehrere Arbeitskarten ziehen kann. Dabei wird klar, daß der Taxierungsprozeß von forschungsautoritativen Direktiven gesteuert wird, die als Zusatzinformation mit dem ursprünglichen Informationsgehalt des Taxandums additiv und subtraktiv interferieren. Hier werden

kommende Dialektometer genauso, wie dies in anderen empirischen Sozialwissenschaften geschehen ist (Cartwright 1953), im intersubjektiven Dialog versuchen müssen, wenigstens innerhalb des Rahmens ihrer jeweiligen Teilphilologie zu einer möglichst grossen 'Objektivität' bei der Taxierung zu gelangen: „Perhaps the foremost challenge is the development of a generally acceptable system for coding and scaling characters.” (Sneath/Sokal 1973, 427). Die Terminologie Taxandum-Taxat wurde nach einem bei Engelen (1971, 12) referierten Vorschlag geprägt.

2.2. Gleichgewichtung oder Ungleichgewichtung der Taxate: zum Problem des Adansonismus

Bei der Erstellung biologischer Klassifikationen wurde vom französischen Botaniker Michel Adanson (1727—1806) erstmals explizit das Prinzip aufgestellt, daß alle an den zu klassifizierenden Objekten beobachtbaren Merkmale gleich zu gewichten seien (vgl. Sneath/Sokal 1973, 5; Sodeur 1974, 42 f.; Vogel 1975, 57 f.). Dieses Prinzip wurde und wird zwar bis heute mehrheitlich befolgt (vor allem bei den numerischen Taxonomen), ist aber im Grunde doch *kontrovers* geblieben. So könnte man sich rechtens fragen, ob im Falle der Dialektometrie nicht phonetisch relevante Taxate innerhalb der Meßmatrix höher zu gewichten seien als etwa lexikale Taxate. Innerhalb der Sprachwissenschaft fehlt es diesbezüglich ja nicht an z. T. höchst divergierenden Meinungen, denen zufolge etwa „Wörter leicht wandern,, (= geringes typologisches Gewicht des Lexikons), die „Dialektsyntax kaum landschaftliche Varianten zeige,, (= geringe Bedeutung der Syntax für typologische Abgrenzungen), aber die „Phonetik überaus stabile Raumkammerungen erbe,, (= hohes typologisches Gewicht der Phonetik). Der hier verwendete dialektometrische Ansatz arbeitet mit dem Prinzip der *Gleichgewichtung*, ist also adansonistisch und folgt damit dem in Taxometrie und Biologie mehrheitlich etablierten Brauch. Im übrigen sind mit gleichgewichteten Taxaten besetzte Meßmatrizen statistisch einfacher zu behandeln und daher für eine Neudisziplin vorzuziehen. Inwieweit aber verschiedene grammatische Kategorien typologisch (unter ansonsten gleichen Meßbedingungen) unterschiedliche Meßerträge liefern, müßte durch entsprechende dialektometrische Untersuchungen geklärt werden.

2.3. Meßmatrix

Diese besteht aus einem durch die Atlaspunkte (= zeilenweise) und die Atlaskarten (= spaltenweise) aufgespannten *Merkmalsraum*. Taxometrischer Terminologie zufolge (Sodeur 1974, 9) spricht man auch von Elementen oder Objekten (= Atlaspunkte) und von Merkmalen oder Variablen (= Atlas-/Arbeitskarten). Wenn man qualitativ beurteilte Daten zum Zweck der taxometrischen Verdichtung in eine Matrix einsortiert, erfolgt damit der erste quantifizierende Schritt. Es erhebt sich dabei das meßtheoretische *Problem*, ob die Relationen der qualitativ aussortierten Taxate zueinander (empirisches Relativ) adäquat durch Relationen von Zahlen (numerisches Relativ) wiedergegeben werden, die diesen Taxaten zugeordnet werden. Es geht um die „homomorphe Abbildung eines empirischen Relativs in ein numerisches,, (Orth 1974, 17). Dabei ist zu entscheiden, welche numerischen Eigenschaften den Taxaten zuerkannt werden können. Es entsteht das Problem der Wahl der *Meßskala* (vgl. Cicourel 1970, Orth 1974, 26 f.; Schaich 1977, 4; Scheuch/Zehnpfennig 1974). Von den vier meistgenannten Meßskalen (Nominal-, Ordinal-, Intervall-, Verhältnisskala) kommen hierfür die Nominal- und die Intervallskala in Frage.

2.3.1. Meßmatrix auf Nominalskalenniveau

Atlaskarten (Merkmale)	3	i	i	i	i	i
	2	f	f	g	×	h
	1	a	b	c	d	e
		1	2	3	4	5

Atlaspunkte (Elemente)
× fehlendes Datum (Nullstelle)

Abb. 45.1: Meßmatrix auf Nominalskalenniveau (erfundener Datensatz)

Die Taxate werden als Träger einer Eigenschaft mit der Ausprägung 'voneinander verschieden' betrachtet. Sie sind sozusagen Träger verschiedener Namen (nomina), hier repräsentiert durch verschiedene Buchstaben. Aber auch Zahlen/Ziffern wären dafür verwendbar. Es handelt sich bei der Nominalskala um die am wenigsten eindeutige und zugleich meßtheoretisch niedrigste Skala. — Zu beachten ist noch der Eintrag eines feh-

lenden Datums, womit schematisch auf einen für die Dialektometrie zwar unangenehmen, aber praktisch kaum vermeidbaren Sachverhalt hingewiesen wird, nämlich jenen fehlender Daten in Sprachatlanten (vgl. 3.1., 3.1.2. und Vogel 1975, 77 f.).

2.3.2. Meßmatrix auf Intervallskalenniveau

Atlaskarten (Merkmale)	3	$\frac{5}{5}$	$\frac{5}{5}$	$\frac{5}{5}$	$\frac{5}{5}$	$\frac{5}{5}$
	2	$\frac{2}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	×	$\frac{1}{4}$
	1	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
		1	2	3	4	5

Atlaspunkte (Elemente)
× fehlendes Datum (Nullstelle)

Abb. 45.2: Meßmatrix auf Intervallskalenniveau (abgeleitet aus Abb. 45.1).

Jedes Taxat wird als Träger der Eigenschaft 'Raumanteil' (meßbar in Atlaspunkten) betrachtet, wobei dieser Raumanteil numerisch verschieden stark ausgeprägt sein kann. Unter dieser Voraussetzung kann die Nominalmatrix (vgl. Abb. 45.1) in eine metrische Matrix (auf Intervallskalenniveau) transformiert werden (vgl. Abb. 45.2). Man beachte, daß dabei die Raumanteilswerte (in Atlaskarte 1: 1; in Atlaskarte 2: 2 bzw. 1; in Atlaskarte 3: 5) zeilenweise relativiert werden, um solcherart den Nullstelleneffekt wenigstens teilweise aufzufangen. Intervallskalen sind definiert durch eine klar erkennbare Rangordnung von Meßwerten (hier: $1/5$, $1/4$, $2/4$, $5/5$) und durch Invarianz der Intervalle zwischen den Meßwerten (hier: natürliche Zahlen). Diese Transformation hat zur Voraussetzung, daß wir aus sprachgeographischer Sicht und Vorkenntnis bereit sind, den Atlaspunkten die meßbare Größe 'Raumanteil' als sprachwissenschaftlich relevantes Merkmal zuzuerkennen.

2.4. Q-Analyse vs. R-Analyse

Man spricht im Fall der Analyse von Zusammenhängen zwischen Elementen (hier von Atlaspunkten) üblicherweise von einer *Q-Analyse*, im Falle der Analyse von Zusammenhängen zwischen Merkmalen (hier von Atlas-/Arbeitskarten) von einer *R-Analyse* (Sneath/Sokal 1973, 114 f.; Sodeur 1974,

98). Gemäß den allgemeinen Postulaten der Sprachgeographie, die die Verteilung sprachlicher Phänomene im Raum, also über verschiedene Sprachatlaspunkte hinweg, untersucht, betreibt die Dialektometrie überwiegend Q-Analyse.

3. Q-analytische Meßmomente

3.1. Identitätstest (Berechnung des Relativen Identitätswerts — RIW; vgl. die Karten 45.1 und 45.2)

Auszugehen ist von der Nominalmatrix (vgl. Abb. 45.1). Ein Atlaspunkt wird als Prüfbezugs punkt vorgewählt. Von diesem sind zu den restlichen Atlaspunkten Paarvergleiche durchzuführen, wobei jedesmal ein vorher festzulegender Ähnlichkeitskoeffizient zur Verrechnung kommt. Bei n Atlaspunkten können also $n-1$ Paarvergleiche durchgeführt werden. Der hier vorgestellte *Ähnlichkeitskoeffizient* ist der einfachst mögliche (entspricht dem „simple matching coefficient“; vgl. Vogel 1975, 95 f.) und beruht auf der Division der Anzahl der taxatgleichen ($I_{x,y}$) durch die Anzahl der insgesamt verglichenen ($U_{x,y}$) Matrixzellen mit anschließender

Prozentuierung:

$$RIW_{x,y} = \frac{I_{x,y}}{U_{x,y}} \cdot 100.$$

Einsetzbeispiel nach Abb. 45.1: $RIW_{1,2}$: (a, f, i/b, f, i): $2/3 \cdot 100 = 66,6\%$; $RIW_{1,3}$: (a, f, i/c, g, i): $1/3 \cdot 100 = 33,3\%$. Die Anwendung anderer Ähnlichkeitskoeffizienten ist möglich und ist bei hybriden und/oder kleineren Datensätzen (Cowan 1964, Ellegård 1959, Houck 1967, Reed/Spicer 1952) oder bei sprachtypologischen Versuchen (Kroeber/Chrétien 1937, Muljačić 1967) schon mehrfach versucht worden. Neu ist allerdings die hier gezeigte Zusammenstellung aller durch paarweise Anwendung eines Ähnlichkeitskoeffizienten ermittelten Ähnlichkeitswerte zu einer empirischen Häufigkeitsverteilung, deren Weiterverarbeitung in Form einer statistischen Wertkarte (vgl. Karte 45.2) und die anschließende kartographische Behandlung nach dem Choroplethenprinzip (vgl. Karte 45.1). In weiterer Zukunft müßten nun die verschiedensten Ähnlichkeitskoeffizienten an verschiedenen Datensätzen ausprobiert werden, um deren Tauglichkeit für dialektometrische Untersuchungen zu prüfen (vgl. etwa Altmann 1978).

3.1.1. Oberitalienischer Datensatz (vgl. die Karten 45.1 und 45.2)

Maximale Matrixmaße: 251 Meßpunkte mal 696 Arbeitskarten (gezogen aus AIS Bd. I, II, IV). Von den 251 Meßpunkten entsprechen 247 originalen AIS-Atlaspunkten (zugleich die Nordhälfte des gesamten AIS-Punktfeldes von insgesamt 405 Atlaspunkten); 4 Stück sind als 'Kunstpunkte' zusätzlich eingesteuert worden (neben Punkt 999 = Hochitalienisch die Punkte 154, 262, 524). Die Meßmatrix ist mit 4836 nach lexikalischen, morphologischen und syntaktischen Kriterien isolierten Taxaten besetzt. Vom Prüfbezugs punkt 172 aus (im südlichen Piemont gelegen; vgl. Jaberg/Jud 1928, 59) können nulleintragsbedingt nur 693 (von insgesamt 696 verkodeten) Arbeitskarten verrechnet werden.

3.1.2. Kartographisches Beispiel (Karte 45.2, links unten)

Paarvergleich	172	→	176.
172 693	→	176 691	
100.000		80.029	
7—, G—, 7—		5+, F+, 6+	

Der Ausdruckblock bei Punkt 176 ist wie folgt zu interpretieren: „Von 693 vergleichbaren Arbeitskarten (bei Punkt 172 und am Kopf von Karte 45.2 vermerkt) sind 691 auch im Meßpunkt 176 nulleintragsfrei. Davon sind 80,029% (= 553 Stück) in der Prüfgabel zwischen dem Prüfbezugs punkt 172 und dem verglichenen Punkt 176 nominal identisch.. Die letzte Zeile des Ausdruckblocks gibt die Intervallzugehörigkeit (vgl. 3.1.3.) an. Punkt 176 gehört gemäß Intervallalgorithmus MINMWMAX zum Intervall 5+.

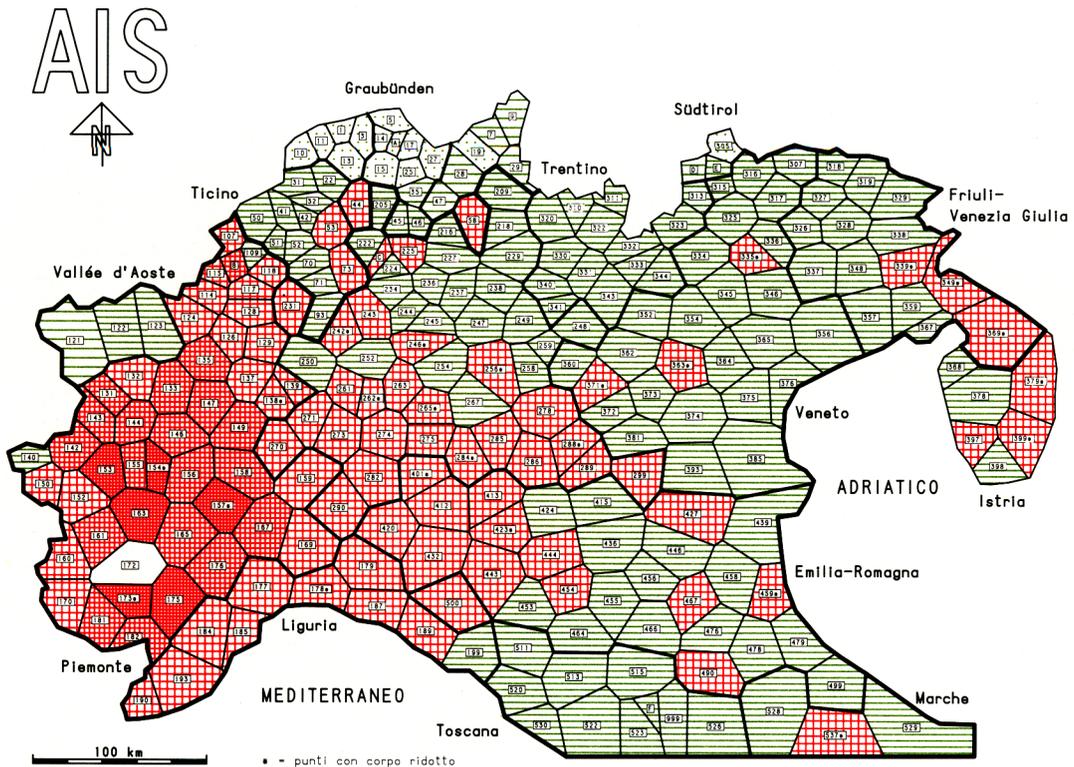
3.1.3. Intervallalgorithmus

Eine statistische Wertkarte und die darin enthaltene empirische Häufigkeitsverteilung führen allein noch nicht zu jener erwünschten Informationsverdichtung, bzw. Raumtypisierung. Das kann erst durch Erstellung einer Choroplethenkarte (oder einer anderen anschaulichen Darstellungsform; vgl. Arnberger 1977) geschehen. Damit wird eine *visuelle Typisierung* durchgeführt und man gelangt — je nach Art des dabei verwendeten Typisierungsverfahrens — zu einem abgestuften empirischen Typusbegriff (bezogen auf 'Dialekte, Sprachlandschaften', etc.) im Sinne von Hempel/Oppenheim (1936). —

Konkret geschieht dies so, daß charakteristische *Kennwerte* der Häufigkeitsverteilung bestimmt werden: Minimalwert, Mittelwert, Maximalwert (daneben auch Standardabweichung und Median). Dann wird die gewünschte *Intervallanzahl* festgelegt. Unter Berücksichtigung sechpsychologischer, arbeitspraktischer und numerischer Gründe fiel die Entscheidung hier auf 6 (bzw. 12) Intervalle. Es folgt die Festlegung der anzuwendenden *Intervallalgorithmen*. Es sind dies in dem Beispiel 3 Stück (MINMWMAX, MEDMW, MED), wobei im Fall des Intervallalgorithmus MINMWMAX (vgl. den Kopf der Karte 45.2) die Spannen zwischen Mittelwert und Minimalwert und zwischen Maximalwert und Mittelwert jeweils gedrittelt (für Sechsfachintervallisierung), bzw. geschestet (für Zwölfachintervallisierung) werden. Im Falle einer Sechsfachintervalleinteilung ergeben sich 7 Intervalleckwerte, die am Kopf der Karte 45.2, Zeile MINMWMAX, vermerkt sind. Der eingedrückt darunterliegende, nur ganze Zahlen enthaltende dreizeilige Ausdruckblock gibt an, wieviele Meßpunkte die einzelnen Intervalle enthalten. Im Falle von MINMWMAX (sechsfach) sind das bei Intervall 1 $11 + 3 = 14$ Meßpunkte, bei Intervall 2 $7 + 1 = 8$ Meßpunkte, etc. Anhand dieser Angaben läßt sich (etwa nach dem bei Schaich 1977, 17—19 vermerkten Vorgang) das dieser Häufigkeitsverteilung entsprechende Säulendiagramm (Histogramm) erstellen. — Die Entscheidung zwischen mehreren Intervallalgorithmen muß autoritativ getroffen werden und ist somit Teil des gesamten taxonomischen Prozesses. Doch wird man dabei so verfahren, daß die algorithmisch ermittelten Raumtypen den aus der klassischen Sprachgeographie bekannten Dialekträumen möglichst angenähert werden. (Zur Technik der statistischen Intervalleinteilung siehe Dickinson 1973, 82 f.; Jenks/Caspall 1971, Jenks/Coulson 1963, Kishimoto 1972; Pudlatz 1976, 16 f. Zur Erklärung der in Karte 45.2 neben MINMWMAX verwendeten Intervallalgorithmen MEDMW und MED vgl. Goebel 1978 a, 346 f.)

3.1.4. Romanistische Interpretation von Karte 45.1

Die Intervalle 6 und 5 geben das nähere piemontesische Umfeld von Punkt 172 wieder. Intervall 4 zeigt die Hauptvernetzungsrichtung der Dialektizität von Punkt 172 in das Nordwest- und Südostlombardische, sowie — gegen Westen — in das Gebiet des Alpin-



LEX - ITALIEN

999-T0T-172

Punkte mit Buchstaben:

- A - 16
- B - 116*
- C - 225*
- D - 312
- E - 314*
- F - 524

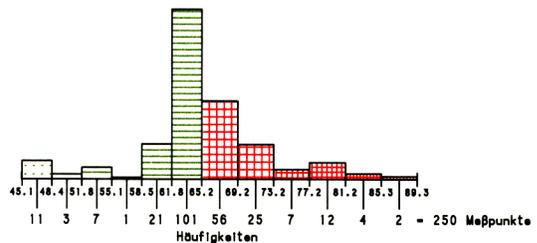


Kartographie: W.-D. Rose

Choroplethenkarte
MINMWMAX 6-fach

1	45.073 - 51.773
2	58.473
3	65.173
4	73.211
5	81.248
6	89.286

Häufigkeitsverteilung
MINMWMAX 12-fach



Karte 45.1: Identitätsprofil zum Prüfbezugspunkt 172 (Villafalletto, Provinz Cuneo) anhand von 693 Arbeitskarten nach AIS; zum numerischen Ertrag vgl. Karte 45.2 und 3.1. — Erstellung der Graurasterkarte (Choroplethenkarte MINMWMAX 6fach): anhand Intervallalgorithmus MINMWMAX 6fach (vgl. 3.1.2.) — Erstellung des Stabdiagramms (Häufigkeitsverteilung MINMWMAX 12fach): anhand Intervallalgorithmus MINMWMAX 12fach, entspricht Halbierung von MINMWMAX 6fach; Rechen- und Darstellungsmodus nach Schaich 1977, 17—19. Elektronische Kartographie: W.-D. Rase, Bonn

provenzalischen (Meßpunkte 170, 160, 152, 150). Vereinzelt Meßpunkte in Intervall 4 (v. a. im venezianischen Raum) verdanken ihren atypisch hohen RI-Wert den in ihren Merkmalsvektoren allzu häufigen Nulleinträgen. Unterdurchschnittliche Ähnlichkeiten (Intervalle 3,2,1) existieren gegenüber dem Frankoprovenzalischen des Aostatals, gegenüber gewissen Teilen des Alpinprovenzalischen (Punkt 140), dem Rätoromani-

schen Graubündens (wobei sich das Engadin als typologisch 'weicheres' Gebiet gegenüber West- und Mittelbünden abhebt) und Südtirols (Punkt 305), dem Friaulischen, dem Venezianischen und dem Toskanischen. Derartige punktbezogene Identitätskarten erlauben es, gewisse traditionelle Fragestellungen der Sprachgeographie („Die Stellung des Dialekts von X zwischen Y und Z,“) in neuem Licht zu sehen.

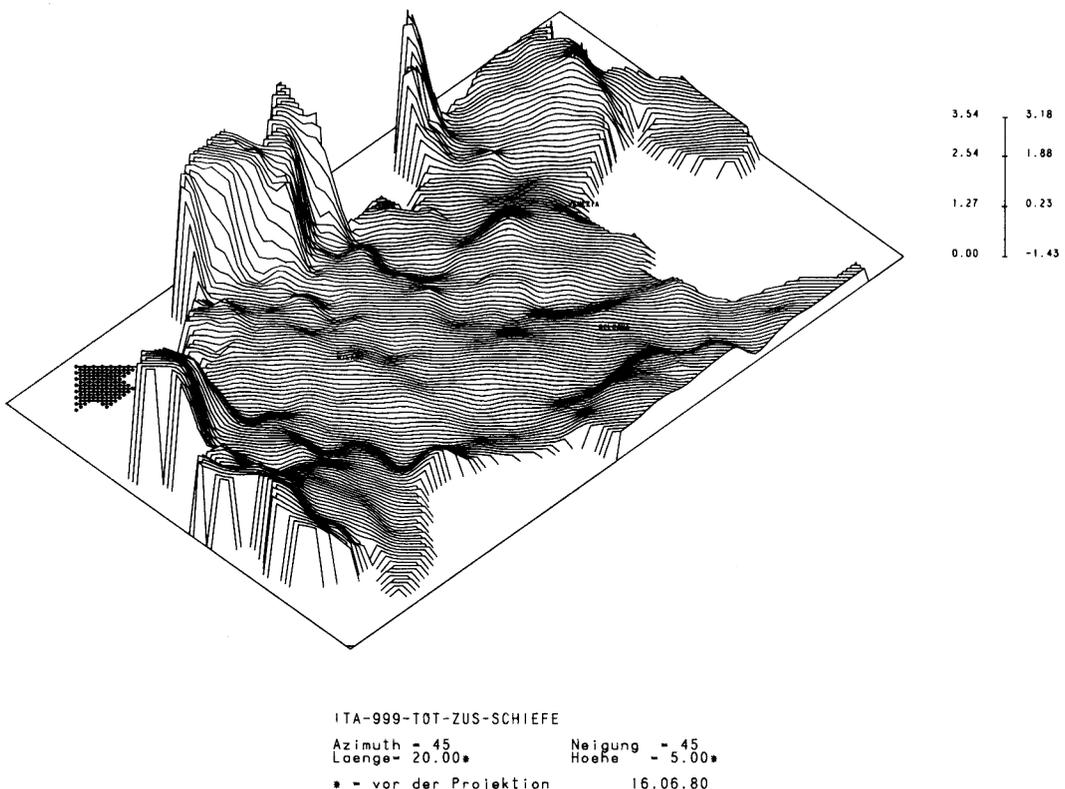
3.1.5. Heuristik der Identitätskarten

Prinzipiell sollen ebenso viele Identitätskarten errechnet werden, wie die Meßmatrix Atlaspunkte enthält. Wenn die vom Dialektometer unterhaltene kartographische Infrastruktur es erlaubt, können durch sukzessive Vorführung (mittels Trickfilmtechnik, Überblendprojektion, etc.; vgl. Goebel 1978 a, 335 f.) einander eng benachbarter Identitätsprofile sehr anschauliche Fließbildeffekte erzielt werden, die für den akademischen Unterricht und die wissenschaftliche Hypothesenbildung überaus ertragreich sind.

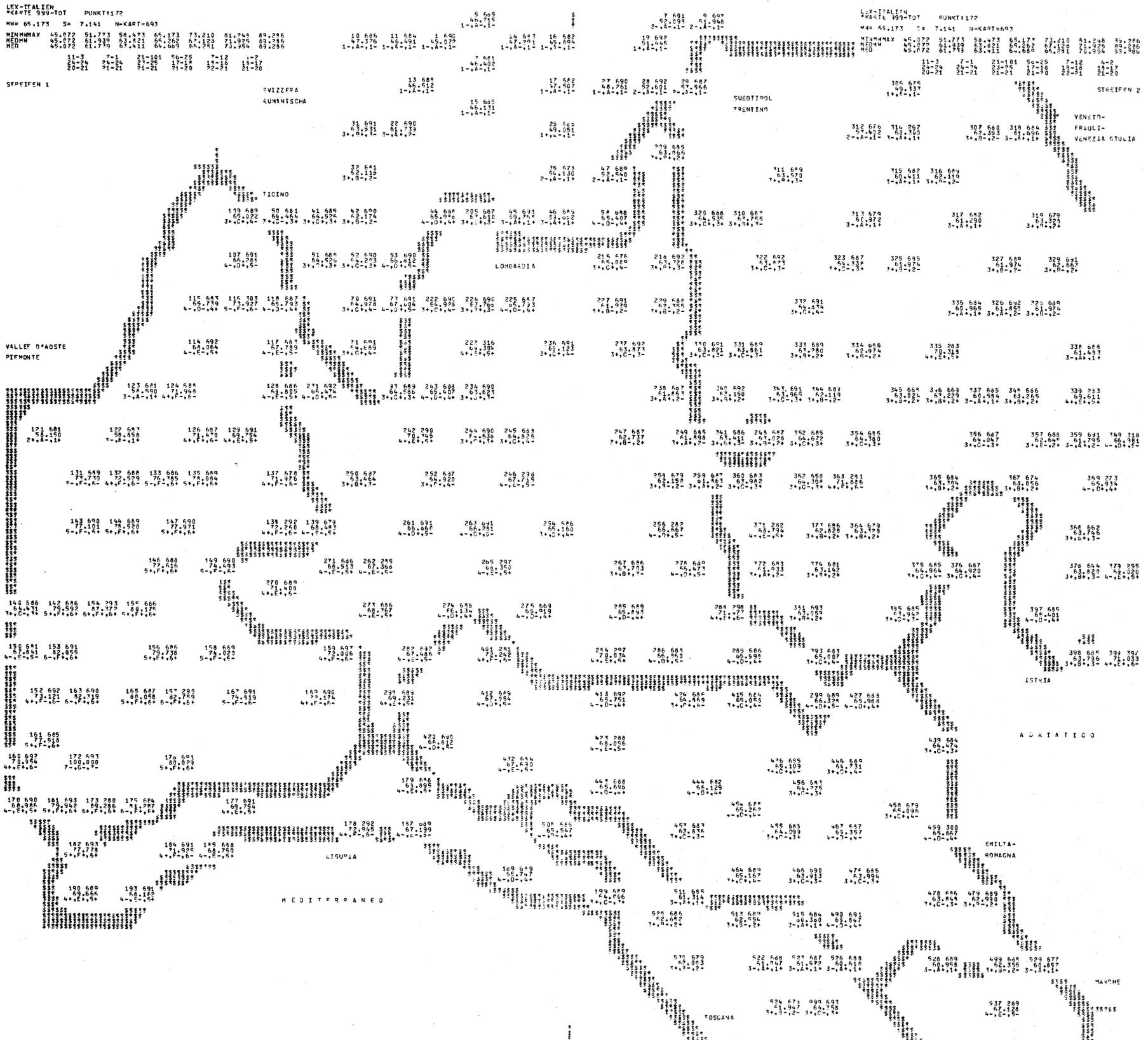
3.1.6. Fortführung des Identitätstests

Da bei n Meßpunkten n Identitätsverteilungen ermittelt werden können, von denen jede spezifische charakteristische Parameter (der Lage: Extremwerte, arithmetisches Mit-

tel, Median; der Streuung: Standardabweichung, Quantilabstände, etc.; der Symmetrie: Schiefe) besitzt, liegt es nahe, *Synopsen* dieser jeweils n Parameter zu erstellen. Besonders anschauliche Resultate liefern die Synopsen der Maximalwerte (zeigt Dialektkerne), der Standardabweichungen (zeigt überaus deutlich den identitätsspezifischen Durchmischungsgrad eines Untersuchungsgebiets, wobei sich zentrale und periphere Dialektstände graduell entflechten), der Schiefe und eines Quotienten aus Maximalwert und arithmetischem Mittel (MAX/\bar{x}) (ergibt eine in beiden Fällen 'automatische' Dialektgruppierung: systemzentrale Zonen haben niedrigere, systemperiphere Rand- und Einschlußgebiete hohe Meßwerte. Die taxometrische Wertigkeit der Schiefe ist dabei durch das Zusammenwirken von Vorzeichenvariation — positive vs. negative Schiefewerte — und Meßwertausprägung besonders hoch.). Aus Karte 45.3 ist die gute Ver-



Karte 45.3: Digitales Geländemodell einer numerischen Synopse von 251 Schiefewerten (Korpus wie bei Karte 45.1; vgl. 3.1.6. und 3.1.7.). Zu den 251 Identitätsverteilungen (zu je 250 Ähnlichkeitswerten -RIW) der (aus der Meß- oder Datenmatrix mittels Ähnlichkeitsmessung abgeleiteten; vgl. 3.1.) Ähnlichkeitsmatrix (vgl. 3.1.7.) wurden 251 Schiefen (Formel nach Bahrenberg/Giese 1975, 54) errechnet und computerkartographisch zu einer geglätteten statistischen Oberfläche synthetisiert (Betrachtung aus Südwesten unter einem Winkel von 45°). Elektronische Kartographie: W.-D. Rase, Bonn



Karte 45.2: Relative Identitätswerte (RIW) zum Prüfbezugs punkt 172 (Villafalletto, Provintz Cuneo) anhand von 693 Arbeitskarten nach AIS (vgl. 3.3. und Karte 45.1)

wendbarkeit dieser Klassifikation mittels Schiefe für die jetzt ein gutes Jahrhundert alte Frage ersichtlich, ob das Frankoprovenzalische, Alpinokitanische, Bündnerromanische, Ladinische und Friaulische 'italienische Dialekte' seien oder nicht. Die höchsten Gebirgserhebungen (positive Schiefe) in diesem Geländemodell entsprechen (im Uhrzeigersinn) dem Alpinprovenzalischen, dem Frankoprovenzalischen des Aostatales (und diesem südlich vorgelagerter Täler) und dem Rätoromanischen Graubündens, Ladinien sowie (weniger deutlich ausgeprägt) Friauls. Damit sind jene Bereiche des AIS erfaßt, in denen aus der Sicht der klassischen romanischen Sprachgeographie schon immer nicht-italienische Sprachstände angenommen worden waren. Demgegenüber sind die Meßwertminima (kleinste negative Schiefen) im Verlauf des Apennins einerseits und parallel zur Etsch andererseits anzutreffen, woraus sich eine eigenartige scherenförmige Gliederung des Zentrums dieses Untersuchungsgebietes ergibt. Auf der Karte 45.3 sind die fraglichen Depressionen aus Gründen der Bildsyntax weniger deutlich als die peripheren Elevationensichtbar. Die genannten Ergebnisse (vgl. dazu Goebel 1979 a, 163 f.) wurden rein empirisch gefunden, woraus mit Popper (1973, 389) einmal mehr deutlich wird, daß der „Gang der Wissenschaft (...) im Probieren, Irrtum und Weiterprobieren,, besteht.

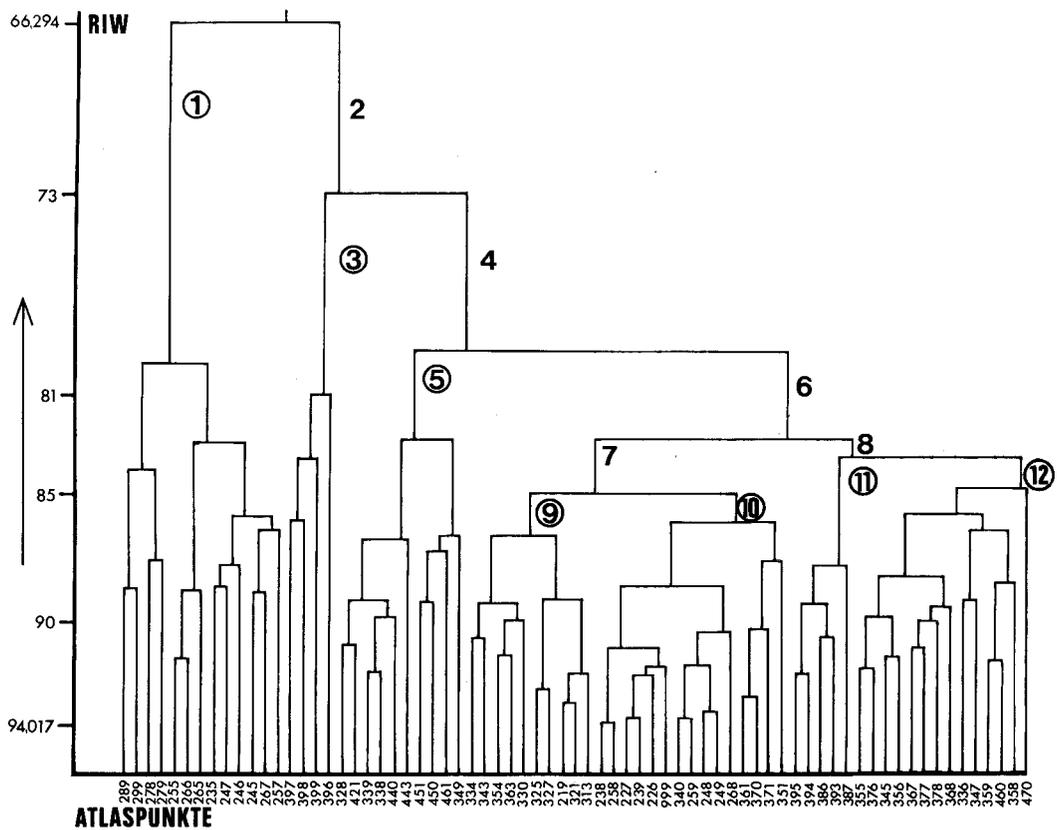
3.1.7. Symmetrische Ähnlichkeitsmatrix

Man kann eine symmetrische (i. e. quadratische) Matrix erstellen, deren Seiten jeweils aus den n Elementen (Meßpunkten) des Untersuchungsgebietes bestehen, so daß darin alle n Identitätskarten (vgl. 3.1.) enthalten sind und daher die RI-Werte aller Meßpunkte zu allen anderen abgelesen werden können. Die Diagonale dieser quadratischen Ähnlichkeitsmatrix enthält den RI-Wert 100%. Eine derartige Ähnlichkeitsmatrix kann unter Verwendung verschiedenster Ähnlichkeits-, Abstands- und Korrelationsmaße gebildet werden. Einzige Voraussetzung: das gewählte Maß sollte die Bedingungen einer Metrik erfüllen (Späth 1975, 11). Im Falle des RI-Werts trifft dies zu (vgl. Späth 1975, 15; Sodeur 1974, 77 f.). Diese symmetrische Ähnlichkeitsmatrix ist Grundlage der Clusteranalyse.

3.2. Clusteranalyse

Dieses sich heute großer Beliebtheit er-

freuende *multivariate Analyseverfahren* versucht, Elemente (hier: Atlaspunkte), die sich untereinander bezüglich eines Ähnlichkeits- oder Distanzmaßes (hier: RI-Wert) unterscheiden, in Klassen (Cluster, Gruppen, Haufen) so einzuteilen, daß die einer Klasse zugeteilten Elemente voneinander weniger differieren als gegenüber 'außen' (d. h. gegenüber Elementen, die außerhalb der Klasse liegen oder gegenüber anderen Klassen). Aus der großen Vielfalt der derzeit verwendeten clusteranalytischen Verfahren (vgl. Anderberg 1973, Bock 1974, Späth 1975, Vogel 1975) eignen sich die hierarchisch-agglomerativen Verfahren (Anderberg 1973, 132 f.; Bock 1974, 383 f.; Späth 1975, 162 f.; Vogel 1975, 249 f.) besonders gut für dialektometrische Klassifikationsaufgaben (Karten 45.4 und 45.5). Üblicherweise starten diese Verfahren den Prozeß der *Klassenbildung* mit n zunächst unverbunden dastehenden Elementen (hier mit 71 Atlaspunkten). Dann folgt der erste Klassifizierungsschritt, dessen Ziel und Ergebnis die Verschmelzung jener beiden Elemente zu einem Elementenpaar (Cluster Nr. 1) ist, die einander in bezug auf das verwendete Ähnlichkeitsmaß am ähnlichsten sind. Als Resultat dieser Fusion wird die Anzahl der für den zweiten Klassifizierungsschritt verbleibenden Elemente zunächst um 2 verringert, dann aber nach erfolgter Verschmelzung dieser zwei Elemente zu einem Cluster um den einen neugebildeten Cluster vermehrt, so daß de facto für den zweiten Klassifikationsschritt $n - 2 + 1 = n - 1$ Elemente und/oder Cluster verbleiben. Es sind insgesamt $n - 1$ Klassifikationsschritte möglich (hier 70), deren Resultat eine in Dendrogrammform ausgegebene Clusterhierarchie ist. Jeder Cluster liegt auf einem genau festgelegten Wertniveau. — Auf Karte 45.4 sind von den 70 berechneten Wertniveaus platzbedingt nur 2 auf drei Kommastellen genau und 4 auf ganze Zahlen gerundet angeführt. Die quantitativen Modalitäten der $n - 1$ Fusionen werden durch bestimmte vorgegebene numerische Kriterien gesteuert (hier: complete linkage; vgl. Anderson 1973, 138 f.; Bock 1974, 392 f.; Vogel 1975, 300 f.). Der Taxometer kommt dabei aber nicht um die Aufgabe herum, bei der Verclustering eines Datensatzes diese numerischen Kriterien — die ja stets die Ergebnisse in eine bestimmte Richtung lenken — experimentell an seine persönlichen klassifikatorischen Erwartungen anzupassen. — Das errechnete Dendrogramm (vgl. Karte 45.4) zeigt deutlich, daß sich die



Karte 45.4: Dendrogramm einer hierarchisch-agglomerativen Clusteranalyse (71 Atlaspunkte nach ALF) anhand RIW und 1468 Arbeitskarten nach ALF (vgl. 3.2.)

70 Cluster zwanglos und de visu in hierarchisch verschiedenrangige Großklassen zusammenfassen lassen. Das hier vorliegende Dendrogramm entspricht weitgehend der in der galloromanischen Dialektologie üblichen Dialektklassifikation dieses Gebiets.

3.2.1. Nordwestfranzösischer Datensatz (vgl. die Karten 45.4—45.7)

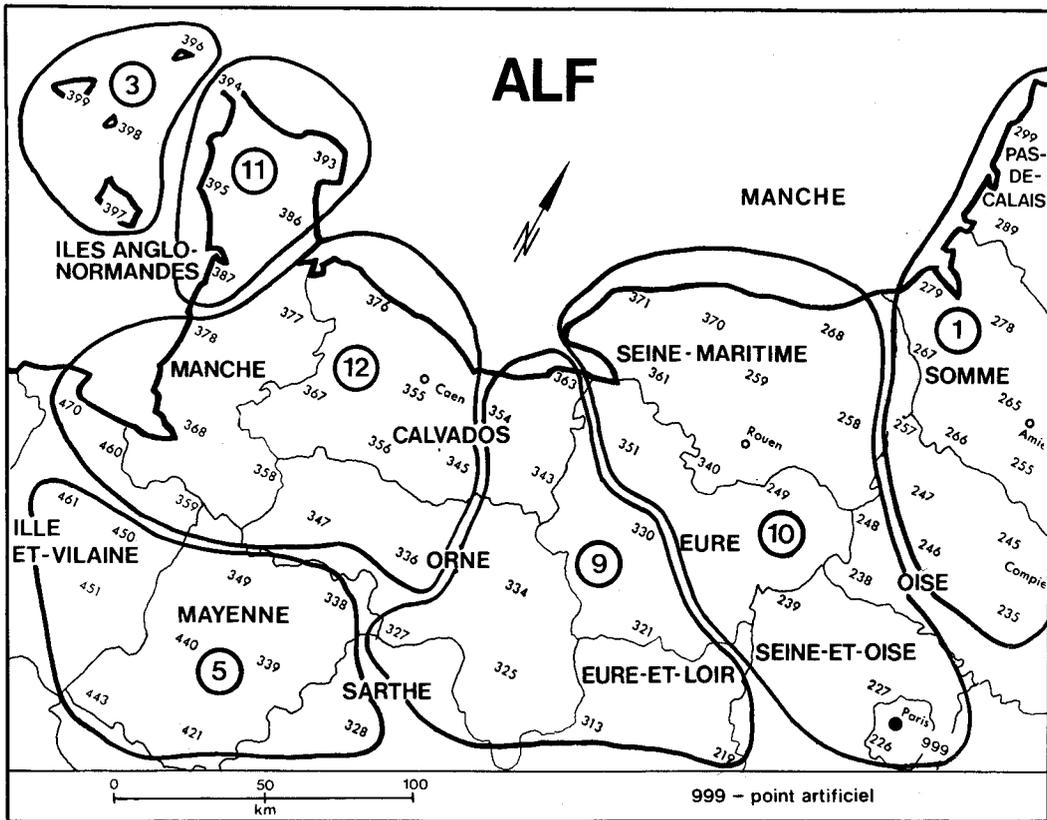
Maximale Matrixmaße: 71 Meßpunkte mal 1468 Arbeitskarten (gezogen aus ALF Bd. I—IX). Von den 71 Meßpunkten entsprechen 70 originalen ALF-Atlaspunkten (zugleich ein kleiner Ausschnitt im NW des gesamten ALF-Punktefeldes von 639 Atlaspunkten); Punkt 999 gibt das Hochfranzösische wieder. Die Meßmatrix ist mit 3959 nach lexikalischen, morphologischen und syntaktischen Kriterien isolierten Taxaten besetzt.

3.2.2. Symmetrische Ähnlichkeitsmatrix

Diese wurde nach den Prinzipien des Identitätstests (vgl. 3.1.) erstellt.

3.2.3. Romanistische Interpretation der Karten 45.4 und 45.5

Der durch unser Prüffenster sichtbare Teil Nordwestfrankreichs umfaßt herrschender *Lehre* zufolge das Westpikardische (Départements Pas-de-Calais, Somme, Oise), das Normandische (wozu die dialektal oft extrem liegenden anglonormandischen Inseln gezählt werden), Anteile des Westfranzösischen (Départements Ille-et-Vilaine, Mayenne, Sarthe) und zentralfranzösische Varietäten (vor allem im SO). Diesem Vorwissen entspricht der clusteranalytische Befund weitestgehend. — Der höchstrangige *Cluster* umfaßt (und trennt) die jeweils wohlseparierten Einzugsgebiete des Westpikardischen (Fläche 1) und des Westfranzösisch-Normandischen (Clusterast 2 auf Karte 45.4). In weiterer Folge fallen die anglonormandischen Inseln (Fläche 3) und das Westfranzösische (Fläche 5) aus. Als wohlseparierte Gruppen erscheinen ferner das Westnormandische (Flächen 11 und 12) und das Ostnormandische (Flächen 9 und 10), womit



Karte 45.5: Kartographische Darstellung des clusteranalytischen Dendrogramms (vgl. Karte 45.4 und 3.2.)

durchaus plausible und grosso modo in dieser Form akzeptable Ergebnisse geliefert werden.

3.3. Kohärenztest (Berechnung des Relativen Kohärenzmittels — RKM; vgl. die Karten 45.6 und 45.7)

Auszugehen ist von der Meßmatrix auf Intervallskalenniveau (vgl. Abb. 45.2). Ein Atlaspunkt wird vorgewählt, das arithmetische Mittel der über ihm in der entsprechenden Matrixspalte enthaltenen Meßwertserie ermittelt und auf Prozente gebracht. Der Vorgang wird für alle anderen Atlaspunkte wiederholt und das Resultat in kartographische Synopse gebracht. Beispiel: $RKM_1 = 100/3$. $(1/5 + 2/4 + 5/5) = 56,6\%$; $RKM_4 = 100/2$. $(5/5 + 1/5) = 60\%$. Eine derartige Synopse liegt in den Karten 45.6 und 45.7 vor.

3.3.1. Datensatz (vgl. die Karten 45.6 und 45.7)

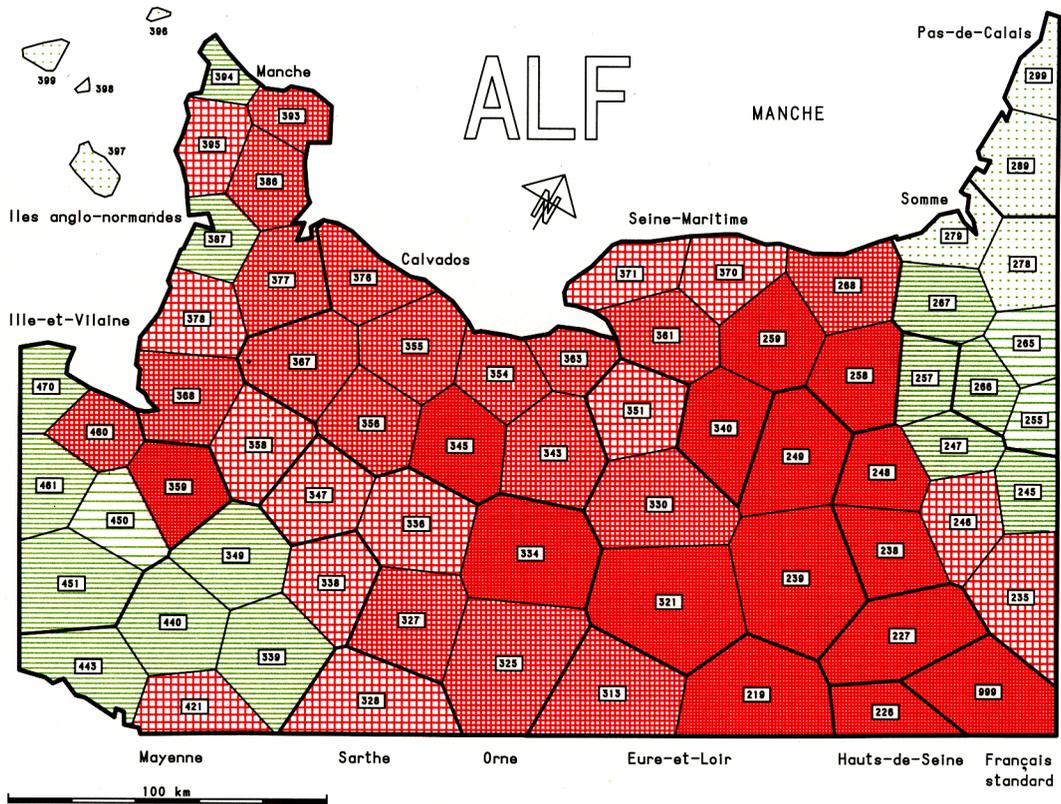
Wie in 3.2.1. beschrieben.

3.3.2. Intervallalgorithmus (Karte 45.6)

Wie in 3.1.3. beschrieben (MINMWMAX).

3.3.3. Romanistische Interpretation von Karte 45.6

Wiederum sind die dialektometrischen Meßergebnisse unter Rückgriff auf dialektologisches Vorwissen plausibel zu interpretieren. Man erkennt sofort, daß das Epizentrum des durch den RKM erfaßten raumgreifenden Dynamismus im SO, also um Paris liegt. Die von dort ausgehende Kohärenztendenz verflacht gegen NW und bricht sich deutlich an den schon auf den Karten 45.4 und 45.5 als extrem ausgewiesenen Dialektgebieten der anglonormandischen Inseln, des Westfranzösischen und des Westpikardischen. Damit ist ein innerhalb der nordfranzösischen Sprachgeschichtsschreibung immer wieder aufgegriffener Vorgang (Homogenisierung der galloromanischen Sprachlandschaften ab Paris) deutlich sichtbar gemacht worden.



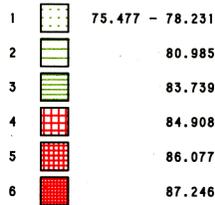
LEX - NORMANDIE

APB-TOT-RKM

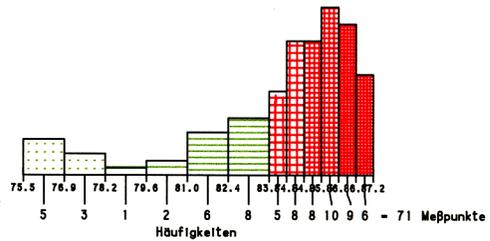


Kartographien W.-D. Rase

Choroplethenkarte
MINMWMAX 6-fach



Häufigkeitsverteilung
MINMWMAX 12-fach



Karte 45.6: Kohärenzprofil der Normandie anhand von 1468 Arbeitskarten nach ALF; zum numerischen Ertrag vgl. Karte 45.7 und 3.3. — Erstellung der Graurasterkarte (Choroplethenkarte MINMWMAX 6fach): anhand Intervallalgorithmus MINMWMAX 6fach (vgl. 3.1.2.) — Erstellung des Stabdiagramms (Häufigkeitsverteilung MINMWMAX 12fach): anhand Intervallalgorithmus MINMWMAX 12fach, entspricht Halbierung von MINMWMAX 6fach; Rechen- und Darstellungsmodus nach Schaich 1977, 17—19. Elektronische Kartographie: W.-D. Rase, Bonn

3.4. Q-analytische Meßmomente anhand der Meßmatrix auf Intervallskalenniveau

Der eigentliche *Mehrwert* der metrischen Meßmatrix gegenüber der nominalen besteht in der Möglichkeit der Durchführung paarweiser Atlaspunktvergleiche (wie unter 3.1.

beschrieben) unter Einsatz korrelations- und regressionsanalytischer Verfahren (Schaich 1977, 255 f.). Konkret bedeutet dies, daß (vgl. dazu die Abb. 45.1 und 45.2) im Paarvergleich z. B. der Atlaspunkte 2 und 3 nicht die nominalen Meßwertserien (b, f, i) und (c, g, i), sondern die metrischen Meßwertserien (1/5, 2/4, 5/5) und (1/5, 1/4, 5/5) ver-

4.1. Das Problem der 'semantischen Kategorien'

Sprachatlanen enthalten die Sprachkarten entweder in alphabetischer (Fall von ALF) oder in sachgruppenbezogener Ordnung (Fall von AIS). Man kann nun — eine entsprechende Indizierung vorausgesetzt — gewisse Q-analytische Meßmomente getrennt nach einzelnen *Sachgruppen* („semantischen Kategorien“; z. B. Wortschatz des Familienlebens, der Körperteile, von Handwerk und Handel, von Zahlen, Zeit und Raum, etc.) durchführen und die Resultate miteinander de visu oder korrelations- bzw. regressionsanalytisch vergleichen. In letzterem Fall ergibt sich wiederum die Möglichkeit, eine symmetrische Korrelationsmatrix (n Sachgruppen mal n Sachgruppen; jeweils auf der Basis eines Q-analytischen Meßmoments) zu erstellen und diese clusteranalytisch weiterzuverarbeiten. — Bislang angestellte de visu-Vergleiche haben eine relative *Invarianz* innerhalb der verglichenen Sachgruppenkorpora ergeben (Goebel 1976). Sollte man allerdings das in 2.2. angesprochene Problem der typologischen Valenz verschiedener grammatischer Kategorien hypothesengestützt untersuchen wollen, muß man darauf achten, pro Subkorpus eine gewisse Mindestanzahl von Arbeitskarten zur Verfügung zu haben und andererseits auch die Mächtigkeit der verschiedenen Subkorpora einander möglichst anzugleichen.

4.2. Das Problem der Korpusmächtigkeit

Alle taxometrischen Handbücher enthalten Angaben darüber, wieviele Merkmale man im Falle einer Q-Analyse mindestens berücksichtigen sollte, um zu verlässlichen *Resultaten* zu gelangen (Sneath/Sokal 1973, 106 f.). Seltener jedoch wird darauf hingewiesen (Vogel 1975, 50 f.), daß dieses Problem am ehesten dadurch untersucht werden kann, daß man die Meßmatrix ausgehend vom Vollstand teilweise reduziert (Zufallsprinzip beachten!) und dabei stets erneut das gewählte Meßmoment anwendet. — Im Falle des oberitalienischen Datensatzes wurden aus dem vollen Korpus (696 Arbeitskarten) nach dem Prinzip eines Urnenmodells mit Zurücklegung (Schaich 1977, 76) und unter Wahrung strenger Zufälligkeit bei der Auswahl *Teilkorpora* von 25, 50, 75, 100, 150, 200, 300 und 450 Arbeitskarten gezogen und jeweils mit mehreren Meßmomenten verrechnet. Dabei zeigt sich, daß sich die Grundzüge des Raumgliederungstypus bereits bei 25 Arbeitskarten einstellen, wäh-

rend eine weitgehende Stabilisierung in den Gliederungsdetails bei ca. 200 bis 300 Arbeitskarten erfolgt. Diese Versuche wurden anhand des nordwestfranzösischen Datensatzes wiederholt und dabei bestätigt. — Gegebenenfalls kann auch — wie in 4.1. beschrieben — eine korrelations- und anschließend clusteranalytische Untersuchung hierzu durchgeführt werden.

5. Literatur (in Auswahl)

AIS: Sprach- und Sachatlas Italiens und der Südschweiz. Hrsg. v. Karl Jaberg. Jakob Jud. Bd. I—VIII. Zofingen 1928—1940.

ALF: Atlas linguistique de la France. Hrsg. v. Jules Gilliéron. Edmond Edmont. Bd. I—IX. Paris 1902—1910.

Altmann 1978 = Gabriel Altmann: Zur Ähnlichkeitsmessung in der Dialektologie. In: Germanistische Linguistik 3—4/77, 305—310.

Altmann/Lehfeldt 1973 = Gabriel Altmann/Werner Lehfeldt: Allgemeine Sprachtypologie. Prinzipien und Meßverfahren. München 1973.

Anderberg 1973 = Michael R. Anderberg: Cluster analysis for applications. New York. San Francisco. London 1973.

Arnberger 1977 = Erik Arnberger: Thematische Kartographie. Mit einer Kurzeinführung über Automation in der thematischen Kartographie. Braunschweig 1977.

Bahrenberg/Giese 1975 = Gerhard Bahrenberg/Ernst Giese: Statistische Methoden und ihre Anwendungen in der Geographie. Stuttgart 1975.

Bock 1974 = Hans Hermann Bock: Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse). Göttingen 1974.

Cartwright 1953 = Darwin P. Cartwright: Analysis of qualitative material. In: Research methods in the behavioral sciences. Hrsg. v. Leon Festinger. Daniel Katz. New York. Chicago. San Francisco 1953. 2. unveränderte Auflage 1966, 421—470.

Cavalli-Sforza/Lorenz 1972 = Luigi Cavalli-Sforza: Biometrie. Grundzüge biologisch-medizinischer Statistik. Bearb. v. Rolf J. Lorenz. 3. durchgesehene Aufl. Stuttgart 1972.

Cicourel 1970 = Aaron V. Cicourel: Methode und Messung in der Soziologie. Übers. v. Frigga Haug. Frankfurt am Main 1970.

Cole/King 1968 = J. B. Cole/C. A. M. King: Quantitative geography. Techniques and theories in geography. London 1968.

Cowan 1964 = H. K. J. Cowan: Old limburgian and the quantitative classification of west germanic dialects. In: Leuvense Bijdragen 53. 1964, 1—21.

Dickinson 1973 = G. C. Dickinson: Statistical mapping and the presentation of statistics. 2. Aufl. London 1973.

Ellegård 1959 = Alvar Ellegård: Statistical measurement of linguistic relationship. In: *Language* 35. 1959, 131—156.

Engelien 1971 = Gerhard Engelien: Der Begriff der Klassifikation. Hamburg 1971 (Forschungsberichte des Institutes für Kommunikationsforschung und Phonetik der Universität Bonn 36).

Goebel 1976 = Hans Goebel: La dialectométrie appliquée à l'ALF (Normandie). In: XIV Congresso internazionale di linguistica e filologia romana (Neapel 1974). Atti. Neapel. Amsterdam 1976. Bd. II, 165—195.

Goebel 1977 = Hans Goebel: Rätoromanisch versus Hochitalienisch versus Oberitalienisch. Dialektometrische Beobachtungen innerhalb eines Diasystems. In: *Ladinia* 1. 1977, 39—71.

Goebel 1978 a = Hans Goebel: Zu Methoden und Problemen einiger dialektometrischer Meßverfahren. In: *Germanistische Linguistik* 3—4/77, 335—365.

Goebel 1978 b = Hans Goebel: Analyse dialectométrique de quelques points de l'AIS (italien standard, valdotain, provençal alpin, turinois, milanais). In: *Convegno internazionale di studio su lingue e dialetti nell'arco alpino occidentale: lo stato presente delle ricerche, diacronia e sincronia* (Turin 1976). Atti. Turin 1978, 282—294.

Goebel 1979 a = Hans Goebel: Dialektometrische Studien. Habil. (masch.). Regensburg 1979.

Goebel 1979 b = Hans Goebel: La méthode des interpoints appliquée à l'AIS (essai de dialectométrie). In: *Mélanges de philologie et de toponymie romanes offerts à Henri Guiter*. Montpellier 1979, 137—172.

Goebel 1980 = Hans Goebel: Dialektgeographie + Numerische Taxonomie = Dialektometrie. Anhand rätoromanischer und oberitalienischer Dialektmaterialien (AIS). In: *Ladinia* 4. 1980, 31—95.

Goebel 1981 = Hans Goebel: Isoglossen, Distanzen und Zwischenpunkte. Die dialektale Kammerung der Rätoromania und Oberitaliens aus dialektometrischer Sicht. In: *Ladinia* 5. 1981, 23—55.

Guiter 1973 = Henri Guiter: Atlas et frontières linguistiques. In: *Dialectes romans de France à la lumière des atlas régionaux* (Colloque du CNRS. Straßburg 1971). Hrsg. v. Georges Straka. Paris 1973, 61—109.

Haggett 1973 = Peter Haggett: Einführung in die kultur- und sozialgeographische Regionalanalyse. Übers. v. Dietrich Bartels/Barbara und Volker Kreibich. Berlin. New York 1973.

Hammond/McCullagh 1974 = Robert Hammond/Patrick McCullagh: Quantitative techniques in geography: an introduction. Oxford 1974.

Hempel/Oppenheim 1936 = Carl G. Hempel/Paul Oppenheim: Der Typusbegriff im Lichte der neu-

en Logik. Wissenschaftstheoretische Untersuchungen zur Konstitutionsforschung und Psychologie. Leiden 1936.

Houck 1967 = Charles L. Houck: A computerized statistical methodology for linguistic geography: a pilot study. In: *Folia linguistica* 1. 1967, 80—95.

Jaberg/Jud 1928 = Karl Jaberg/Jakob Jud: Der Sprachatlas als Forschungsinstrument. Kritische Grundlegung und Einführung in den Sprach- und Sachatlas Italiens und der Südschweiz. Halle 1928.

Jenks/Caspall 1971 = George F. Jenks/Fred C. Caspall: Error on choroplethic maps: definition, measurement, reduction. In: *Annals of the association of American geographers* 61. 1971, 217—244.

Jenks/Coulson 1963 = George F. Jenks/Michael R. C. Coulson: Class intervalls for statistical maps. In: *Internationales Jahrbuch für Kartographie* 3. 1963, 119—134.

Kishimoto 1972 = Haruko Kishimoto: Ein Beitrag zur Klassenbildung in statistischer Kartographie unter besonderer Berücksichtigung der maschinellen Herstellung von Choroplethenkarten. In: *Kartographische Nachrichten* 22. 1972, 224—239.

König 1974 = René König (Hrsg.): Handbuch der empirischen Sozialforschung. Bd. III a. III b. 3. Auflage. Stuttgart 1974.

Kroeber/Chrétien 1937 = A. L. Kroeber/C. D. Chrétien: Quantitative classification of indo-european languages. In: *Language* 13. 1937, 83—103.

Muljačić 1967 = Žarko Muljačić: Die Klassifikation der romanischen Sprachen. In: *Romanistisches Jahrbuch* 18. 1967, 23—37.

Orth 1974 = Bernhard Orth: Einführung in die Theorie des Messens. Stuttgart 1974.

Popper 1973 = Karl Raimund Popper: Objektive Erkenntnis. Ein evolutionärer Entwurf. Übers. v. Hermann Vetter. Hamburg 1973.

Pudlatz 1976 = Hilmar Pudlatz: GEOMAP. Ein FORTRAN-Programm zur Erzeugung von Choroplethen- und Isolinienkarten auf dem Schnelldrucker. Münster 1976 (Schriftenreihe des Rechenzentrums der Universität Münster 16).

Reed/Spicer 1952 = David W. Reed/John L. Spicer: Correlation methods of comparing idiolects in a transition area. In: *Language* 28. 1952, 348—359.

Schaich 1977 = Eberhard Schaich: Schätz- und Testmethoden für Sozialwissenschaftler. München 1977.

Scheuch 1974 = Erwin K. Scheuch: Auswahlverfahren in der Sozialforschung. In: Handbuch der empirischen Sozialforschung. Hrsg. v. René König. Stuttgart 1974. Bd. III a, 1—96.

Scheuch/Zehnpfennig 1974 = Erwin K. Scheuch/Helmut Zehnpfennig: Skalierungsverfahren in der Sozialforschung. In: Handbuch der empirischen Sozialforschung. Hrsg. v. René König. Stuttgart 1974. Bd. III a, 97—203.

Séguy 1971 = Jean Séguy: La relation entre la distance spatiale et la distance lexicale. In: *Revue de linguistique romane* 35. 1971, 335—357.

Séguy 1973 a = Jean Séguy: La dialectométrie dans l'Atlas linguistique de la Gascogne. In: *Revue de linguistique romane* 37. 1973, 1—24.

Séguy 1973 b = Jean Séguy: Atlas linguistique et ethnographique de la Gascogne. Bd. VI (1 + 2). Paris 1973.

Sneath/Sokal 1973 = Peter H. A. Sneath/Robert R. Sokal: Numerical taxonomy. The principles and practice of numerical classification. San Fran-

cisco 1973.

Sodeur 1974 = Wolfgang Sodeur: Empirische Verfahren zur Klassifikation. Stuttgart 1974.

Späth 1975 = Helmut Späth: Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion. München. Wien 1975.

Vogel 1975 = Friedrich Vogel: Probleme und Verfahren der numerischen Klassifikation. Unter besonderer Berücksichtigung von Alternativmerkmalen. Göttingen 1975.

Hans Goebel, Regensburg

46. Entwürfe zu dialektalen Informationssystemen

1. Einleitung
2. Der Begriff des Informationssystems
3. Datenaspekte dialektologischer Materialbearbeitung
4. Zur Systematisierung des DV-Einsatzes in der Dialektologie
5. Perspektiven dialektaler Informationssysteme
6. Literatur (in Auswahl)

1. Einleitung

Die explosive Entwicklung, die sich in der elektronischen Datenverarbeitung vor allem auch im nichtnumerischen Bereich vollzieht, hat dazu geführt, daß *computergestützte Forschungsmethoden* immer stärker auch in geisteswissenschaftliche Disziplinen eindringen. Dies gilt nicht zuletzt für die Dialektologie, die es, soweit sie sich als empirische Wissenschaft versteht, mit beträchtlichen Materialmengen zu tun hat und deswegen für einen weitreichenden DV-Einsatz geradezu prädestiniert ist.

Der vorliegende Beitrag will versuchen, eine Skizze der sich gegenwärtig abzeichnenden Möglichkeiten eines *Rechnereinsatzes* bei dialektologischen Materialbearbeitungen nachzuziehen und Perspektiven aufzuzeigen. Ausgegangen wird dabei von einer ansatzweisen Klärung des Informationssystembegriffs, von einer Analyse der Struktur und Bearbeitung dialektaler Ausgangsdaten und einem kursorischen Überblick über rechnergestützte Aktivitäten in der Dialektologie. Bei allen diesen Punkten sind starke inhaltliche Verkürzungen und Beschränkungen auf exemplarische Einzelphänomene unvermeidlich. Dies liegt zum einen an der hier gebotenen Kürze, zum anderen und ent-

scheidend aber an dem Umstand, daß die Thematik im Schnittpunkt zweier sich fremder Disziplinen liegt, der Dialektologie und der Informatik. Gerade bei letzterer ist in jüngster Zeit eine dynamische Entwicklung zu beobachten, in Verläufe derer selbst Grundbegriffe wie z. B. der des Informationssystems neue Konturen anzunehmen beginnen. Dies bedeutet für das hier zu behandelnde Thema, daß der Prozeß, mit dem die Informatik die empirischen Wissenschaften durchdringt, nicht problemlos zu fassen ist. Hinzu kommt, daß Computeranwendungen innerhalb der Dialektologie, aufs Ganze gesehen, bisher eher sporadisch und punktuell erfolgten; eine durchgängige Systematik läßt sich hier gegenwärtig noch nicht erkennen.

Für die weitergehende Behandlung des Themas hat dies zur Folge, daß sie sich nicht ausschließlich auf dem Boden erprobter und anerkannter Tatbestände bewegen kann, sondern auch über sie hinaus extrapolieren muß. Es wird versucht, aufgrund der erwähnten Erörterungen — besonders zur dialektalen Datenstruktur —, die Umriss eines *Systemrahmens* zu entwickeln, in den sich die wichtigsten Arbeiten einordnen lassen, aus dem heraus aber auch weitergehende Perspektiven sichtbar werden. Zielpunkt dabei ist die Definition eines dialektalen Informationssystems im Rahmen einer dafür zu entwickelnden Terminologie und deren Anwendung auf sich abzeichnende Weiterentwicklungsmöglichkeiten.

Dieses Vorgehen hängt mit der Spezifik dialektologischer *Forschungsinteressen* zusammen, die, soweit sie den empirischen Aspekt betreffen, durch Arealität und ausdrucksseitige Orientierung geprägt sind. Eine systematische Standortbestimmung im