

Quantitative Linguistik / Quantitative Linguistics

HSK 27



Handbücher zur Sprach- und Kommunikations- wissenschaft

Handbooks of Linguistics
and Communication Science

Manuels de linguistique et
des sciences de communication

Mitbegründet von Gerold Ungeheuer (†)
Mitherausgegeben 1985–2001 von Hugo Steger

Herausgegeben von / Edited by / Edités par
Herbert Ernst Wiegand

Band 27

Walter de Gruyter · Berlin · New York

Quantitative Linguistik Quantitative Linguistics

Ein internationales Handbuch
An International Handbook

Herausgegeben von / edited by
Reinhard Köhler · Gabriel Altmann
Rajmund G. Piotrowski

Walter de Gruyter · Berlin · New York

⊗ Gedruckt auf säurefreiem Papier, das die
US-ANSI-Norm über Haltbarkeit erfüllt.

ISBN-13: 978-3-11-015578-5

ISBN-10: 3-11-015578-8

Library of Congress Cataloging-in-Publication Data

Quantitative Linguistik : ein internationales Handbuch = Quantitative
linguistics : an international handbook / edited by Reinhard Köhler,
Gabriel Altmann, Rajmund G. Piotrowski.

p. cm. – (Handbücher zur Sprach- und Kommunikationswissen-
schaft = Handbooks of linguistics and communication science ; 27)
Parallel text in German and English.

Includes bibliographical references and indexes.

ISBN-13: 978-3-11-015578-5 (alk. paper)

ISBN-10: 3-11-015578-8 (alk. paper)

1. Mathematical linguistics. I. Title: Quantitative linguistics.
II. Köhler, Reinhard. III. Altmann, Gabriel. IV. Piotrowski, Ra-
eimon Genrikhovich. V. Handbücher zur Sprach- und Kommu-
nikationswissenschaft ; Bd. 27.

P138.Q36 2005

410.1'51—dc22

2005027011

Bibliografische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet
über <<http://dnb.ddb.de>> abrufbar.

© Copyright 2005 by Walter de Gruyter GmbH & Co. KG, D-10785 Berlin

Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der
engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das
gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und
Verarbeitung in elektronischen Systemen.

Printed in Germany

Satz: Meta Systems, Wustermark

Einbandgestaltung und Schutzumschlag: Rudolf Hübler, Berlin

Vorwort

Die quantitative Linguistik als wissenschaftliche Disziplin unterscheidet sich hinsichtlich ihres Gegenstands und Erkenntnisinteresses prinzipiell nicht von den anderen Forschungsrichtungen innerhalb der Sprachwissenschaft. Ihr Charakteristikum ist ein methodisches: die Einbeziehung sämtlicher mathematischer Hilfsmittel in ihr Instrumentarium, d. h. vor allem der quantitativen Methoden in Erweiterung und Ergänzung der in anderen Bereichen der Linguistik eingeführten und üblichen formalen Methoden (Logik, Mengentheorie, Algebra etc.), von einfachen, elementaren Verfahren wie dem Zählen von Einheiten und dem Feststellen von (metrisierten) Merkmalsausprägungen bis hin zu anspruchvollsten Modellbildungen.

Dennoch zeichnet sie sich in der Wissenschaftspraxis gegenüber formalen, qualitativen Ansätzen auch dadurch aus, dass sie aufgrund intensiver wissenschaftstheoretischer Reflexion den in den Naturwissenschaften etablierten, bewährten und in der Wissenschaftstheorie philosophisch abgesicherten Theorie- und Gesetzesbegriff übernommen hat. Dadurch steht sie auch methodologisch den Naturwissenschaften näher als die formalen Ansätze, die sich an nicht-empirischen, axiomatischen Systemen orientieren. In der theoretischen quantitativen Linguistik geht es entsprechend um die Konstruktion einer Sprachtheorie in Form nomologischer Netze, die zur Einlösung eines sprachwissenschaftlichen Erklärungsanspruchs zwingend erforderlich ist, während die Aufstellung ebenfalls meist „Theorien“ genannter axiomatischer Systeme in Form von Algebren oder mengentheoretischen Konstrukten nicht zu einer Erklärung empirischer Befunde zu führen vermag. Eine dem entsprechend zentrale Rolle kommen in der quantitativen Linguistik der deduktiven Ableitung von Hypothesen und dem exakten Testen dieser Hypothesen durch Gegenüberstellung mit geeigneten empirischen Daten zu.

Wie in allen anderen empirischen Wissenschaften nimmt in der Praxis der quantitativen Linguistik der Bereich der Theoriebildung und der Suche nach Gesetzen – gemessen an der insgesamt geleisteten Forschungsarbeit und der veröffentlichten Literatur – nur einen relativ geringen Raum ein. Der Versuch eines Überblicks in Form des vorliegenden Handbuchs durfte sich nicht auf die wissenschaftstheoretisch anspruchsvollsten Teilbereiche konzentrieren; vielmehr sollten ihre verschiedenen Aspekte ohne Wertung umfassend dokumentiert werden. Zu diesen Aspekten gehören insbesondere

1. die Metrisierung (Quantifizierung, Meßbarmachung) von sprachlichen Entitäten und Eigenschaften, also die Schaffung der Möglichkeit, aufgrund von Operationalisierung und Messung quantitative Daten aus sprachlichem Rohmaterial zu erzeugen,
2. die quantitative Analyse und Beschreibung sprachlicher und textueller Objekte,
3. die numerische Klassifikation sprachlicher und textueller Objekte zum Zweck weitergehender Untersuchungen oder für praktische Zwecke,
4. die Entwicklung und Anwendung statistischer Testverfahren zum diagnostischen Vergleich von sprachlichen und textuellen Objekten und zur Entdeckung von Tendenzen,
5. die Modellbildung mit Hilfe quantitativer Modelle und Methoden der Mathematik,

6. die Theoriebildung durch Suche nach universellen Sprach- und Textgesetzen und ihre Einbettung in ein umfassendes nomologisches Netz,
7. die Erklärung sprachlicher Erscheinungen mit Hilfe einer Theorie,
8. die Einbettung der Linguistik in ein allgemeines Gefüge der Wissenschaften, d.h. die Herstellung bzw. Aufdeckung interdisziplinärer Bezüge in Form von Generalisierungen, Analogien oder Spezialisierungen,
9. die Ausarbeitung einer spezifisch linguistischen Methodologie unter Berücksichtigung der besonderen Eigenheiten des linguistischen Untersuchungsgegenstands,
10. praktische Anwendungen in verschiedenen Bereichen, u. a. in Lehr- und Lernzusammenhängen, Psychologie/Psycholinguistik/Psychiatrie, Stilistik/Forensik, Computerlinguistik und Sprachtechnologie, Dokumentationswissenschaften, Inhaltsanalyse, Sprachplanung, Medienwissenschaften u. v. m.

Mit diesem Handbuch soll ein möglichst umfassender Überblick über den Gegenstand, die Geschichte, die wissenschaftstheoretischen Voraussetzungen und den Stand der Forschung der quantitativen Linguistik gegeben werden. In einem einleitenden Kapitel (Kapitel I) werden allgemeine Informationen zum Gegenstand, zu den wissenschaftstheoretischen Grundlagen und zur Geschichte der quantitativen Linguistik mit Einzeldarstellungen von Werken hervorragender Forscher (G. K. Zipf und W. Fucks) sowie Übersichtsartikel über die Forschungsarbeiten in ausgewählten Ländern zusammengefasst. Die Kapitel II bis X widmen sich den wichtigsten Gebieten und Phänomenen, geordnet nach linguistischen Analyseebenen und Forschungsansätzen: Phonologie (Kapitel II), Morphologie (Kapitel III), Syntax (Kapitel IV), Lexikologie und Lexikographie (Kapitel V), Text (Kapitel VI), Semantik und Pragmatik (Kapitel VII), Geolinguistik und Dialektologie (Kapitel VIII), Typologie (Kapitel IX), und Diachronie (Kapitel X). In Kapitel XI wird die Arbeitsweise der quantitativen Linguistik exemplarisch an einigen der wichtigsten Modelle, Hypothesen und Gesetze verdeutlicht. Kapitel XII stellt Bezüge der quantitativen Linguistik zu anderen wissenschaftlichen Disziplinen vor und präsentiert Beispiele für Gebiete, in denen Methoden und Modelle der quantitativen Linguistik oder Resultate wie Daten oder Verfahren angewendet werden. Das abschließende Kapitel XIII bietet einen Überblick über die wichtigsten Aspekte der Organisation der quantitativen Linguistik als Fach und über ihre Hilfsmittel (Bibliographien, Publikationsforen etc.). Alle Artikel schließen mit einer Auswahlbibliographie. Am Ende des Bandes findet sich ein Personen- und Sachregister.

Einige Gebiete und Forschungsrichtungen (wie Soziolinguistik und Sprachvariation, Psycholinguistik, Inhaltsanalyse u. a.) wurden in dem vorliegenden Handbuch nicht berücksichtigt, obwohl sie – zumindest in Teilbereichen – aus methodologischer Sicht mit gleichem Recht wie die hier aufgenommenen zum Gegenstand dieses Handbuchs hätten gezählt werden können. Ein Grund dafür liegt darin, dass einigen ihnen bereits eigene Handbücher gewidmet wurden oder eine andere Form ausführlicher Dokumentation ihres Forschungsstands vorliegt oder in Vorbereitung ist.

Adressaten des Handbuchs sind neben Forschern, Lehrenden und Studierenden aller Zweige der Sprachwissenschaften einschließlich der Philologien auch Wissenschaftler benachbarter Fächer, deren theoretische und empirische Forschung sprachwissenschaftliche Fragen berührt (wie Psychologie und Soziologie) bzw. die bewährte Methoden oder Resultate aus der quantitativen Linguistik für die eigenen Fragestellungen nutzbar machen wollen (wie die Musikologie).

Für die meisten Artikel des Handbuchs konnten die weltweit angesehensten Fachleute aus Europa, Asien und Amerika als Autorinnen und Autoren gewonnen werden. Wie es wohl unvermeidlich im Zusammenhang mit solchen großen Projekten ist, schieden im Laufe der Vorbereitungen aus verschiedensten Gründen einige der ursprünglich vorgesehenen Autoren aus. Denen, die freundlicherweise (zum Teil bereits zu einem späten Zeitpunkt) vakant gewordene Artikel übernahmen und unter Zeitdruck arbeiten mussten, sei an dieser Stelle besonders gedankt. Auch denen, die ihre Beiträge termingerecht eingereicht hatten und sie später aktualisieren mussten, gebührt Dank, vor allen denen, die uns Geduld und Verständnis entgegenbrachten. Ebenso möchten wir dem Reihenherausgeber, Herbert Ernst Wiegand, sowie Barbara Karlson und Monika Wendland vom Verlag Walter de Gruyter für die gute Zusammenarbeit danken.

Oktober 2004

Reinhard Köhler,
Gabriel Altmann,
Rajmund G. Piotrowski

Preface

Quantitative Linguistics as a scientific discipline actually does not differ substantially from common research in linguistics in terms of its subject nor in its main issues. Its characteristic is rather a methodological one: the scope of its instruments includes all of the mathematical tools, i. e. especially quantitative methods expanding and supplying the formal methods introduced and practised in other areas of linguistics (logic, set theory, algebra etc) ranging from the most simple elementary procedures like counting of units and recording of (metric) forms of features up to the most sophisticated modelling.

In fact, compared with formal approaches in the qualitative paradigm, its particular advantage in scientific work are the concepts of theory and law imported from the natural sciences based on careful epistemic consideration: concepts that have been well-established and proven there, capable of referring to epistemology for philosophical foundation. Thereby it is closer related to the natural sciences also in respect of its methodological framework than those formal approaches that refer to non-empirical, axiomatic systems. Accordingly theoretical quantitative linguistics is concerned with the construction of a theory of language in terms of nomological nets which is inevitably necessary to meet the linguistic requirements of explanation; while claiming axiomatic systems in terms of algebraic or set theoretical constructs, though often called “theories”, is not able to provide explanation of empirical findings. Correspondingly, deductive derivation of hypotheses just as the exact testing of these hypotheses by comparing them to the appropriate data is of central importance in quantitative linguistics.

Like in any other empirical science, however, the field of theory construction and search for laws does not take up a lot of space in scientific practice of quantitative linguistics – relative to the total amount of research work and publications. Trying to give a survey with the present handbook by concentrating on the epistemologically most sophisticated parts would have been unfair; instead it was rather aimed to cover extensively all of its various aspects without making value judgements. In particular, these aspects include

1. metricizing (scaling, quantifying, making measurable, or quantitation, as M. Bunge calls it) of linguistic entities and qualities, and thereby providing the possibility of generating quantitative data from speech material based on operationalisation and measurement,
2. quantitative analysis and description of linguistic and textual objects,
3. numerical classification of linguistic and textual objects for the purpose of further investigation or for practical reasons,
4. development and application of statistical test procedures for diagnostic comparison of linguistic and textual objects and for trend detection,
5. modelling of linguistic structures, functions, and processes by means of quantitative models and mathematical methods,
6. theory construction by searching for universal laws of language and text and their embedding into an extensive nomological net,
7. explanation of linguistic phenomena (properties, structures, processes) by means of a theory,

8. embedding of linguistics into a general system of sciences, i.e. establishing resp. exploring interdisciplinary relations in the shape of generalization, analogy or specification,
9. elaborating a genuine linguistic methodology with regard to the particular characteristics of the linguistic subject,
10. practical applications to various areas such as those in contexts of learning and teaching, psychology/psycholinguistics/psychiatry, stylistics/forensics, computational linguistics and language technology, documentation science, content analysis, language planning, mass communication research and more

This handbook is aimed to provide a maximum comprehensive survey of subject, history, epistemological conditions and current state of research of quantitative linguistics. In an introductory chapter (chapter I) general information is summarized concerning object of research, epistemologic foundations and history of quantitative linguistics including individual accounts on the works of outstanding researchers (G. K. Zipf and W. Fucks) as well as survey articles on the research work in selected countries. Chapters II to X attend to the central areas and phenomena, sorted by levels of linguistic analysis and research frameworks: phonology (chapter II), morphology (chapter III), syntax (chapter IV), lexicology and lexicography (chapter V), text (chapter VI), semantics and pragmatics (chapter VII), geolinguistics and dialectology (chapter VIII), typology (chapter IX), and diachronics (chapter X). In chapter XI, the technique of quantitative linguistics will be illustrated exemplarily by some of the main models, hypotheses and laws. Chapter XII establishes the relations of quantitative linguistics to other scientific disciplines and presents areas where methods and models of quantitative linguistics or results like data or procedures are applied. The final chapter XIII provides a survey on the most important aspects of the structure of quantitative linguistics as a discipline and its tools (bibliographies, places of publication etc). Each article provides selected references. The remainder of the volume presents an index of persons and subjects.

Some fields and branches of research (as sociolinguistics and language variation, psycholinguistics, content analysis and others) are not regarded in the present handbook, though they – at least in component parts – would have deserved to be counted as belonging to the topic of this book as rightly as the ones included here. One reason for this is that some of them are already attended in proper handbooks or that some other kind of/inlineo comprehensive coverage of their results is otherwise available or in preparation.

The handbook is addressed not only to research workers, teachers and students of all areas of linguistics including philology but also to scientists of closely related fields, whose theoretical and empirical investigations are concerned with linguistic questions (as psychology and sociology) or attempt to make use of proven methods or of results from quantitative linguistics for their own issues (e. g., musicology).

For the majority of contributions in this handbook we could gain for authorship the most respected international experts with the highest reputation from Europe, Asia and America. As seems to be inevitably joined with such extensive projects, in the course of preparation some of the initially designated authors were lost for many different reasons. At this point, we wish to express our special appreciation to those who kindly took over (partly even at a later time) vacant contributions and had to work under pressure. Acknowledgements are also due to those who had submitted

their contributions on time and had to update them later, particularly those who met us with patience and understanding. We also wish to express our thanks to the series editor, Herbert Ernst Wiegand, and to Barbara Karlson and Monika Wendland of Walter de Gruyter publishing company for good cooperation.

Oktober 2004

Reinhard Köhler,
Gabriel Altmann,
Rajmund G. Piotrowski

VIII. Gebiete und Phänomene: Geolinguistik und Dialektologie / Fields and phenomena: geolinguistics and dialectology

37. Dialektometrie

1. Dialektometrie: Typodiagnose anhand sprachgeographischer Daten
2. Methodische Voraussetzungen: vom Sprachatlas zur Datenmatrix
3. Meßmomente
4. Nachbemerkung
5. Literatur (in Auswahl)

1. Dialektometrie: Typodiagnose anhand sprachgeographischer Daten

Im Jahr 1973 kreierte der Tolosaner Romanist Jean Séguy (1973) den Terminus „dialectométrie“, worunter ab ovo ein numerisch-mathematisches Verfahren zur Auswertung und Komprimierung von in Sprachatlanten enthaltenen geolinguistischen Daten zu verstehen war. *Forschungsintention* (= Datensynthese, Datenklassifikation, Typodiagnose, Mustererkennung etc.) und *Struktur der Daten* (= Sprachatlanten) rücken die Dialektometrie in unmittelbare Nähe anderer *Metrien* (wie. z. B. *Soziometrie*, *Psychometrie*, *Ökonometrie*, *Anthropometrie* etc.) oder sonstiger quantifizierender Disziplinen (wie z. B. der quantitativ arbeitenden Geographie), wo überall – freilich in methodisch und methodologisch meist um vieles anspruchsvollerer Form – ein empirisch erhobener Datensatz aus N Elementen (oder Dingen) und p Merkmalen (oder Attributen, Eigenschaften etc.) auf konstitutive Ordnungsmuster verschiedenen Ranges untersucht wird. „In vielen wissenschaftlichen Disziplinen gibt es folgendes Problem: gegeben ist eine Menge von Objekten. Jedes dieser Objekte ist gekennzeichnet durch die Ausprägungen einer bestimmten Anzahl von Merkmalen. Ist es möglich, diese Objekte so zu klassifizieren, daß die einer Klasse zugehörenden Objekte einander in einem bestimmten, noch näher zu definierenden Sinne möglichst ‚ähnlich‘ sind, während gleichzeitig die verschiedenen Klassen zugehörenden Objekte einander möglichst ‚unähnlich‘ sind?“ (Vogel 1975, 1).

Heute steht zur Lösung derartiger Probleme eine numerische Disziplin zur Verfügung, die unter den verschiedensten Namen auftritt: numerical classification (Sneath/Sokal 1973), automatische Klassifikation (Bock 1974), numerische Klassifikation (Vogel 1975), Taxometrie, analyse typologique (Chandon/Pinson 1981), analyse des données (Benzécri 1980) etc. Es existieren hierzu Lehr- und Handbücher in allen gängigen Großsprachen sowie – was besonders unterstrichen werden soll – weltweit kooperierende Klassifikationsgesellschaften (dazu Dachverband seit 1986: International Federation of Classification Societies) mit einer entsprechenden Publikations- und Kongreßtätigkeit. Leider haben Séguy und die Mehrzahl der ihm nachfolgenden Dialektometer den Kontakt zur numerischen Klassifikation weder gesucht noch gefunden, so daß viele der von ihnen gesetzten Bemühungen methodisch und methodologisch höchst unscharf geblieben sind. Es soll daher mit allem Nachdruck betont werden, daß es heutzutage einfach unverantwortlich und anachronistisch wäre, Dialektometrie ohne engste Anlehnung an die allgemeine Entwicklung der numerischen Klassifikation zu betreiben. Überdies sind numerisch-taxonomische Verfahren bereits 1973 erstmals explizit von G. Altmann und W. Lehfeldt zur Bearbeitung *sprachtypologischer* Fragestellungen herangezogen worden.

Jedes typologische Diagnostizieren kennt verschiedene Rang- und Erkenntnisstufen. Es gibt mehr oder weniger *globale* Typodiagnosen sowie Typodiagnosen mit und zu *speziellen* Zielsetzungen. Wichtig ist es stets, daß der Typodiagnostiker sich seines *Erkenntniszieles* voll bewußt ist. Im Rahmen der sprachatlasgestützten Sprachgeographie sind diese Erkenntnisziele allerdings wegen der methodisch eher hemdsärmeligen Vorgeschichte der Sprachgeographie nur vage formulierbar. Es geht dort grosso modo um die Erkenntnis geolinguistischer *Gliederungen*, die Herausarbeitung sprachgeographi-

scher *Raumstrukturen* oder die Problematik dialektgeographischer Abstufungen, alles freilich nur anhand von *einzelnen* Atlaskarten. Damit lag und liegt der Interessenschwerpunkt der traditionellen Sprachgeographie eher im Bereich *partikulärer* als *synthetischer* Sehweisen. Dieser Umstand war und ist für die Entwicklung der Dialektometrie keineswegs förderlich.

Bibliographische Übersichten zur internationalen Dialektometrie: Goebel 1993a, 277–278 und CLAE 2, 9.0. Appendix, ferner Viereck 1980 und 1984 sowie Schneider 1984.

2. Methodische Voraussetzungen: vom Sprachatlas zur Datenmatrix

In Anbetracht verschiedentlich aufgetauchter Mißverständnisse rund um die Dialektometrie formuliere ich die Sache prägnant: *extra atlantes linguisticos nulla salus dialectometrica* [„Außerhalb von Sprachatlanten gibt es kein dialektometrisches Heil.“]. Die Dialektometrie steht und fällt mit ihrer Datenbasis. Und hierfür kommen nur *Sprachatlanten* in Frage, die durch die Struktur ihrer Daten (N Meßpunkte, p Atlaskarten) und die Art von deren Erhebung (N standardisierte Interviews mit jeweils p Fragen aus einem feststehenden Fragebuch) die Gewähr dafür bieten, daß bei der vergleichenden Auswertung „nur Vergleichbares miteinander verglichen“ wird. Unter „Sprachatlas“ soll hier jede nach dem N -mal- p -Prinzip erhobene und organisierte geolinguistische Datenkompilation verstanden werden. Die strikte Anwendung des Atlasprinzips bei der Publikation soll dabei aber keine Rolle spielen. Auch die in der Romanistik gut bekannten „Tableaux phonétiques des patois suisses romands“ (TPPSR) gelten demnach als vollwertiger Sprachatlas. Dies betrifft überdies auch viele andere geolinguistische Erhebungen, die vorwiegend in den ersten Jahrzehnten des 20. Jahrhunderts durchgeführt und nur in *Tabellen-* statt in *Atlasform* veröffentlicht wurden.

2.1. Dialektometrisches Messen bzw. Abbilden

Wer Dialektometrie betreiben will, muß messen. Wer mißt, bildet ab. Und wer abbildet, reduziert, vereinfacht, faßt zusammen – und irrt. Dies gilt sowohl für die empirische Sozialforschung als auch für die Satellitenphotographie. Sprachatlanten sind *Abbil-*

dungen (und damit *Reduktionen*) der unter einem bestimmten Blickwinkel (diatopische Sprachvariation) gesehenen sprachlichen *Realität*. Dieses sprachatlasgebundene Abbild muß, um dialektometrisch faßbar zu werden, weiter reduziert werden. Mit anderen Worten: es muß die jeder Sprachatlas-karte innewohnende Information nach einem bestimmten *Reduktions-* oder *Abbildungsprinzip* zusammengefaßt werden. Dieser Vorgang ist seit dem Erscheinen der ersten (romanischen) Sprachatlanten den Linguisten durchaus vertraut. Es handelt sich dabei um das „Typisieren“ von Sprachatlas-karten mit Hilfe von stummen Karten. Zahllose konkrete Anwendungsfälle finden sich hierzu im Bereich der romanischen Onomasiologie und Wortforschung.

Ziel dieser Reduzierung ist die Erstellung der Datenmatrix, die aus p Zeilen und N Spalten besteht: siehe Abb. 37.1, linke Hälfte. Dabei wird unter Ansetzung eines bestimmten Meßprinzips (Messung der *phonetischen* und/oder *morphosyntaktischen* und/oder *lexikalischen* Variation) pro Atlas-karte (mit N Belegen) eine bestimmte Anzahl von Belegtypen (genauer Taxat-types) ermittelt, die zumeist kleiner als N ist, theoretisch aber alle Werte zwischen 1 und N annehmen kann. Aus meßtheoretischer Sicht handelt es sich dabei um eine Messung auf der *Kardinal-*, *Nominal-* oder *Kategorialskala*: cf. dazu Sodeur 1974, 40 f. oder Orth 1974, 24 f. Dabei wird vorausgesetzt, daß die gemessenen Einheiten (Taxate bzw. Taxat-types oder -tokens) voneinander gleichrangig verschieden sind. Diese Messung ist freilich sehr grob. Allerdings sind höher-rangige Messungen an Sprachatlasdaten (auf der *Ordinalskala* bzw. auf *metrischen* Skalen) nur sehr schwer möglich.

2.2. Exkurs: Zur quantitativen Struktur dialektometrischer Daten

Bei der nominalen Vermessung der Karten eines Sprachatlases können – stets vorausgesetzt, daß keine Mehrfachbelege erfaßt werden (müssen) – bei N Meßpunkten pro Atlaskarte theoretisch zwischen 1 und N Taxat-types auftreten. Bei 1, 2, 3 (etc.) Taxat-types pro Atlaskarte spricht man von *mono-*, *di-*, *trinymer* (etc.) Taxierung. Die Abb. 37.5 (im Anhang) zeigt die entsprechenden Taxierungsbilanzen für Italien (anhand von 696 Arbeitskarten aus dem AIS) und Frankreich (anhand von 1687 Arbeitskarten aus dem ALF). Die jeweils erfaßten Variationen sind beim AIS nur *lexikalisch* und *morpho-*

syntaktisch, beim ALF dagegen *phonetisch*, *morphosyntaktisch* und *lexikalisch*.

Bilanzen dieser Art hängen natürlich von den folgenden zwei Grundfaktoren ab:

- (1) von Art und Variabilität der Rohdaten (i. e. des Sprachatlases),
- (2) von Zielsetzung und Art des vom Dialektometer verwendeten Meßverfahrens, wobei der „Selektivität“ dieses Meßverfahrens (*Splitting* versus *Lumping*) eine besondere Bedeutung zukommt.

Die in Abb. 37.5 visualisierten Verteilungen – die entfernt an die von G. K. Zipf entdeckten Wortschatzverteilungen erinnern – haben G. Altmann (1985) veranlaßt, deren anscheinende Regularität auf einen Poisson'schen Geburts- und Todesprozeß zurückzuführen, ein entsprechendes Gesetz („Goebbl'sches Gesetz“) zu erstellen und anhand empirischer Daten (aus dem französischen Sprachatlas ALF: $N = 71$, $p = 1468$ Atlaskarten) zu verifizieren. Es ist allerdings sehr schwer, für die auf einer Sprachatlaskarte aufscheinende Typen- bzw. Taxat-Vielfalt die verursachenden *Wirkfaktoren* (Kräfte, Synergien etc.) in eindeutig quantifizierbarer Form zu erfassen. Zwar haben viele Einzelanalysen von Sprachatlasdaten (vor allem im Rahmen der Onomasiologie) diesbezüglich eine Vielzahl anekdotischer Information beigebracht. Doch beschränkt sich deren Erklärungskraft jeweils nur auf eine einzige Sprachatlaskarte. Die Einbindung dieser disparaten Einzelphänomene in eine mathematisch faßbare Gesamtsynergetik ist bis jetzt noch nicht geglückt bzw. scheint zur Zeit auch außer Griffweite zu sein. Gangbar ist dagegen der induktiv-empirische Weg. Dieser setzt voraus, daß jede Sprachatlaskarte mit der ihr eigenen Strukturierung als ein *Puzzle-Stein* eines *übergeordneten Globalmusters* aufgefaßt wird, das als Gesamtbilanz der in der betreffenden Gegend abgelaufenen (Sprach-)Geschichte gelten kann. Eine Zusammenführung vieler Puzzle-Steine (mittels taxometrischer Methoden) würde es nunmehr gestatten, diesem übergeordneten Globalmuster auf die Spur zu kommen. Daraus erwächst auch die besondere *interdisziplinäre Relevanz* dialektometrischer Synthesen. In diesem Zusammenhang muß allerdings ergänzt werden, daß entsprechende Versuche gezeigt haben, daß sich die Gestalt (Musterung, Struktur etc.) von dialektometrischen Synthesen jenseits

einer bestimmten Mindestmenge an synthetisierten Einzeldaten (= Atlas-, genauer: „Arbeits“-Karten) nur mehr wenig verändert. Diese Mindestmenge liegt bei rund 200–300 (thematisch voneinander unabhängigen) Arbeitskarten. Die hier beigelegten AIS-Karten basieren auf rund 700 und die ALF-Karten auf rund 1700 Arbeitskarten: daher sind die auf ihnen aufscheinenden Musterungen stark redundant und damit „sehr solide“.

Die N Spalten der Datenmatrix werden auch „Meßpunktvektoren“, die p Zeilen „Merkmalsvektoren“ genannt. Einzelne Meßpunktvektoren werden mit den Indizes j und/oder k angesprochen, während für einzelne Merkmalsvektoren die Indizierung mittels i üblich ist. Der untersuchte Sprachatlas kann ganz oder teilweise vermessen werden. Es ist möglich, dem Netz der explorierten Meßpunkte einen oder mehrere *künstliche* Meßpunkte (z. B. die jeweilige Hochsprache: siehe hier die Abb. 37.11 [Italienisch] und 37.12 [Französisch]) hinzuzufügen. Einer Karte eines Sprachatlases können fallweise auch *mehrere* Merkmalsvektoren („Arbeitskarten“) entnommen werden (Doppel- oder Mehrfachtaxierung). Man muß allerdings streng darauf achten, daß das gewählte Meßverfahren (Taxierung) in sich konsistent ist und über die Dauer der gesamten Messung dieselbe Validität behält. Zu allen hierher gehörenden theoretischen und praktischen Problemen sei exemplarisch auf Goebel 1984 I und Schiltz 1996 verwiesen.

3. Meßmomente

Die einmal konstituierte Datenmatrix kann nunmehr in der verschiedensten Weise ausgewertet werden. Die *Wahl* des im Zuge der taxometrischen Auswertung der Datenmatrix verwendeten Meßmoments hängt vom *Untersuchungsziel* ab. Man sollte es vor Arbeitsbeginn definiert haben. Diese prinzipielle *Zielgerichtetheit* jeder taxometrischen Arbeit soll hier nochmals explizit unterstrichen werden. Im Rahmen dieser vorwiegend methodisch ausgerichteten Kurzdarstellung können nur einige Meßmomente dargestellt und ihre jeweiligen Einsatzbereiche post festum kurz angedeutet werden. Der bei taxometrisch „reifen“ Wissenschaften übliche Vorgang wäre umgekehrt: zuerst Reflexion über das Erkenntnisziel, dann Wahl der taxometrischen Mittel.

3.1. Ähnlichkeitsmessung
(Abb. 37.6 – Abb. 37.15)

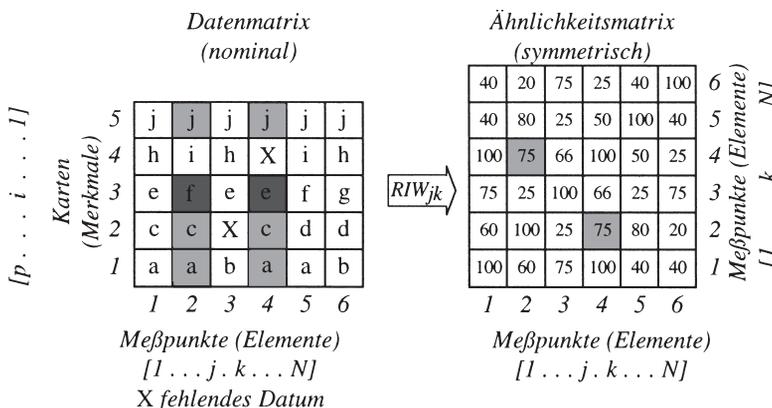
Seit es Sprachatlanten gibt, wird versucht, die darin enthaltenen punktgestützten Daten zu kompakteren Aggregaten zusammenzufassen. Dies geschah und geschieht unter Ansehung der *Ähnlichkeiten* und/oder *Verschiedenheiten (Distanzen)* der Einzelbelege zueinander. In der Terminologie der Taxometrie wird die Messung der (paarweisen) Ähnlichkeiten und/oder Verschiedenheiten (Distanzen) von Elementvektoren der Datenmatrix „Q-Analyse“ genannt. Was im Zuge einer Q-Analyse unter *Ähnlichkeit* bzw. unter *Distanz* zu verstehen ist, muß zunächst genau definiert werden. Die beiden Begriffe – *Ähnlichkeit* und *Distanz* – gehören zur Klasse der *logischen Relationen*. Ihre Anzahl ist theoretisch unbegrenzt. Es ist daher Aufgabe des Dialektometers, in der Sprachgeographie bereits mehr oder weniger explizit vorhandene Ähnlichkeits- und/oder Distanzvorstellungen mit taxometrischen Mitteln nachzuvollziehen bzw. derartige sprachgeographische Ähnlichkeits- und/oder Distanzvorstellungen in das taxometrische Messinstrumentarium einfließen zu lassen: cf. dazu auch Goebel 1983 (a und b) und 1987. Das ist ein besonders heikler Punkt innerhalb der taxometrischen Verfahrenskette, vor allem deshalb, weil im Rahmen linguistischer Klassifikationen (oder Typologien) der Problembereich der *prinzipiellen Vielfalt* der Relationen und des notwendigerweise damit verbundenen Definitions-

zwanges bislang als solcher noch nicht erschlossen worden ist. *Dinge* (hier: Meßpunktvektoren), *Eigenschaften* (hier: Merkmalsvektoren) und *Relationen* (hier: gewähltes Meßmoment): diese *Trias* ist die *logische Basis* jeglicher Klassifikation: cf. dazu Ujmov 1965.

3.1.1. Erstellung einzelner Ähnlichkeitskarten

3.1.1.1. Methodisches, Verfahrenstechnisches

Der französische Naturforscher und Anthropologe Joseph-Pierre Durand (1826–1900) hat vor weit mehr als 100 Jahren einen geolinguistisch brauchbaren Ähnlichkeitsbegriff wie folgt definiert: „Et maintenant, qu’est-ce qui constitue le degré de ressemblance qui rapproche deux langues entre elles, et le degré de dissemblance qui les éloigne l’une de l’autre? La ressemblance se mesure à la proportion des caractères communs, la dissemblance à la proportion des caractères particuliers“ [„Und was definiert nun das Ausmaß der Ähnlichkeit, die zwei Sprachen miteinander verbindet, und was das Ausmaß der Unähnlichkeit, die sie voneinander scheidet? Die Ähnlichkeit wird durch den Prozentsatz der übereinstimmenden Sprachmerkmale erfaßt, die Unähnlichkeit aber durch jenen der differierenden Sprachmerkmale.“]. (Durand 1889, 63). Dem darin formulierten Ähnlichkeitsbegriff entspricht der „Relative Identitätswert“ (RIW_{jk}). Siehe dazu Abb. 37.1.



Datenmatrix: heller Raster: Koidentitäten ($KOI_{2,4}$) i in $i = 1, 2, 5$.
 dunkler Raster: Kodifferenzen ($KOD_{2,4}$) i in $i = 3$.
 Ähnlichkeitsmatrix: heller Raster: $RIW_{2,4}$ und $RIW_{4,2}$.

Abb. 37.1: Generierung einer Ähnlichkeitsmatrix aus der Datenmatrix durch Messung der paarweisen Ähnlichkeiten der Meßpunktvektoren der Datenmatrix mittels RIW_{jk} .

Die Formel dazu lautet:

$$RIW_{jk} = 100 \cdot \frac{\sum_{i=1}^{\tilde{p}} (KOI_{jk})_i}{\sum_{i=1}^{\tilde{p}} (KOI_{jk})_i + \sum_{i=1}^{\tilde{p}} (KOD)_{jk})_i} \quad (1)$$

In dieser Formel bedeuten:

RIW_{jk}	Maß für die Ähnlichkeit zwischen den Meßpunktvektoren j und k ($0 < RIW_{jk} < 100$)
$\sum_{i=1}^{\tilde{p}} (KOI_{jk})_i$	Anzahl der im Meßpunktvektorenpaar jk vorhandenen paarweisen Taxatübereinstimmungen (Koidentitäten)
$\sum_{i=1}^{\tilde{p}} (KOD)_{jk})_i$	Anzahl der im Meßpunktvektorenpaar jk vorhandenen paarweisen Taxatunterschiede (Kodifferenzen)
\tilde{p}	Anzahl der pro Meßpunktvektorenpaar verrechenbaren Merkmalsebenen (nach Abzug der fehlenden Daten)

Der *RI*-Wert nimmt stets positive Werte an, pendelt zwischen 0 und 100, ist unabhängig von der Meßrichtung (symmetrisch: $RIW_{jk} = RIW_{ki}$) und ergibt bei reflexiver Messung (RIW_{ji}) stets 100. Für jede Koidentität ($(KOI_{jk})_i$) und für jede Kodifferenz ($(KOD)_{jk})_i$) wird definitionsgemäß der Wert 1 eingesetzt. Insofern ist der *RI*-Wert gleichgewichtend (adansonistisch). Immerhin wäre es ja denkbar, jede Koidentität nicht von vorneherein mit 1 gleichzusetzen, sondern ihr in Abhängigkeit von ihrer Lage innerhalb der Datenmatrix einen variablen Wert beizumessen (cf. Goebel 1983a und b sowie 1987, alle passim). Der *RI*-Wert ist ein Standardmaß der numerischen Klassifikation und hat dort verschiedene Namen. Zur Problematik cf. Goebel 1984 I, 74 f.

Einsetzbeispiel anhand Abb. 37.1:

Man suche auf: $j = 2, k = 4$.

Hiefür gelten:

$$\sum_{i=1}^3 (KOI_{2,4})_i = 3$$

$$\sum_{i=1}^3 (KOD)_{2,4})_i = 1$$

$$RIW_{2,4} = 100 \cdot \frac{1}{3 + 1}$$

$$RIW_{2,4} = 75$$

Der Wert $RIW_{2,4} = 75$ ist in der Ähnlichkeitsmatrix (Abb. 37.1, rechte Hälfte) durch Grauraster hervorgehoben. Die Ähnlichkeitsmatrix hat stets die Dimensionen N mal N und entsteht durch N mal N nach demselben Schema durchgeführte Messungen. Vermöge der metrischen Eigenschaften des ver-

wendeten Ähnlichkeitsmaßes (RIW_{jk}) enthält die Diagonale der Ähnlichkeitsmatrix nur den Wert 100. Ferner sind die beiden Hälften der Ähnlichkeitsmatrix spiegelgleich. Daher verbleiben zur weiteren taxometrischen Exploration: $1/2(N^2 - N) = N/2(N - 1)$ Meßwerte. Die Ähnlichkeitsmatrix ist jene Drehscheibe, von der die weiteren taxometrischen Schritte ausgehen: dazu gehört zunächst die Erstellung von „Ähnlichkeitskarten“. Darunter versteht man Visualisierungen von „Ähnlichkeitsverteilungen“, die ihrerseits wiederum je einer Spalte (oder Zeile) der Ähnlichkeitsmatrix entsprechen. Siehe dazu Abb. 37.2, links oben.

Eine Ähnlichkeitsverteilung besteht immer aus N Meßwerten, wovon immer ein Meßwert 100 beträgt, der das Resultat der reflexiven Messung (RIW_{jj}) ist. Eine ganze Ähnlichkeitsmatrix enthält stets N Ähnlichkeitsverteilungen. Dies bedeutet, daß zu jedem der N Meßpunkte einer Datenmatrix eine eigene Ähnlichkeitsverteilung errechnet werden kann. Jede Ähnlichkeitsverteilung ist also definiert durch einen Prüfbezugs punkt j . Aus Abb. 37.2 wird ferner ersichtlich, daß die Exhaustion einer Ähnlichkeitsverteilung nur eine von vielen Möglichkeiten darstellt, die Daten der Ähnlichkeitsmatrix taxometrisch auszuwerten. Die taxometrisch-klassifikatorische Auswertung der Ähnlichkeitsverteilung kann *visuell* und/oder *numerisch* erfolgen. Für Sprachgeographen ist selbstverständlich der *visuelle* Weg besonders wichtig. Für die *Visualisierung* der N Werte einer Ähnlichkeitsverteilung empfiehlt sich ganz besonders das thema-

Ähnlichkeitsmatrix
symmetrisch

40	20	75	25	40		
40	80	25	50		40	
100	75	66		50	25	
75	25		66	25	75	
60		25	75	80	20	
	60	75	100	40	40	

1 2 3 4 5 6
Meßpunkte (Elemente)
[1 j N]

6
5
4
3
2
1
Meßpunkte (Elemente)
[1 k N]

Links oben, gerastert: Ähnlichkeitsverteilung zu $j (= k) = 4$.

Rechts unten, gerastert: an der Berechnung von Kennwertsynopsen und Dendrogrammen beteiligter Bereich der Ähnlichkeitsmatrix.

Die (symmetrische) Ähnlichkeitsmatrix wurde entlang der Diagonale in ihre zwei (jeweils asymmetrischen) identischen Hälften zerlegt.

Abb. 37.2: Beispiele für verschiedene Exhaustionen der Ähnlichkeitsmatrix.

kartographische Instrument der *Flächenmosaik-* und *Choroplethenkarte*. Hiefür sind erforderlich:

- (a) die Verwandlung der im Raum diskret verteilten Meßpunktmenge in ein fugelos und überlappungsfrei gestaltetes Flächenmosaik, das zur Aufnahme von Farb- oder Graurastermarkierungen geeignet ist und solcherart als „Pseudokontinuum“ dient;
- (b) die Umlegung der *numerischen Variation* der Ähnlichkeitsmeßdaten in eine für das menschliche Auge sehspsychologisch günstige *optische* (ikonische) *Variation*.

Zu (a): Hiefür eignet sich optimal ein geometrisches Verfahren, das – obschon in mathematischen Fachkreisen seit Jahrhunderten bekannt – 1898 vom Germanisten Carl Haag erstmals auf sprachgeographische Fragestellungen angewendet wurde und im Rahmen der Geographie seit geraumer Zeit mit dem Namen des Meteorologen Alfred H. Thiessen (1911) und in jenem der Mathematik mit den Namen der Mathematiker B. Delaunay und G. F. Voronoi verbunden wird (cf. dazu Okabe/Boots/Sugihara 1992). Dieses Verfahren sieht die folgenden Prozeduren vor:

- (1) Dreiecksvernetzung der Meßpunkte des Untersuchungsgebietes.
- (2) Errichtung der Streckensymmetralen über den Dreiecksseiten.
- (3) Verschmelzung jeweils dreier Streckensymmetralen zum Mittelpunkt des dem fraglichen Dreieck umschriebenen Kreises (Umkreismittelpunkt).
- (4) Die Umkreismittelpunkte stellen die Eckpunkte von Polygonen dar, innerhalb deren der jeweilige Meßpunkt liegt.

Zu weiteren Details (mit Konstruktions-skizze) cf. Goebel 1984 I, 90f. und Okabe/Boots/Sugihara 1992 (passim).

Zu (b): Kartographen, Semiotikern und Psychologen ist gut bekannt, daß das menschliche Auge bei der Betrachtung von Flächenmosaiken differentiell nur 6–8 Tonwert- oder Farbstufen verarbeiten kann. Unser Problem besteht nun darin, die *Variation* von N Meßwerten in einer (verhältnismäßigen) *Variation* von 6–8 *Tonwert-* oder *Farbstufen* abzubilden. Dieser Abbildungsvorgang kann wiederum auf die verschiedenste Art und Weise erfolgen und stellt demnach ein vollwertiges Glied der taxometrischen Kette dar. Die Wahl der Abbildungsmodalitäten muß daher mit besonderer Sorgfalt getroffen werden. Auch hiefür steht im Rahmen der statistischen Kartographie erprobtes Handbuchwissen zur Verfügung: cf. Dickinson 1973. Bei der Erstellung der Abb. 37.6–37.15 wurde die Visualisierung mittels des Algorithmus MINMWMAX durchgeführt. Der fragliche Algorithmus sieht vor: 6 Graurasterstufen, 3 davon unterhalb und 3 oberhalb des arithmetischen Mittels der zu visualisierenden Ähnlichkeitsverteilung. Die unterdurchschnittlichen Intervallbreiten errechnen sich durch Drittelung der Spanne zwischen Minimalwert und arithmetischem Mittel, die überdurchschnittlichen Intervallbreiten durch Drittelung der Spanne zwischen Maximalwert und arithmetischem Mittel.

Dieser Algorithmus (MINMWMAX) hat sich aus vielerlei Gründen für choroplethische Zwecke als brauchbarster Abbildungsschlüssel erwiesen. Freilich gibt es für andere Zwecke und besondere Fragestellungen daneben noch zahlreiche andere Umsetzungsalgorithmen. Zur Problematik cf. Goebel 1984 I, 93f. Die Abb. 37.16–37.19 beruhen auf den Intervallalgorithmen MEDMW und MED: beide stützen sich auf mit möglichst gleich vielen Meßpunkten be-

setzte Klassen: bei MEDMW zu beiden Seiten des Mittelwerts, bei MED zu beiden Seiten des Medians: cf. dazu Goebel 1984 I, 95–97.

Auf den Abb. 37.6–37.19 sind rechts unten im Legendenbereich Stabdiagramme sichtbar, die eine konkrete Vorstellung von der Gestalt der visualisierten Häufigkeitsverteilung geben. Die Konstruktion dieser Stabdiagramme (Histogramme) wird ebenso mit Hilfe des jeweils eingesetzten Intervallalgorithmus durchgeführt: zu Details cf. Goebel 1984 I, 97 f.

3.1.2. Allgemeines zur Interpretation der Abb. 37.6–37.14

Zu beachten ist, daß die neun Karten auf zwei Datenmatrizen beruhen, die anhand des AIS (696 Arbeitskarten aus den Bänden 1, 2 und 4; lexikalische und morphosyntaktische Variation) und des ALF (1687 Arbeitskarten aus 626 ALF-Originalkarten; phonetische, lexikalische und morphosyntaktische Variation) generiert wurden. Zudem enthalten die beiden Meßpunktnetze je einen Kunstpunkt, der der jeweiligen Hochsprache entspricht: AIS: P. 999 (in der Nähe von Florenz), ALF: P. 999 (in der Nähe von Paris). Die Kenntnis dieser datenseitigen Voraussetzungen ist wichtig, um die Kartenbilder richtig einzuschätzen. Zu beachten ist ferner, daß gewisse Meßpunkte des AIS-Netztes einen Sternchenvermerk tragen. Es sind das Meßpunkte, in deren Vektoren zahlreiche Daten (oft bis zu zwei Dritteln des Gesamtpotentials) fehlen. Die an ihnen auftretenden Meßwerte fallen daher aus dem Rahmen des sonst Üblichen. Sie sind meist unverhältnismäßig hoch. Bei der Interpretation der Choroplethenprofile muß man diesen Umstand berücksichtigen und sollte die fraglichen Polygone am besten „geistig ausblenden“.

Die Abb. 37.6–37.14 stellen fünf von insgesamt 251 (AIS) bzw. vier von insgesamt 641 (ALF) Ähnlichkeitskarten dar. Sie enthalten jeweils einen bestimmten Bildtyp, der dem Betrachter in der Form eines „Gebirges“ (warmfarbene Polygone: große Reliefhöhen, kaltfarbene Polygone: geringe Reliefhöhen) entgegentritt. Bei der Betrachtung der Karten sollte man seine Aufmerksamkeit auf die folgenden Bildteile richten: Lage des „Gebirgssockels“ (Farben 1–2: Dunkelblau, Hellblau), der „Gebirgsflanken“ (Farben 3–4: Grün, Gelb) und der „Spitzen des Gebirges“ (Farben 5–6: Orange, Rot).

Das Zusammenspiel der sechs nach der Logik des Regenbogens gestaffelten Farben erzeugt Raumbgliederungen, die dialektologisch bedeutsam sind. So erkennt man auf Abb. 37.6 durch das Zusammenspiel der Farben Rot und Orange einen als „lombardisch“ anzusprechenden Zentralkern, dann das Einzugsgebiet des Piemontesischen (Grün), ferner das Aostatal (Dunkel- und Hellblau), das Bündnerromanische (ebenso Dunkel- und Hellblau), den Stoß zwischen dem Festlandvenezianischen und umgebenden Dialektzonen etc.

Eine volle Würdigung der typodiagnostischen Leistung der Ähnlichkeitsmessung kann eigentlich nur nach Einsicht aller *N* Flächenmosaikkarten und aller dazugehörigen Zahlenwertkarten (auf deren Präsentation hier aus Platzgründen zur Gänze verzichtet werden muß) vorgenommen werden. Wichtig ist die Erkenntnis, daß zwischen linguistischer Ähnlichkeit und geographischer Distanz ein variabler (und keineswegs ein fester) Zusammenhang besteht, der für den vorgewählten Prüfbezugspunkt typisch ist und daher zu typodiagnostischen Zwecken herangezogen werden kann.

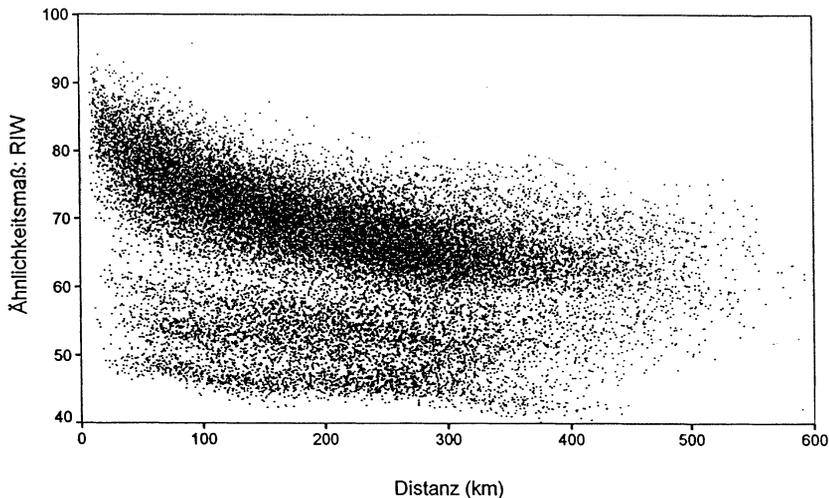
Siehe dazu die Abb. 37.3 die diesen Zusammenhang zwischen linguistischer Ähnlichkeit und geographischer Distanz für die AIS-Daten-Matrix anhand zweier Ähnlichkeitsmaße (RIW und GIW(1)) in sehr anschaulicher Form zeigt.

Man unterscheidet auf den beiden Scattergrammen deutlich einen Hauptschwarm und mindestens einen Nebenschwarm. Der Hauptschwarm gehört zu den anhand der genuin italienischen Meßpunkte vorgenommenen Ähnlichkeitsmessungen, der Nebenschwarm zu jenen Ähnlichkeitsmessungen, die zu den rätoromanischen und frankoprovenzalischen Meßpunkten des AIS-Netztes durchgeführt wurden. Man beachte den mittels der Gewichtung (bei GIW(1)) erzielbaren „Kompaktierungseffekt“!

Wiewohl zu *N* Meßpunkten stets *N* Ähnlichkeitskarten erstellt werden können, bedeutet dies nicht, daß jede dieser *N* Ähnlichkeitskarten einen gesonderten Bildtyp repräsentiert. So sind die Prüfbezugspunkte der Abb. 37.6 und 37.7 sowie 37.8 und 37.9 einander direkt benachbart. Man erkennt, daß die betreffenden Choroplethenprofile einander sehr ähnlich sind. Bei Versetzung des Prüfbezugspunktes erfährt der jeweilige Bildtyp nur dann eine tiefgreifende Veränderung, wenn dabei der Prüfbezugspunkt in eine typologisch stark abweichende Zone

Relation zwischen linguistischer Ähnlichkeit
und geographischer Distanz

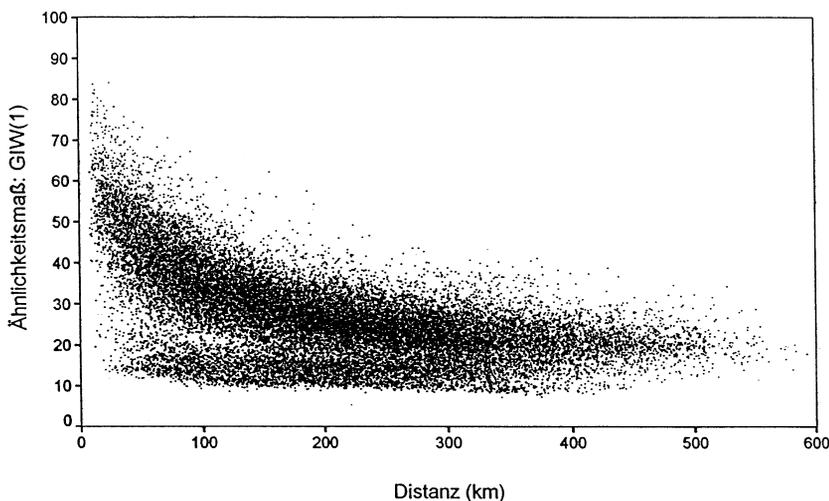
(anhand AIS vol. I, II und IV)



Gesamtkorpus: 251 Meßpunkte (31.375 RI-Werte)

Relation zwischen linguistischer Ähnlichkeit
und geographischer Distanz

(anhand AIS vol. I, II und IV)



Gesamtkorpus: 251 Meßpunkte (31.375 GI(1)-Werte)

Abb. 37.3: Zwei Ausprägungen (anhand RIW und GIW (1)) der Relation zwischen linguistischer Ähnlichkeit und geographischer Distanz.

gerät: so liegen die Prüfbezugs-
punkte der Karten 37.6 und 37.7
einerseits und der Abb. 37.8
und 37.9 andererseits in solchen
„typologisch stark voneinander
abweichenden

Zonen“. Diese von Ähnlichkeits-
karte zu Ähnlichkeitskarte mehr
oder weniger stark schwankende
Bildtypendifferenz kann – bei
Vorhandensein entsprechender photo-

oder EDV-graphischer Hilfsmittel (Überblendprojektion, Trickfilm, Computer-Animation) – zur Erzeugung kinematographischer Effekte (Fliehbild-Effekte) in der Form einer „dynamischen Bildtypologie“ verwendet werden: cf. dazu Goebel 1984 I, 104 f.

Die sprachgeographisch bedeutsamste Aussagekraft von Ähnlichkeitskarten liegt in der präzisen Beantwortung einer explizit und implizit im Rahmen von Dialektologie und Variationslinguistik sehr oft gestellten Frage, nämlich nach der „Stellung der Varietät X im Rahmen eines umgebenden Varietätenraumes“. Die von den Abb. 37.6 und 37.7 gelieferten Informationen zeigen deutlich, **wie** die Dialektizitäten der AIS-Meßpunkte 252 respektive 1 innerhalb des untersuchten Meßpunktfeldes gelagert sind. Die Frage nach dem WIE wird hier *quantitativ-typologisch* beantwortet und setzt damit den Rückgriff auf eine *quantitative* Begrifflichkeit voraus, die in der Sprachwissenschaft im allgemeinen und in der sprachwissenschaftlichen Typologie im besonderen keine große Tradition hat; cf. dazu Goebel 1984 I, 98 f.

3.1.2.1. (Geo)Linguistische Interpretation der Abb. 37.6–37.14

Vorbemerkung: Zur Numerierung der Polygone der ALF-Karten siehe http://ald.sbg.ac.at/dm/germ/Allgemein/ALF_Netz.pdf.

Die Abb. 37.6 und 37.7 haben einander direkt benachbarte Prüfbezugspunkte und zeigen beide ein typisch *westlombardisches* Ähnlichkeitsprofil. Der Profilverlauf gestaltet sich gegen Norden (Rumantschia bzw. Rätoromanisch der Schweiz) sehr abrupt, gegen Westen, Süden und Osten dagegen um einiges weicher. Aus der Sicht der romanischen Geolinguistik kann die Lage aller sechs Wertstufen plausibel interpretiert werden. Der Prüfbezugspunkt der Abb. 37.6 (252-Monza) ist stärker städtisch beeinflusst als jener von Abb. 37.7 (250-Maniago: ein kleines Dorf, 30 km westlich von Mailand). Man erkennt dies vor allem an der größeren Ähnlichkeit des Toskanischen (Wertklasse 4, Gelb, statt 3, Grün) und des Hochitalienischen zum Prüfbezugspunkt 252.

Die Abb. 37.8 und 37.9 haben erneut einander benachbarte Prüfbezugspunkte, die beide im Einzugsbereich des *Surselvischen*, einer westlichen Variante des Bündnerromanischen, liegen. Bemerkenswert sind die Lage der „Antipoden“ (Wertklasse 1, Dunkelblau) in der Toskana und in der Romagna

sowie die nach Osten weisenden Ähnlichkeitsbezüge zum Ladinischen und Friaulischen (Wertklasse 4, Gelb). Die Bildsyntax wird an einigen Stellen (*AIS-Punkte mit reduziertem Korpus*) auf Grund des Fehlens von Daten durch Polygone in (der überhöhten) Wertklasse 4 (Gelb) gestört.

Die Abb. 37.10 stellt – da ihr Prüfbezugspunkt das Hochitalienische (ein in das AIS-Netz post festum eingesteuerter „Kunstpunkt“) ist – letztendlich eine Bilanz der Italianisierung des ganzen Untersuchungsraumes dar. Die „anti“-italienischen Residuen liegen erwartungsgemäß an der Peripherie.

Die auf dem französischen Sprachatlas ALF beruhenden Abb. 37.11–37.14 haben nacheinander als Prüfbezugspunkte das Hochfranzösische (Abb. 37.11), einen pikardischen (Abb. 37.12), einen frankoprovenzalischen (Abb. 37.13) und einen languedokischen (Abb. 37.14) Lokolekt. Die darauf sichtbaren Choroplethenprofile sind für die Einzugsbereiche des zentralen Nordfranzösischen, des Pikardischen, Frankoprovenzalischen und Languedokischen als „typisch“ zu bezeichnen und treten daher auch bei der Ähnlichkeitsmessung ausgehend von ALF-Punkten auf, die in der Nähe zu den hier gewählten Prüfbezugspunkten liegen.

Die Summe der warmfarbenen Polygone einerseits und jene der kaltfarbenen Polygone andererseits ist auf den Abb. 37.11 und 37.12 ungefähr gleich groß und deckt eine in der Galloromanistik traditionsgemäß als „Domaine d’Oil“ bezeichnete Sprachlandschaft ab. Umgekehrt verweisen die warmfarbenen Polygone der Abb. 37.14 auf die galloromanische Sprachlandschaft „Domaine d’Oc“. Auf Abb. 37.13 entsprechen die roten und orangefarbenen Polygone weitgehend der traditionell dem Frankoprovenzalischen zugesprochenen Fläche.

Die Lage der jeweiligen „Antipoden“ (Dunkel- und Hellblau) ist räumlich deutlich markiert.

Noch zwei Hinweise:

- (1) Am Südrand der katalanophonen Landschaft Roussillon befindet sich der Kunstpunkt Hochkatalanisch (P. 997) und am äußersten Ostrand der Provence – und zwar direkt an der Küste – der Kunstpunkt Hochitalienisch (P. 998). Es hat sich allerdings gezeigt, daß die diesen beiden Kunstpunkten zugeordneten Hochsprachen bei der Ausbildung der (süd)galloromanischen Land-

schaften keine auch nur annähernd mit dem Hochfranzösischen vergleichbare Rolle gespielt haben.

- (2) Auf den Abb. 37.11–37.13 ist südlich von Bordeaux ein gelb markiertes Polygon sichtbar, das im Westen und Süden von hellblauen und im Osten und Norden von grünen Polygonen umgeben und von der warmfarbenen Zone deutlich abgesetzt ist. Es handelt sich dabei um den zur (nordfranzösischen) Sprachinsel der „Petite Gavacherie“ gehörenden ALF-Meßpunkt 635 (Andraut, Département Gironde).

3.1.2.2. Außerlinguistische Interpretation der Abb. 37.6–37.14

Die Abb. 37.6–37.14 können auch aus *außerlinguistischer* Sicht interpretiert werden. So erlaubt es die algebraische Logik des verwendeten Ähnlichkeitsmaßes (RIW_{jk}), eine Ähnlichkeitskarte *fernsprechtechnisch* zu interpretieren. Dabei entspricht dem Prüfbezugs- punkt ein Telephonanschluß: der Profilverlauf der Ähnlichkeitskarte zeigt an, wohin der betreffende Telephoninhaber am meisten telephonierte bzw. woher er die meisten Gespräche empfangen hat (cf. dazu Rauh 1998, passim). Diese *funktionale* Interpretation *per analogiam* vollzieht grosso modo jene Prozesse nach, die bei einer räumlichen *Diffusion* von sprachlicher Affinität (repräsentiert hier vor allem durch lexikalisch und morphosyntaktisch affine Einheiten) ablaufen bzw. bis dato abgelaufen sind.

Eine andere interdisziplinäre Analogieinterpretation der Ähnlichkeitskarten bezieht sich auf den in der Humangeographie gebräuchlichen Begriff der „Zentralität“ bzw. der „zentralen Lage“ eines Orts (hier: eines Meß- bzw. Prüfbezugs punktes). Ein Vergleich der Abb. 37.6 und 37.8 zeigt, daß die Dialektizität des AIS-Punkts 252 (i. e. der Vektor des Meßpunkts 252) besser in die Datengesamtheit integriert ist als die Dialektizität des AIS-Punkts 1. Als Maßzahl hierfür kann beispielsweise das arithmetische Mittel (oberer Eckwert der Farbstufe 3, Grün) der jeweiligen Ähnlichkeitsverteilung genommen werden (Abb. 37.6: 69,39; Abb. 37.8: 48,37). In diesem Zusammenhang ist auch ein Vergleich der Silhouetten der beiden Histogramme angebracht: auf Abb. 37.6 ist eine leicht *rechtsschiefe*, auf Abb. 37.8 eine stark *linksschiefe* Verteilung sichtbar. Die Ähnlichkeitsverteilungen sind also nicht nur *ikonisch-kartographisch*, sondern offenbar auch *mathematisch-statistisch* typisiert.

3.2. Erstellung von Kennwertsynopsen (Abb. 37.15–37.17, Anhang)

Die Ähnlichkeitskarte repräsentiert eine Ordnungsstruktur *niederen* Ranges. Will man *höherrangige* und damit komplexere, umfassendere, globalere Ordnungsstrukturen sichtbar machen, muß man entsprechende Komplexitätskriterien definieren bzw. – mit anderen Worten – in komplexerer Weise auf die Daten der Ähnlichkeitsmatrix zugreifen. Siehe dazu Abb. 37.2, rechte untere Hälfte.

3.2.1. Synopse der Schiefen von N Ähnlichkeitsverteilungen (Abb. 37.15–37.17, Anhang)

Unter *Schiefe* versteht man einen Kennwert der deskriptiven Statistik, mit dem die *Symmetrie* bzw. *Asymmetrie* einer Häufigkeitsverteilung erfaßt wird. Ein Vergleich der Silhouetten der Histogramme der Abb. 37.6 und 37.8 ist hierfür lehrreich. Das Prinzip der Synopse der Schiefen besteht darin, für jede der Ähnlichkeitsverteilungen die Schiefe zu errechnen und anschließend den errechneten Schiefenwert an der Stelle des jeweiligen Prüfbezugs punktes einzutragen. Dieser Vorgang wird N mal wiederholt. Solcherart entsteht eine Häufigkeitsverteilung aus N Schiefenwerten, die in üblicher Weise visualisiert wird: siehe dazu die Abb. 37.15–37.17. Die Formel für die Schiefe lautet wie folgt:

$$g_j = \frac{\sum_{i=1}^n (RIW_{jk} - \bar{X}_j)^3}{n \cdot s_j^3} \quad (2)$$

In dieser Formel bedeuten:

- g_j Schiefe einer Ähnlichkeitsverteilung zum Prüfbezugs punkt j
 s_j Standardabweichung einer Ähnlichkeitsverteilung zum Prüfbezugs punkt j
 RIW_{jk} einzelner *RI*-Wert einer Ähnlichkeitsverteilung zum Prüfbezugs punkt j
 k Index für die jeweils mit j verglichenen Meßpunktvektoren: läuft zwischen 1 und N
 n $N - 1$ (hier: 250, bei $N = 251$ (bzw. 640, bei $N = 641$))
 \bar{X}_j arithmetisches Mittel einer Ähnlichkeitsverteilung zum Prüfbezugs punkt j

Der Wert g_j kann positive und negative Werte annehmen. Bei vollständiger Symme-

trie der Ähnlichkeitsverteilung ist der Wert von g_j Null. Für taxometrische Zwecke können sowohl die Meßwertausprägung als auch nur die Vorzeichenvariation benutzt werden. Für die Abb. 37.15–37.17 gilt ersteres. Zum Verständnis der drei Karten muß man sich zuvorderst über die *geolinguistische* Bedeutung des Kennwerts „Schiefe“ Rechenschaft ablegen. Bei Ähnlichkeitsverteilungen mit stark *negativer* Schiefe (d. h. mit *rechtssteiler* Kurvensilhouette) liegt die Mehrzahl der einzelnen *RI*-Werte *über* dem arithmetischen Mittel. Bei Ähnlichkeitsverteilungen mit stark *positiver* Schiefe (d. h. mit *linkssteiler* Kurvensilhouette: siehe das Histogramm auf den Abb. 37.8 und 37.9) liegt die Mehrzahl der einzelnen *RI*-Werte *unter* dem arithmetischen Mittel. Je größer die Anzahl der *RI*-Werte *über* dem arithmetischen Mittel ist, desto *besser* ist die fragliche Ähnlichkeitsverteilung (und damit die Dialektizität des dazugehörigen Meßpunktvektors) in die Datengesamtheit *integriert*. Die fragliche Dialektizität kann demnach als mehr oder weniger *gut*, *akzeptiert*, „im Trend liegend“, *anpassungsfähig*, *durchsetzungskräftig* etc. betrachtet werden. Dieser Umstand läßt sich durch eine *kommunikationstheoretische Analogie* verdeutlichen: jeder Merkmalsvektor sei einer von N Spielern. Er kann gegen jeden seiner $N-1$ Mitspieler jeweils einmal spielen. Aus diesen $N-1$ Spielen eines jeden Spielers erwächst ein numerisches Spielresultat (hier: die Schiefe g_j). Die N möglichen Spielresultate der N Spieler werden verglichen. Schlechte Spieler erzielen „nur“ positive Schiefewerte, gute Spieler erzielen negative Werte. Diese Problematik knüpft an in der Dialektologie oft gebrauchte, aber stets sehr vage verbliebene Begriffe wie „*laterale* vs. *zentrale* Dialektzonen“, „*konservative* vs. *innovatorische* Gebiete“ etc. an.

3.2.2. (Geo)linguistische Interpretation der Abb. 37.15

Die Abb. 37.15 zeigt ein bildsyntaktisch überaus harmonisch aufgebautes Profil. An der Peripherie liegen die traditionell als „nicht-italienisch“ eingestuftes Dialektzonen des Alpinokzitanischen, Frankoprovenzalischen (Aostatal und südliche Vorlande), Bündnerromanischen, Sellaladinischen und Friaulischen. Sie haben alle sehr hohe Schiefewerte. Dagegen bilden die kleinsten Meßwerte (Farbstufe 1, Dunkelblau) zusammen eine Art Zange, deren Arme *grosso*

modo einerseits dem Tal der Etsch und andererseits dem Verlauf des Apennin entsprechen. Es sind das Zonen, die gemäß unserer Definition Meßpunktvektoren enthalten, deren Dialektizitäten als überaus „anpassungsfähig“ bzw. „durchschlagskräftig“ angesehen werden können. Daß damit zugleich markante Sprachgrenzgebiete erfaßt werden, die aus der klassischen Sprachgeographie her bekannt sind, sollte nicht erstauen, da Sprachgrenzgebiete immer auch dem Sprachaustausch dienen und solcherart eine gewisse „Doppelnatur“ bzw. „Janusköpfigkeit“ besitzen: man vergleiche dazu die Abb. 37.15 und 37.18 bzw. die Abb. 37.16 – 37.17 und Abb. 37.19! *Grosso modo* darf festgestellt werden, daß durch die in dunkelblau kolorierten Polygone jene Gebiete erfaßt werden, in denen die intensivsten Sprachaustausch-, Sprachausgleichs- und Sprachkontaktphänomene stattfinden bzw. stattgefunden haben.

3.2.3. (Geo)linguistische Interpretation der Abb. 37.16 und 37.17

Die Abb. 37.16 und 37.17 wurden zum einen nicht mit 6, sondern mit 8 Farbstufen erzeugt und beruhen zum anderen auf verschiedenen Teilkorpora unserer ALF-Daten. Ersteres geschah, um die hochinteressanten Musterstrukturen der beiden Karten einzeln und im Vergleich miteinander besser sichtbar zu machen; letzteres, um zu zeigen, wie sich kategorielle Unterschiede bei den zu synthetisierenden Daten auf die dialektometrischen Choroplethenprofile auswirken.

Wie bei Abb. 37.15 ist zunächst auf die Lage der dunkelblauen Zonen zu achten, wo bekanntlich die intensivsten Sprachausgleichsprozesse ablaufen. Auf beiden Karten erkennt man dabei eine Nordfrankreich bzw. das *Domaine d’Oïl* umschließende zirkuläre Struktur, die ihr Irradiationszentrum im (grün markierten) Zentrum Nordfrankreichs zu haben scheint. Auch das Frankoprovenzalische (im Südosten) ist Gegenstand einer von Norden, Westen und Süden wirksamen Umfassung. Die vierfache Abstufung der kalten Farben läßt die Zirkularität der Sprachausgleichsgebiete besonders deutlich zutage treten. Demgegenüber signalisieren die in Rot, Orange etc. markierten Gebiete (v. a. Gascogne, Languedoc, Roussillon, Provence) das weitgehende Fehlen weitstrahliger angelegter Sprachausgleichsvorgänge und repräsentieren demnach Relikt- bzw. „Außenseiter“-Zonen. Überaus be-

merkwürdig ist die weitgehende Ähnlichkeit der beiden Choroplethenprofile, die immerhin auf Korpora beruhen (Abb. 37.16: Phonetik, Abb. 37.17: Lexikon), denen traditionell nicht dieselbe Neigung zu Gesetzmäßigkeiten und Strukturbildung zugeschrieben wird. Ich erinnere in diesem Zusammenhang an die in Germanistik und Romanistik hinsichtlich der angeblichen „Spontaneität“ und „Wanderfreudigkeit“ des Wortschatzes seit Jacob Grimm immer wieder zitierte Maxime der zufolge „jedes Wort seine eigene Geschichte“ habe (cf. Christmann 1971). Solcherart wurden Wortschatzstudien dominant im Zeichen kasuistischer Einzelanalysen betrieben. Demgegenüber belegen die hier gezeigten synthetischen Resultate, daß es möglich ist, durch die konsequente Addition vieler Einzel-„Geschichten“ zu einer Global-„Geschichte“ der lexikalischen Bewirtschaftung einer bestimmten Gegend durch die betreffenden Sprecher vorzustoßen bzw. diese zu rekonstruieren.

Auf beiden Karten ist ferner die jeweils überaus charakteristische Gestaltung der Kontaktzonen zwischen dem *Domaine d’Oil* und dem *Domaine d’Oc* bemerkenswert, wobei der breite dunkelblaue Gürtel auf Abb. 37.17 räumlich besonders tief gestaffelte Sprachaustauschvorgänge signalisiert. Siehe dazu auch Goebel 2002, 25 f. und 2003, 81 f.

3.3. Distanzmessung und Isoglossensynthese (Abb. 37.18–37.19, Anhang)

3.3.1. Methodisches, Verfahrenstechnisches

Das klassische Isoglossenprinzip (cf. dazu Händler/Wiegand 1982) kann dialektometrisch unschwer nachmodelliert werden. Zwischen den Meßpunktvektoren werden statt Ähnlichkeiten *Distanzen* gemessen. Dies bedeutet, daß Koidentitäten und Kodifferenzen ihre Stelle wechseln und in der Formel (1) füreinander eintreten. Das dabei entstehende Distanzmaß sei „Relativer Fündigkeitswert“ (RFW_{kj}) genannt. Zum „Relativen Identitätswert“ (RIW_{jk}) verhält sich der RFW_{kj} komplementär:

$$RFW_{kj} + RIW_{jk} = 100 \quad (3)$$

Mit dieser Formel kann aus der Ähnlichkeitsmatrix der Abb. 37.1 und 37.2 unschwer die Distanzmatrix der Abb. 37.4 hergeleitet werden.

Von ausschlaggebender Bedeutung ist ferner, daß aus der solcherart generierten Di-

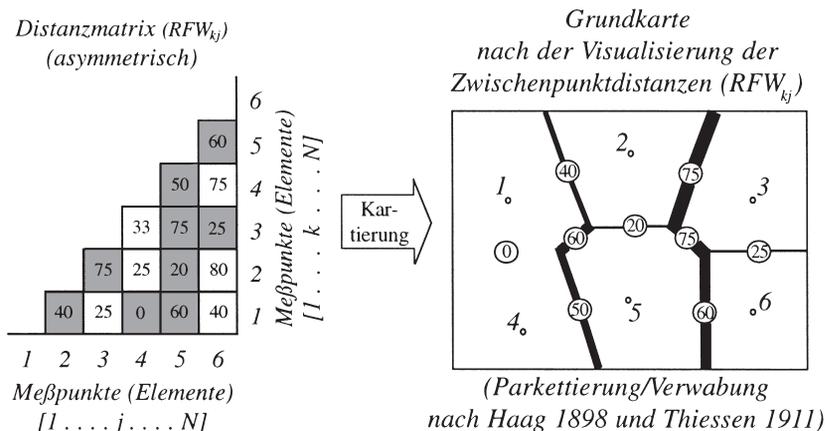
stanzmatrix nur eine relativ kleine Anzahl von Meßwerten zur Kartierung bzw. Visualisierung abgegriffen wird. Es sind dies jene Meßwerte, die auf Grund der durch die Thiessen- (bzw. Delaunay-Voronoi-) Parkettierung festgelegten Nachbarschaftslage (Kontiguität) hierfür in Frage kommen. Siehe dazu wiederum Abb. 37.4 (linke Hälfte). Alle anderen Meßwerte werden beiseitegelassen bzw. unterdrückt. Die fraglichen Relationen sehen in unserem Fall wie folgt aus: in der italienischen Distanzmatrix (nach AIS) sind 31.375 Meßwerte enthalten ($= N/2 (N - 1)$, bei $N = 251$). Davon werden nur 665 Meßwerte (für ebensoviel „Polygonseiten“) ($= 2,12\%$) abgegriffen und visualisiert. Im Falle Frankreich (nach ALF) sind das bei 205.120 Meßwerten der Distanzmatrix und 1791 Polygonseiten gar nur 0,87% aller Meßwerte. Die *Isoglossenklassifikation* ist also sehr *selektiv*.

Die gesamte ikonische Klassifikationsausage entsteht durch das bildsyntaktische Zusammenspiel von 665 bzw. 1791 Polygonseiten sowie durch deren Färbung und Strichstärke. Die strenge Einhaltung der durch das Haag/Thiessen/Delaunay-Voronoi-Prinzip vorgegebenen geometrischen Normen ist für das Gelingen und die Aussagekraft der Isoglossensynthese von entscheidender Bedeutung. Leider haben viele, auch genuin dialektometrisch angelegte, Isoglossensynthesen (z. B. Guiter 1973) diesen Grundsatz nicht beachtet.

Bedeutsam ist ferner – genau wie bei den Choroplethenkarten – die Wahl des passenden Intervallalgorithmus, der die Umsetzung der 665 bzw. 1791 Distanzwerte in eine sechspologisch verkräftbare Anzahl von Strichstärken steuert. Der hier gewählte Intervallalgorithmus (MED) beruht auf dem Prinzip der Bildung von hinsichtlich der Zahl der Meßwerte (bzw. Polygonseiten) möglichst gleich großen Klassen.

Allgemein ist festzuhalten, daß Aussagekraft und Bildqualität von Isoglossensynthesen von der Größe und Dichte des Meßpunktfeldes und damit von der Anzahl der verfügbaren Grenzsegmente (bzw. Polygonseiten) abhängig sind. Dabei gilt die Faustregel: je größer die Anzahl der Grenzsegmente, desto differenzierter die Bildsyntax und damit desto größer der typologische Wert der Isoglossensynthese.

Isoglossensynthesen mittels Wabenkarte gehören seit Haag (1898) zum festen Repertoire der *germanistischen* Dialektologie. *Romanistische* Isoglossensynthesen haben dagegen bis vor kurzem auf die Anwendung



Distanzmaß: Relativer Fündigkeitswert (RFW_{kj}).

Die Werte der Distanzmatrix wurden aus jenen der Ähnlichkeitsmatrix in Abb. 37.1 nach der folgenden Formel abgeleitet: $RFW_{kj} + RIW_{jk} = 100$.

Abb. 37.4: Exhaustion und Kartierung jener RF -Werte (RFW_{kj}) der Distanzmatrix, die hierfür auf Grund der Nachbarschaftsgeometrie der Grundkarte in Frage kommen.

der Haag/Thiessen/Delaunay-Voronoi-Parkettierung verzichtet und sind aus diesem Grund graphisch mehr oder weniger mißlungen (typisch dafür Etmayer 1924, Atwood 1955, Goebel 1981 und zahlreiche Arbeiten von Guiter seit 1973, z. B. Guiter 1984; graphisch bessere Resultate liefern dagegen Polanco Roig 1983 und Saramago 1986). Einen gesonderten Hinweis verdient eine mit einer Wabenkarte anhand von AIS-Daten (phonetische Variation) durchgeführte Isoglossensynthese von Haag selber (1930). Die dabei erzielte Bildaussage ist jener von Abb. 37.18 durchaus ähnlich. Ich habe seit 1985 damit begonnen, vor allem für romanistische Sprachatlanten stumme Karten mit grau oder gerastert aufgedruckter Haag/Thiessen/Delaunay-Voronoi-Parkettierung zu produzieren (cf. Goebel 1998). Die Verwendung derartiger stummer Karten würde es gestatten, in Hinblick graphischen Insuffizienzen der inkriminierten Art zu begegnen und das für methodisch korrekte Isoglossensynthesen notwendige Problembewußtsein der Sprachgeographen zu schärfen.

3.3.2. (Geo)linguistische Interpretation der Abb. 37.18 und 37.19

Auf die Beigabe von Listen aller 665 bzw. 1791 kartierten RF -Werte mußte verzichtet werden: zu den AIS-Werten cf. Goebel 1983b, 396 f. und 1984 III, 285 f. Die bildtypologische Aussagekraft der Abb. 37.18 und 37.19 beruht auf dem sechsfach gestuften Zusammenspiel von 665 bzw. 1791 Grenzsegmenten (Schot-

ten, Wabenkanten, Polygonseiten etc.). An charakteristischen Stellen der beiden Karten ergibt sich dabei aus der mehr oder weniger kompakten Abfolge ähnlich dicker Grenzsegmente der Eindruck einer Begrenzung. Nirgends aber entsteht durch das Zusammenwirken mehrerer Grenzsegmente eine gleichmäßig dicke, durchgehende Linie, die den in der traditionellen Sprachgeographie üblichen Begriff der „Dialektgrenze“ rechtfertigen würde. Dieser undifferenzierte (qualitative) Begriff ist für Zwecke von Typodiagnose und Datensynthese völlig unbrauchbar. Er sollte vermöge der ihm inhärenten Indifferenziertheit nur zur Bezeichnung jener Grenzlinien (= Isoglossen im eigentlichen Sinne) verwendet werden, die das Einzugsgebiet eines *einzigsten*, räumlich kompakt auftretenden geolinguistischen Merkmals (Taxat-types) umspannen. Beim Sprechen von „Dialektgrenzen“ im dialektometrischen Kontext muß man sich dessen bewußt sein, daß eine dialektometrische Dialektgrenze einem *quantitativen* Begriff entspricht und de facto einen höchst differenziert verlaufenden „Höhenzug“ repräsentiert. Überdies hat man diesen Umstand bereits früh erkannt, doch bis heute daraus nicht die nötigen begrifflichen und terminologischen Konsequenzen gezogen: „[...] mais encore faut-il s'entendre sur le mot frontière, limite, muraille, ou tout autre que l'on voudra employer. Entre le basque et le béarnais vous trouverez en effet une limite en quelque sorte matérielle et tangible, parce que ces langues sont absolument différentes; mais, quand il

s'agit de langues ou de dialectes de même origine, la limite sera une zone intermédiaire. Il y aura transition et non un brusque saut“ [... doch müßte man sich hinsichtlich der Bedeutung von *frontière, limite, muraille* oder anderen Wörtern verständigen. Zwischen dem Baskischen und dem Bearnesischen kann man eine in gewisser Weise materielle und greifbare Grenze finden, weil diese beiden Sprachen von Grund auf verschieden sind; aber wenn es sich um genetisch verwandte Sprachen oder Dialekte handelt, wird die Grenze eine Übergangszone sein. Es wird also ein Hinübergleiten und keinen abrupten Sprung geben.“] (Castets 1888, 307).

Die Abb. 37.18 zeigt deutlich komplex gestaltete Raumkammerungen, die alle traditionell bekannten Dialektlandschaften Oberitaliens und der Südschweiz hervortreten lassen. Die intensivsten Abschottungen finden sich zwischen Graubünden und der Lombardei einerseits sowie zwischen dem Aostatal und Piemont andererseits. Die bekannte Apenninengrenze tritt demgegenüber numerisch zurück. Durch den bei der Erstellung der Abb. 37.18 verwendeten Intervallalgorithmus (MED) wird diese *numerische* Unterschiedlichkeit allerdings im oberen Meßwertebereich zugunsten differenzierterer Gliederungen im mittleren Meßwertebereich *ikonisch* verwischt.

Auf Abb. 37.19 repräsentiert der breite Grenzgürtel zwischen dem Domaine d'Oil und dem Domaine d'Oc das Zentrum der Bildaussage. Daneben sind auch die dunkel und hellblau markierten Verschottungseffekte im Nordosten (Wallonie, Lothringen) und Osten (Frankoprovenzalisch) des ALF-Netzes bemerkenswert. Deutlich ist ferner der Stoß zwischen den Dialektlandschaften Languedoc und Roussillon markiert (dunkelblaue, dicke Grenzsegmente), welche letzterem gegenüber sich sogar das homogenetische Hochkatalanische deutlich absetzt. Die im Bereich des Nordrands des Massif Central in der Gestalt einer querliegenden Mondsichel auftretenden Abschottungen markieren die seit etwa 100 Jahren in der Galloromanistik immer wieder diskutierte Übergangszone (zwischen Oil und Oc) des „Croissant“ (cf. Brun-Trigaud 1990). Sehr deutlich ist auch südlich von Bordeaux die vorhin schon erwähnte Sprachinsel der Petite Gavacherie (ALF-Punkt 635) zu sehen. Als isoglottisch weniger gegliedert (dünne, rote Grenzsegmente) erscheinen drei Zonen: das Zentrum Nordfrankreichs rund um Paris, das südliche Languedoc und die südliche Provence.

3.4. Hierarchisch-agglomerative Klassifikation

Dieses Verfahren ist im Rahmen der Dialektometrie sehr gut einsetzbar. Fast alle der in der einschlägigen Literatur dazu beschriebenen Methoden ergeben brauchbare Resultate. Aus Platzgründen werden hier nur die Ergebnisse (anhand von ALF) der Methode „complete linkage“ vorgeführt. Die Kenntnis des Funktionierens von „complete linkage“ muß dabei als bekannt vorausgesetzt werden (cf. Bock 1974, 392 f. und Vogel 1975, 300 f.; zu einer ausführlichen Diskussion im Rahmen der Dialektometrie cf. Goebel 1992, 1993a und b sowie 1997, 2002, 29 f. und 2003, 84 f.). Der Gesamtertrag der Klassifikation nach „complete linkage“ liegt in einem Baumgraphen vor, der „bottom up“ – d. h. in der Richtung von den *Blättern* des Baumes zum *Stamm* – generiert wurde. Die Verzweigungsstruktur des Baumes ist streng *binär* und *hierarchisch*. Aus dialektometrischer bzw. geolinguistischer Sicht interessieren nur bestimmte (in sich kohärente) Abschnitte des Gesamtbaumes („Dendreme“) und deren räumliche Umlegung in das betreffende Sprachatlasnetz („Choreme“). Der Einfachheit halber wurden für das ALF-Netz nur 11 Dendreme (alle in der Nähe des Stammes) ausgewählt und räumlich umgelegt: siehe dazu das Abbildungspaar 37.20 und 37.21. Die dabei jeweils zustandekommenden Raumpartitionen sind aus geolinguistischer Sicht überaus interessant und zugleich hochplausibel. Bemerkenswert ist die große *räumliche Kohärenz* der Choreme: Ausreißer sind sehr selten. Die Dendrogramme können im Prinzip sowohl *synchron* (zu Zwecken der *Dialektklassifikation*) als auch *diachron* (zu Zwecken der *Rekonstruktion* der dialektalen Ausgliederung) interpretiert werden. Dabei entspricht die *diachrone Interpretation* sowohl inhaltlich wie methodisch voll und ganz der *Lexikostatistik*. Dieser in der Fachwelt viel zu wenig bekannte Sachverhalt soll hier besonders hervorgehoben werden (cf. auch Sankoff 1973 sowie Goebel 1992, 70 f. und 1993a, 75 f.).

Die auf den Abb. 37.20 (Baum) und 37.21 (dazu Spatialisierung) aufscheinenden Gruppierungen wurden mittels des von E. Haimers (siehe hier Art. 38) geschriebenen Dialektometrie-Programms VDM („Visual Dialectometry“) erstellt und können die mit der hierarchisch-agglomerativen Klassifikation (mittels Baumgraphen) einerseits und dem Programm VDM andererseits gegebenen Mög-

lichkeiten nur andeuten. Der Einsatzbereich der hierarchisch-agglomerativen Klassifikation ist nämlich durch die Kombination einer Vielzahl von *Ähnlichkeitsmaßen* (hier: RW_{jk} – Relativer Identitätswert) mit einer Vielzahl von *Baumalgorithmen* (hier: Complete Linkage) überaus breitgefächert. Entsprechende Selektionen müssen demnach vom Linguisten hinsichtlich ihrer Nützlichkeit für bestimmte Forschungsziele genau geprüft werden.

An den beiden obersten Astgabeln oder Bifurkationen (A und B) werden mit dem Dendrem-Chorem 1 die katalanischen Lokalekte des Roussillon, das Hochkatalanische und das Hochitalienische vom Rest des ALF-Netztes (Bifurkation B) abgetrennt. Dieser Rest zerfällt seinerseits in das Domaine d'Oc (Bifurkation C: Dendreme-Choreme 2–5) und das Domaine d'Oil (Bifurkation D: Dendreme-Choreme 6–11). Das Domaine d'Oc gliedert sich sehr schlüssig in die Dendreme-Choreme 5 (Gascogne), 4 (Languedoc), 3 (Provence) und 2 (Alpinprovenzalisch, z. T. auf italienischem Staatsboden). Gleiches gilt für das Domaine d'Oil (Bifurkation D). Dabei entsprechen die von der Bifurkation D wegstrebenden Verzweigungen G, H und I der Reihe nach dem „Croissant“, dem Frankoprovenzalischen und dem eigentlichen Domaine d'Oil (Dendreme-Choreme 8–11). Auffällig ist daran die Anbindung des „Croissant“ (Dendrem-Chorem 6) und des Frankoprovenzalischen (Dendrem-Chorem 7) an das Astwerk des Domaine d'Oil. Im Zuge anders angelegter hierarchisch-agglomerativer Analysen können beide einzeln oder getrennt auch an das Domaine d'Oc angeschlossen werden.

Das Domaine d'Oil zerfällt im Rahmen des gewählten Klassifikationskontextes in räumlich überaus kompakte Makrochoreme (8: Pikardie, Wallonie, Nordlothringen; 9: Südlothringen, Franche-Comté; 10: Poitou, Saintonge samt Petite Gavacherie; 11: Rest des Domaine d'Oil). Selbstverständlich kann auch das Dendrem-Chorem 11 bei Bedarf weiter aufgliedert werden.

Unsere bislang mit der hierarchisch-agglomerativen Klassifikation nordfranzösischer (1983b), italienischer (1992 und 1993a, b) und englischer (1997) Daten gemachten (sehr positiven) Erfahrungen werden nunmehr durch die Resultate der Analyse der ALF-Daten – welche zugleich umfangsmäßig das Maximum unserer bisherigen Klassifikationen repräsentieren – vollauf bestätigt.

4. Nachbemerkung

Die Dialektometrie zählt als typodiagnostisches Verfahren unzweifelhaft zur Domäne der linguistischen Klassifikation und/oder Typologie. Von Begrifflichkeit, Methode und Methodologie steht sie den von Greenberg (z. B. 1957, 1960) vorgelegten klassifikatorischen Arbeiten noch am nächsten. Aus der Sicht der Dialektometrie und der ihr zugrundeliegenden methodischen Postulate sind *Klassifikation* und *Typologie* nicht notwendigerweise getrennt zu halten, wie das ansonsten im Rahmen der Linguistik geschieht. Die typologische Erkenntnisabsicht der Dialektometrie zielt auf die Erstellung bzw. Entdeckung *geolinguistischer Organisations- und Musterstrukturen*. Die in der Dialektometrie mittels Klassenbildung (Klassifikation) generierten Typen sind *quantitativ* strukturierte *räumliche Ordnungsmuster* verschiedenen Ranges. Damit setzt die Dialektometrie überkommene sprachgeographische Fragestellungen meist *qualitativer* Ausrichtung mit *quantitativen* Mitteln fort. Insofern diese traditionellen Fragestellungen von *syn-* und *diachroner* Relevanz waren oder sind, trifft das auch auf die Fragestellungen der Dialektometrie zu. Wie schon bei der klassischen Sprachgeographie sind auch bei der Dialektometrie die Sach- und Analogiebezüge zu den Geo- und Sozialwissenschaften unübersehbar. Diese *interdisziplinäre Brücke* sollte mit aller Sorgfalt und Umsicht ausgebaut werden. Siehe dazu beispielsweise den Vergleich zwischen *geo-genetischen* und *geo-linguistischen* Datensynthesen in Goebel 1996. Dialektometrische Methoden können nur sinnvoll auf genuine *Sprachatlasdaten* angewandt werden. Eine Heranziehung oder Anwendung dialektometrischer Methoden auf andere als geolinguistische Datensorten führt automatisch in eine Sackgasse und sollte unterbleiben. Vor der Durchführung dialektometrischer Projekte ist ein sorgfältiges Studium der numerischen und graphischen Methoden anhand der Originalliteratur unerlässlich. Ich schließe mit einer Paraphrase auf ein bekanntes Werk („On ne badine pas avec l'amour“ [„Man spielt nicht mit der Liebe.“]) des französischen Dichters Alfred de Musset (1810–1857): „On ne badine pas avec la *dialectométrie*“.

5. Literatur (in Auswahl)

AIS: Jaberg, Karl/Jud, Jakob (Hrsg.) (1928–1940), *Sprach- und Sachatlas Italiens und der Südschweiz*,

8 vol. Zofingen: Ringier (Neudruck: Nendeln: Kraus 1971).

ALF: Gilliéron, Jules/Edmond, Edmont (Hrsg.) (1902–1910), *Atlas linguistique de la France* [*Sprachatlas Frankreichs*], 10 vol. Paris: Champion (Neudruck: Bologna: Forni 1968–69).

Altmann, Gabriel/Lehfeldt, Werner (1973), *Allgemeine Sprachtypologie. Prinzipien und Meßverfahren*. München: Fink (1973).

Altmann, Gabriel (1985), Die Entstehung diatopischer Varianten. Ein stochastisches Modell. In: *Zeitschrift für Sprachwissenschaft* 4, 139–155.

Atwood, E. B. (1955), The Phonological Division of Belgo-Romance. In: *Orbis* 4, 367–389.

Benzécri, J.-P./Benzécri, F. (1980), *Pratique de l'analyse des données. 1. Analyse des correspondances. Exposé élémentaire* [*Praxis der Datenanalyse. 1. Die Korrespondenzanalyse: Einführung*]. Paris: Dunod.

Bock, Hans Hermann (1974), *Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse)*. Göttingen: Vandenhoeck & Ruprecht.

Brun-Trigaud, Guylaine (1990), *Le croissant: le concept et le mot. Contribution à l'histoire de la dialectologie française au XIX^e siècle* [*Das « Croissant » : der Begriff und das Wort. Ein Beitrag zur Geschichte der französischen Dialektologie im 19. Jahrhundert*]. Lyon: Université-Lyon III Jean Moulin.

Castets, F. (1988), Rezension zu: Gaston Paris, *Les parlers de France* [*Die Dialekte Frankreichs*] [1888]. In: *Revue des langues romanes* 32, 303–314.

Chandon, Jean-Louis/Pinson, Suzanne (1981), *Analyse typologique. Théories et applications* [*Typen-Analyse. Theorien und Anwendungen*]. Paris, New York, Barcelona, Mailand: Masson.

CLAE: Viereck, Wolfgang/Ramesch, Heinrich (Hrsg.) (1991, 1997), *The Computer Developed Linguistic Atlas of England*, 2 Bände, Tübingen: Niemeyer.

Christmann, Hans Helmut (1971), Lautgesetze und Wortgeschichte. Zu dem Satz „Jedes Wort hat seine eigene Geschichte“. In: Stempel, Wolf-Dieter/Coseriu, Eugenio (Hrsg.), *Sprache und Geschichte. Festschrift für Harri Meier zum 65. Geburtstag*. München: Fink, 11–124.

Dickinson, G. C. (1973), *Statistical mapping and the presentation of statistics*, London: Arnold, 2. Auflage.

Durand, Joseph-Pierre (1889), Notes de philologie rouergate (suite) [Anmerkungen zur Philologie des Rouergue, Fortsetzung]. In: *Revue des langues romanes* 33, 47–84.

Ettmayer, Karl von (1924), Über das Wesen der Dialektbildung, erläutert an den Dialekten Frankreichs. In: *Denkschriften der Akademie der Wissenschaften in Wien, phil.-hist. Klasse*, 66. Band, 3. Abhandlung, Wien, 1–56, 7 Tafeln.

Goebel, Hans (1981), La méthode des interpoints appliquée à l'AIS (essai de dialectométrie interponctuelle) [Die Interpunktmethod, auf den AIS angewendet: ein Versuch zur interpunktuellen Dialektometrie]. In: *Mélanges de philologie et de toponymie romanes offertes à Henri Guiter*, Perpignan, 137–172.

Goebel, Hans (1983a), Parquet polygonal et treillis triangulaire. Les deux versants de la dialectométrie interponctuelle [Polygonmosaik und Dreiecksvernetzung. Die zwei Seiten der Zwischenpunktdialektometrie]. In: *Revue de linguistique romane* 47, 353–412.

Goebel, Hans (1983b), „Stammbaum“ und „Welle“. Vergleichende Betrachtungen aus numerisch-taxonomischer Sicht. In: *Zeitschrift für Sprachwissenschaft* 2 (1), 3–44.

Goebel, Hans (1984), *Dialektometrische Studien. Anhand italo-romanischer, rätomanischer und gallo-romanischer Sprachmaterialien aus AIS und ALF*, 3 Bände, Tübingen: Niemeyer.

Goebel, Hans (1987), Points chauds de l'analyse dialectométrique. Pondération et visualisation [Brennpunkte der dialektometrischen Analyse: Gewichtung und Visualisierung]. In: *Revue de linguistique romane* 51, 63–118.

Goebel, Hans (1992), Dendrogramme im Dienst der Dialektometrie. Zwei hierarchisch-agglomerative Klassifikationen von Daten des Sprachatlases AIS. In: Klenk, Ursula (Hrsg.), *Computatio linguae. Aufsätze zur algorithmischen und quantitativen Analyse der Sprache*, Stuttgart: Steiner, 54–73.

Goebel, Hans (1993a), Dialectometry: A Short Overview of the Principles and Practice of Quantitative Classification of Linguistic Atlas Data. In: Köhler, Reinhard/Rieger, Burghard B. (Hrsg.), *Contributions to Quantitative linguistics*. Dordrecht: Kluwer, 277–315.

Goebel, Hans (1993b), Eine hierarchisch-agglomerative Klassifikation von Daten des Sprachatlases AIS. In: Schmidt-Radefeldt, Jürgen/Harder, Andreas (Hrsg.), *Sprachwandel und Sprachgeschichte. Festschrift für Helmut Lüdtke zum 65. Geburtstag*, Tübingen: Narr, 63–78.

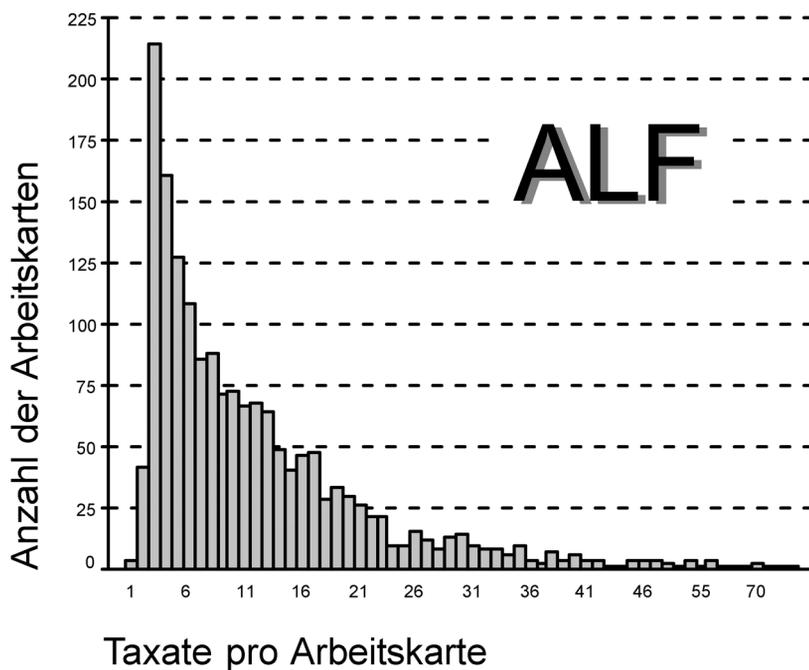
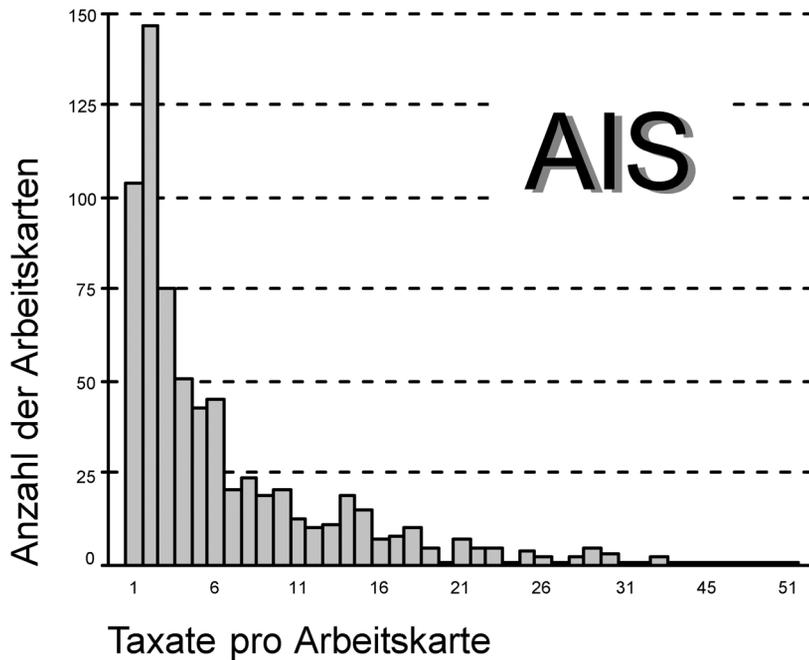
Goebel, Hans (1996), La convergence entre les fragmentations géo-linguistique et géo-génétique de l'Italie du Nord [Die Übereinstimmung zwischen der geo-linguistischen und der geo-genetischen Gliederung Norditaliens]. In: *Revue de linguistique romane* 60, 25–49.

Goebel, Hans (1997), Some Dendrographic Classifications of the Data of CLAE 1 and CLAE 2. In: *CLAE* 2, 23–32.

Goebel, Hans (1998), Due cartine mute poligonizzate per lo spoglio dell'Atlante linguistico italiano [Zwei stumme Karten mit Polygonnetz zur Auswertung des Atlante linguistico italiano]. In: *Bollettino dell'Atlante linguistico italiano, III serie, Dispensa N. 22*, 205–212 (+ 2 Faltkarten).

- Goebel, Hans (2002), Analyse dialectométrique des structures de profondeur de l'ALF [Dialektometrische Analyse der Tiefenstrukturen des ALF]. In: *Revue de linguistique romane* 66, 5–63.
- Goebel, Hans (2003), Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur [Dialektometrische Blicke auf die Daten des Atlas linguistique de la France (ALF): quantitative Relationen und Tiefenstrukturen]. In: *Estudis Romànics* 25, 59–121.
- Greenberg, Joseph H. (1957), The Nature and Uses of Linguistic Typologies. In: *International Journal of American Linguistics* 23, 68–77.
- Greenberg, Joseph H. (1960), A Quantitative Approach to the Morphological Typology of Language. In: *International Journal of American Linguistics* 26, 178–194.
- Guitier, Henri (1973), Atlas et frontières linguistiques [Atlanten und Sprachgrenzen]. In: Straka, Georges/Gardette, Pierre (Hrsg.), *Les dialectes romans de France à la lumière des atlas régionaux* (Colloque de Strasbourg 1971), Paris: Klincksieck, 61–109.
- Guitier, Henri (1984), Confrontation de méthodes géolinguistiques en domaine normand [Vergleich von geolinguistischen Methoden im Bereich des Normannischen]. In: *Actes du 105^e Congrès national des Sociétés savantes (Caen 1980)*, Section de philologie et d'histoire, Paris, tome II, 19–28.
- Haag, Carl (1898), *Die Mundarten des oberen Neckar- und Donaulandes (schwäbisch-alemannisches Grenzgebiet: Baarmundarten)*, Beilagen zum Programm der königlichen Realanstalt zu Reutlingen. Reutlingen: Hutzler.
- Haag, Karl (1930), Die Sprachlandschaften Oberitaliens. In: *Germanisch-romanische Monatsschrift* 18, 458–478.
- Händler, Harald/Wiegand, Herbert Ernst (1982), Das Konzept der Isoglosse. In: Besch, Werner et al. (Hrsg.), *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Berlin/New York: Walter de Gruyter, erster Halbband, 501–527.
- Okabe, Atsuyuki/Boots, Barry/Sugihara, Kokichi (1992), *Spatial Tessellations. Concepts and Applications of Voronoi Diagrams*. Chichester/New York/Brisbane/Toronto/Singapur: Wiley.
- Orth, Bernhard (1974), *Einführung in die Theorie des Messens*. Stuttgart: Kohlhammer.
- Polanco Roig, Lluís B. (1984), Lléngua o dialecte: solucions teòriques i aplicació al cas català [Sprache oder Dialekt: theoretische Lösungen mit einer Anwendung auf das Katalanische]. In: *Actes du XVII^e Congrès international de linguistique e philologie romanes (Aix-en-Provence 1983)*, Aix-en-Provence, Band 5, 13–31.
- Rauh, Jürgen (1998), *Telekommunikation und Raum. Informationsströme im internationalen, regionalen und individuellen Beziehungsgefüge*. Hamburg/Münster/London: LIT.
- Sankoff, David (1973), Mathematical Developments in Lexicostatistic Theory. In: *Current Trends in Linguistics* 11, 93–111.
- Saramago, João (1986), Différenciation lexicale (un essai dialectométrique appliqué aux matériaux de l'A.L.E.) [Lexikalische Unterschiede (ein dialektometrischer Versuch mit Anwendung auf die Materialien des A.L.E.)]. In: *Géolinguistique* 2, 1–31.
- Schiltz, Guillaume (1996), *Der Dialektometrische Atlas von Südwest-Baden (DASB). Konzepte eines dialektometrischen Informationssystems*. 4 Bände, Marburg: Elwert.
- Schneider, Edgar (1984), Methodologische Probleme der Dialektometrie. In: Goebel, Hans (Hrsg.), *Dialectology*, Bochum: Brockmeyer, 314–335.
- Séguy, Jean (1973), La dialectométrie dans l'Atlas linguistique de la Gascogne [Die Dialektometrie im Atlas linguistique de la Gascogne]. In: *Revue de linguistique romane* 37, 1–24.
- Sneath, Peter H. A./Sokal, Robert R. (1973), *Numerical Taxonomy. The Principles and Practice of Numerical Classification*. San Francisco: Freeman.
- Sodeur, Wolfgang (1974), *Empirische Verfahren zur Klassifikation*. Stuttgart: Teubner.
- Thiessen, Alfred H. (1911), Precipitation Averages for large Areas. In: *Monthly Weather Review* 39, 1082–1084.
- TPPSR: Gauchat, Louis/Jeanjaquet, Jules/Tappolet, Ernest (Hrsg.) (1925), *Tableaux phonétiques des patois suisses romands. Relevés comparatifs d'environ 500 mots dans 62 patois-types [Lauttabellen zu den Welschschweizer Dialekten. Vergleichende Aufnahmen von rund 500 Wörtern in 62 ausgewählten Ortschaften]*. Neuchâtel: Attinger.
- Ujomov, A. I. (1965), *Dinge, Eigenschaften und Relationen*. Berlin: Akademie-Verlag.
- Viereck, Wolfgang (1980), Dialektometrie und englische Dialektologie. In: *Grazer linguistische Studien* 11/12, 335–355.
- Viereck, Wolfgang (1985), Linguistic Atlases and Dialectometry: the Survey of English Dialects. In: Kirk, J. M./Sanderson, St. F./Widdowson, J. D. A. (Hrsg.), *Studies in Linguistic Geography*, London: Croom Helm, 94–112.
- Vogel, Friedrich (1975), *Probleme und Verfahren der numerischen Klassifikation unter besonderer Berücksichtigung von Alternativmerkmalen*. Göttingen: Vandenhoeck & Ruprecht.
- Siehe auch die Internetseite: <http://ald.sbg.ac.at/dm/>
- Danksagung: Für die EDV-technische Erstellung aller Abbildungen dieses Beitrags danke ich herzlichst Roland Bauer, Edgar Haimerl, Heide Marie Pamminger und Slawomir Sobota.

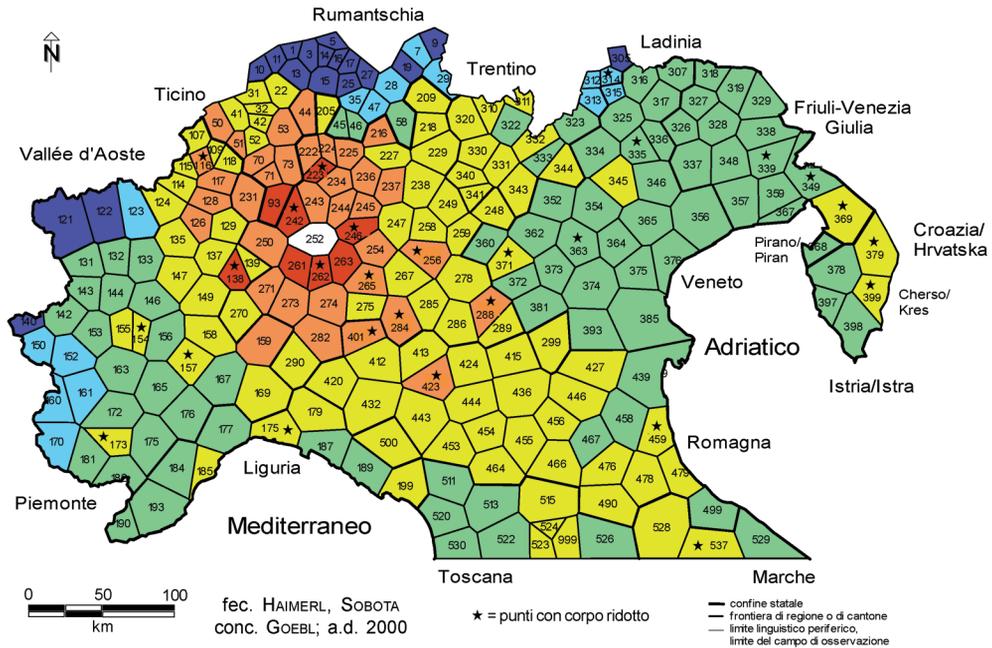
Hans Goebel, Salzburg (Österreich)



Kommentar: Entlang der x-Achse sind die pro Atlaskarte erhobenen Anzahlen der Taxat-types aufgetragen. Diese variieren beim AIS zwischen 1 und 51 und beim ALF zwischen 1 und 91. Entlang der y-Achse sind die Anzahlen der betreffenden Arbeitskarten ablesbar. So gibt es beim AIS 147 bi-nyme (also stets 2 Taxat-types enthaltende) und beim ALF 211 tri-nyme (also stets 3 Taxat-types enthaltende) Arbeitskarten. Zum AIS: cf. Goebel 1984 I, 41 f.; zum ALF: cf. <http://ald.sbg.ac.at/dm/germ/default.htm> und Goebel 2002, 10–15.

Abb. 37.5: Histogramm zu den numerischen Strukturen der Daten des AIS (696 Arbeitskarten) und des ALF (1687 Arbeitskarten).

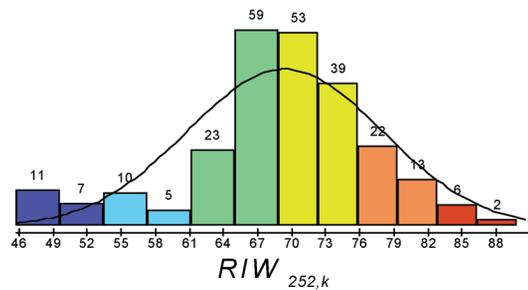
AIS



Choroplethenkarte
MINMWMAX 6-fach
 $R/W_{252,k}$

Häufigkeitsverteilung
MINMWMAX 12-fach

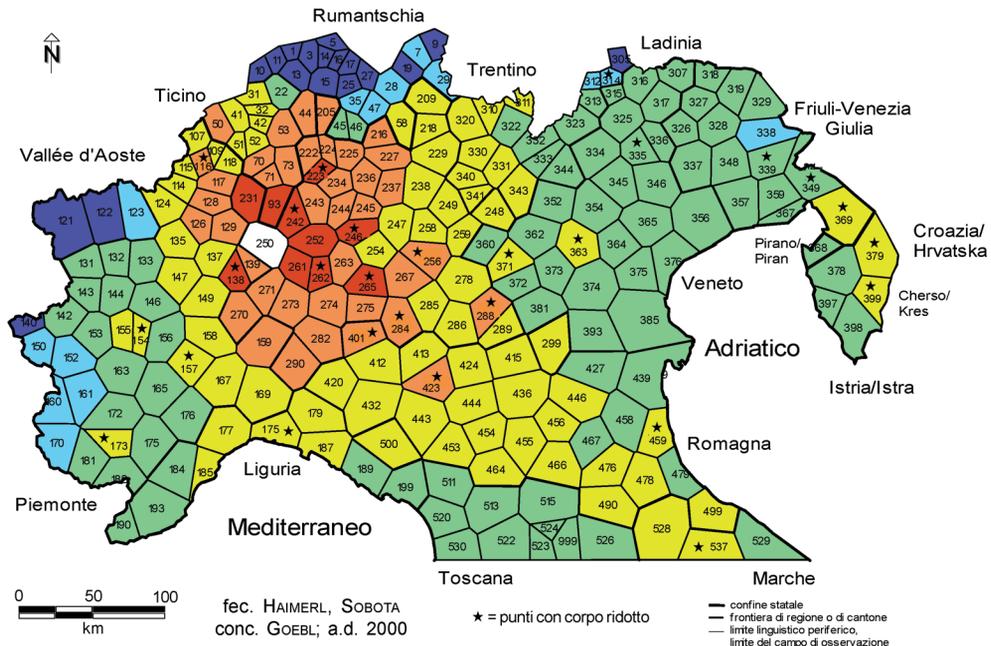
1	45,71 - 53,60	(18)
2	- 61,50	(15)
3	- 69,40	(82)
4	- 76,59	(92)
5	- 83,78	(35)
6	- 90,97	(8)
$\Sigma = 250$		
Meßpunkte		



R/W – Relativer Identitätswert.
Totalkorpus (696 Arbeitskarten).

Abb. 37.6: Ähnlichkeitskarte zum Prüfbezugspunkt 252 des AIS (Monza).

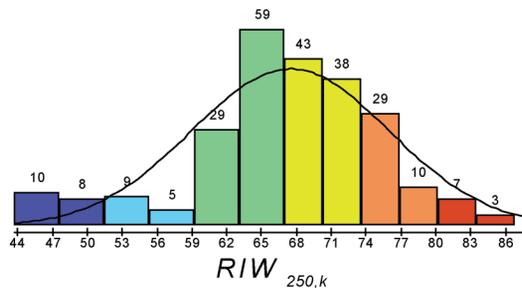
AIS



Choroplethenkarte
MINMWMAX 6-fach
 R/W
 $250, k$

Häufigkeitsverteilung
MINMWMAX 12-fach
 R/W
 $250, k$

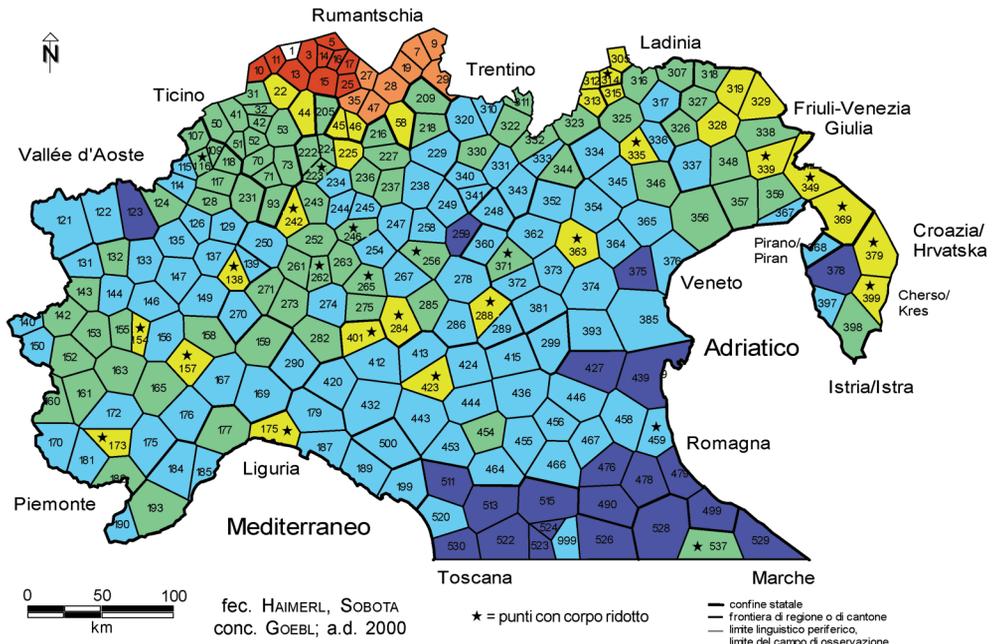
1	43,67	-	51,60	(18)
2		-	59,54	(14)
3		-	67,47	(88)
4		-	74,29	(81)
5		-	81,11	(39)
6		-	87,93	(10)
$\Sigma = 250$				
Meßpunkte				



R/W – Relativer Identitätswert.
Totalkorpus (696 Arbeitskarten).

Abb. 37.7: Ähnlichkeitskarte zum Prüfbezugs punkt 250 des AIS (Magnago).

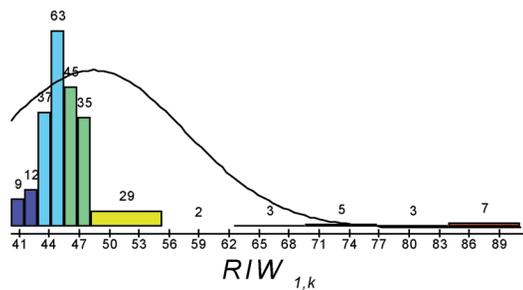
AIS



Choroplethenkarte
MINMWMAX 6-fach
 $R/W_{1,k}$

Häufigkeitsverteilung
MINMWMAX 12-fach

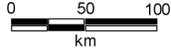
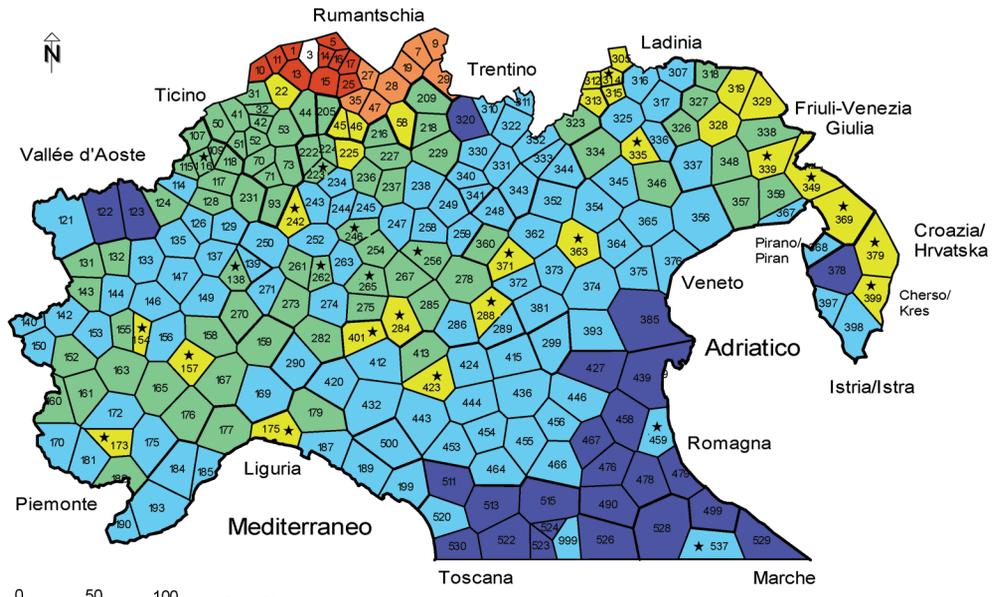
1	40,20 - 42,93	(21)
2	- 45,65	(100)
3	- 48,37	(80)
4	- 62,77	(31)
5	- 77,17	(8)
6	- 91,57	(10)
$\Sigma = 250$		
Meßpunkte		



R/W – Relativer Identitätswert.
Totalkorpus (696 Arbeitskarten).

Abb. 37.8: Ähnlichkeitskarte zum Prüfbezugspunkt 1 des AIS (Brigels/Breil, Graubünden).

AIS



fec. HAIMERL, SOBOTA
conc. GOEBL; a.d. 2000

* = punti con corpo ridotto

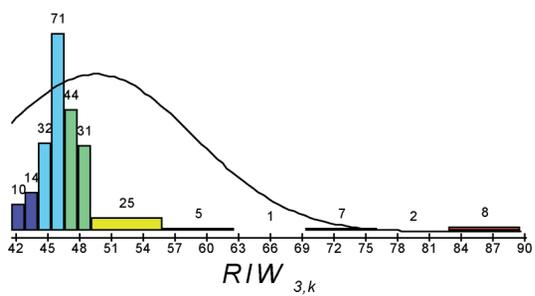
— confine statale
— frontiera di regione o di cantone
— limite linguistico periferico,
limite del campo di osservazione

Choroplethenkarte
MINMWMAX 6-fach
 $RIW_{3,k}$

Häufigkeitsverteilung
MINMWMAX 12-fach

1	41,53 - 44,18	(24)
2	- 46,84	(103)
3	- 49,49	(75)
4	- 63,01	(30)
5	- 76,53	(8)
6	- 90,04	(10)

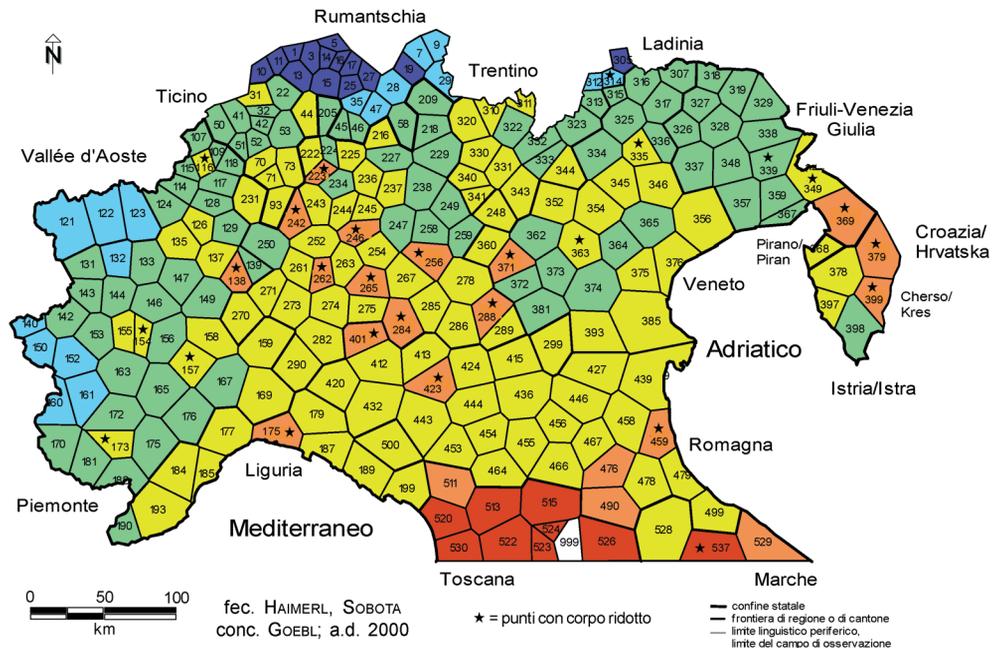
$\Sigma = 250$
Meßpunkte



RIW – Relativer Identitätswert.
Totalkorpus (696 Arbeitskarten).

Abb. 37.9: Ähnlichkeitskarte zum Prüfbezugs punkt 3 des AIS (Pitasch, Graubünden).

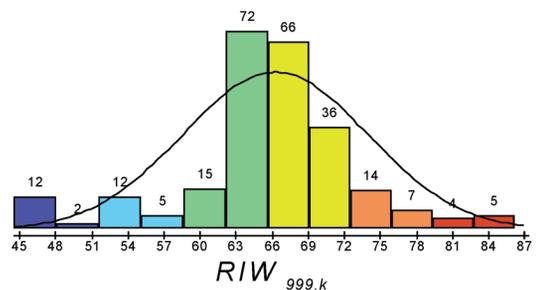
AIS



Choroplethenkarte
MINMWMAX 6-fach
 $R/W_{999,k}$

Häufigkeitsverteilung
MINMWMAX 12-fach

1	44,49 - 51,75	(14)
2	- 59,00	(17)
3	- 66,26	(87)
4	- 73,25	(102)
5	- 80,25	(21)
6	- 87,24	(9)
$\Sigma = 250$		
Meßpunkte		



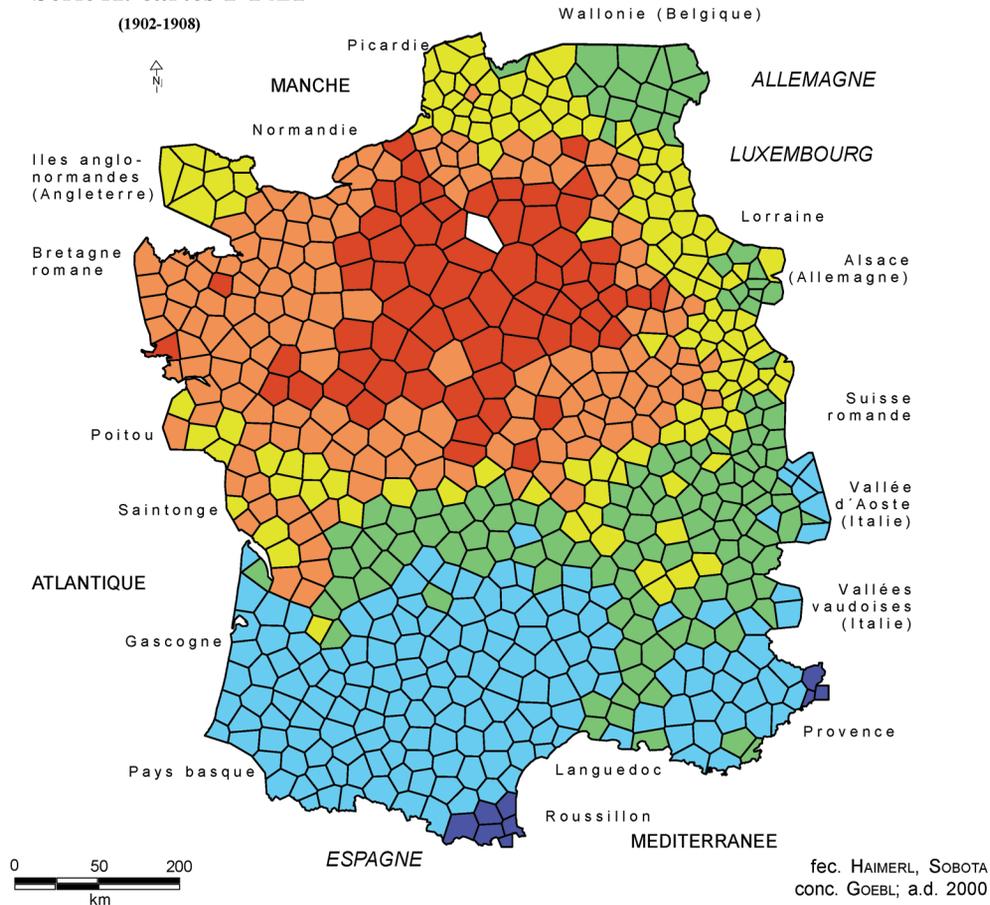
R/W – Relativer Identitätswert.
Totalcorpus (696 Arbeitskarten).

Abb. 37.10: Ähnlichkeitskarte zum künstlichen Prüfbezugspunkt 999 des AIS-Netzes (Standarditalienisch).

ALF

Série A: cartes 1-1421

(1902-1908)

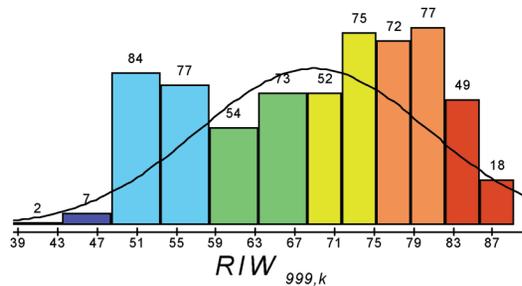


Choroplethenkarte
MINMWMAX 6-fach

Häufigkeitsverteilung
MINMWMAX 12-fach

$R/W_{999,k}$

1	38,52 - 48,65	(9)
2	- 58,78	(161)
3	- 68,91	(127)
4	- 76,11	(127)
5	- 83,30	(149)
6	- 90,50	(67)
		$\Sigma = 640$
Meßpunkte		



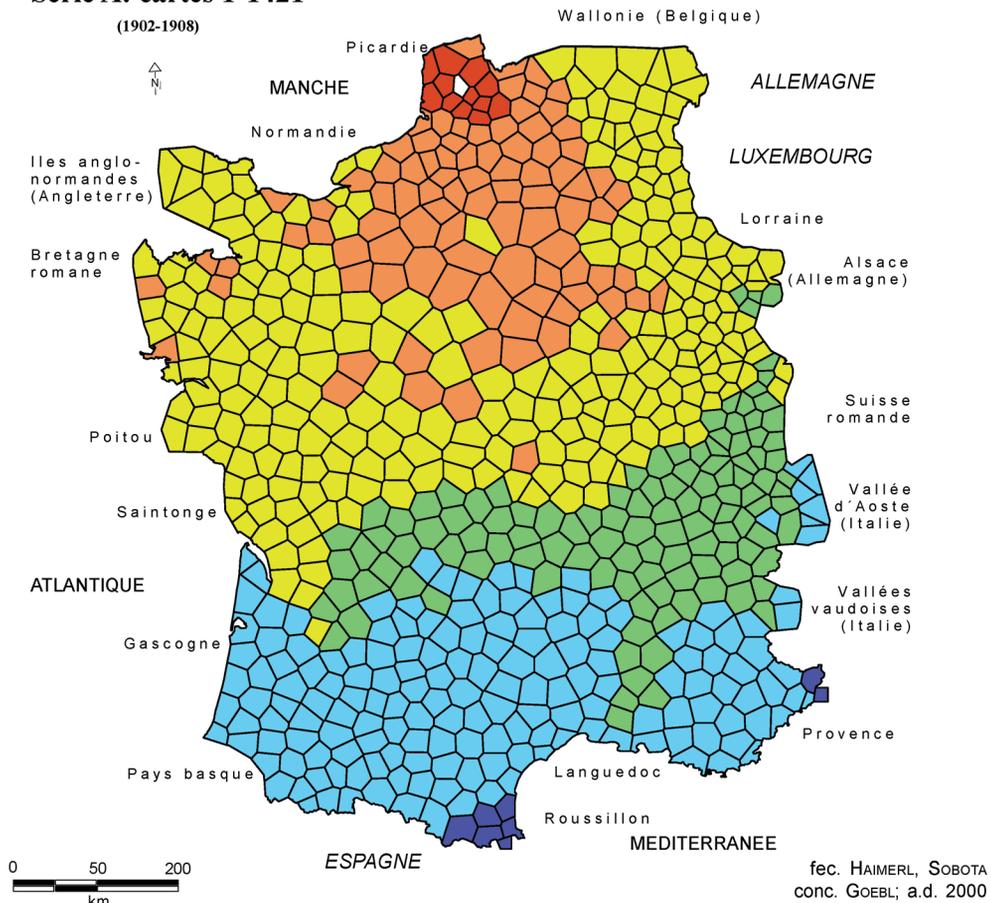
R/W – Relativer Identitätswert.
Totalkorpus (1687 Arbeitskarten).

Abb. 37.11: Ähnlichkeitskarte zum Prüfbezugspunkt 999 des ALF (Hochfranzösisch).

ALF

Série A: cartes 1-1421

(1902-1908)



fec. HAIMERL, SOBOTA
conc. GOEBL; a.d. 2000

Choroplethenkarte
MINMWMAX 6-fach

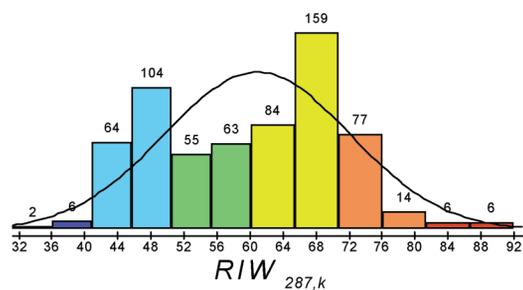
Häufigkeitsverteilung
MINMWMAX 12-fach

$RIW_{287,k}$

1	31,19 - 41,09	(8)
2	- 50,99	(168)
3	- 60,89	(118)
4	- 71,77	(243)
5	- 82,65	(91)
6	- 93,52	(12)

$\Sigma = 640$

Meßpunkte



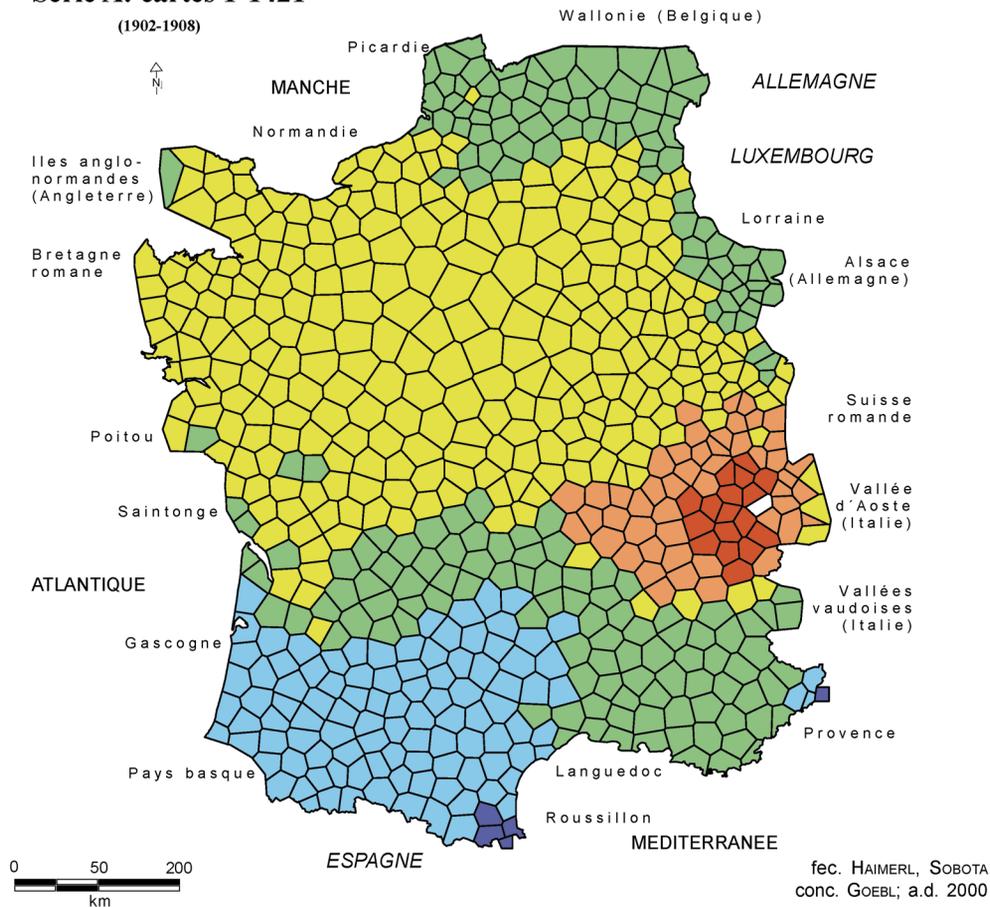
RIW – Relativer Identitätswert.
Totalkorpus (1687 Arbeitskarten).

Abb. 37.12: Ähnlichkeitskarte zum Prüfbezugspunkt 287 des ALF (Teneur, Département Nord).

ALF

Série A: cartes 1-1421

(1902-1908)

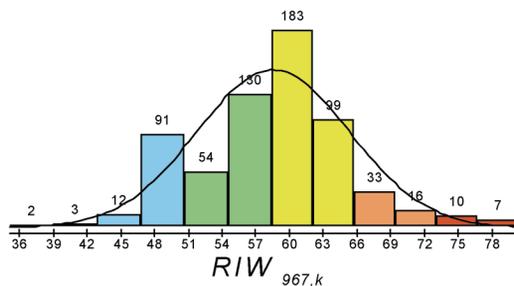


Choroplethenkarte
MINMWMAX 6-fach

$RIW_{967,k}$

1	35,09 - 42,90	(5)
2	- 50,72	(103)
3	- 58,54	(184)
4	- 65,97	(282)
5	- 73,40	(49)
6	- 80,83	(17)
$\Sigma = 640$		
Meßpunkte		

Häufigkeitsverteilung
MINMWMAX 12-fach



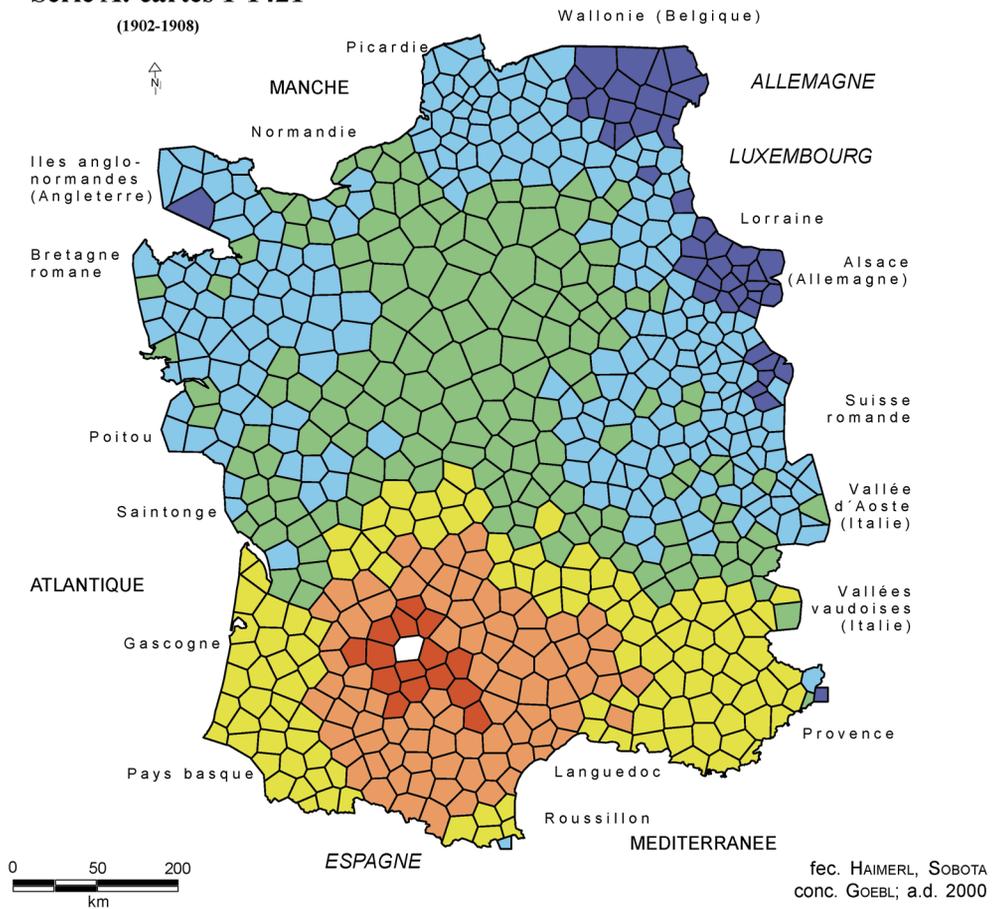
RIW – Relativer Identitätswert.
Totalkorpus (1687 Arbeitskarten).

Abb. 37.13: Ähnlichkeitskarte zum Prüfbezugspunkt 967 des ALF (Chamonix, Département Haute-Savoie).

ALF

Série A: cartes 1-1421

(1902-1908)



fec. HAIMERL, SOBOTA
conc. GOEBL; a.d. 2000

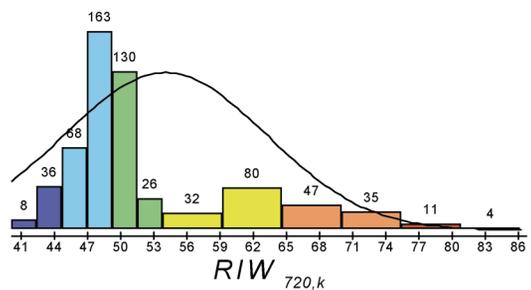
Choroplethenkarte
MINMWMAX 6-fach

Häufigkeitsverteilung
MINMWMAX 12-fach

$R/W_{720,k}$

1	40,11 - 44,76	(44)
2	- 49,41	(231)
3	- 54,07	(156)
4	- 64,94	(112)
5	- 75,81	(82)
6	- 86,68	(15)
		$\Sigma = 640$

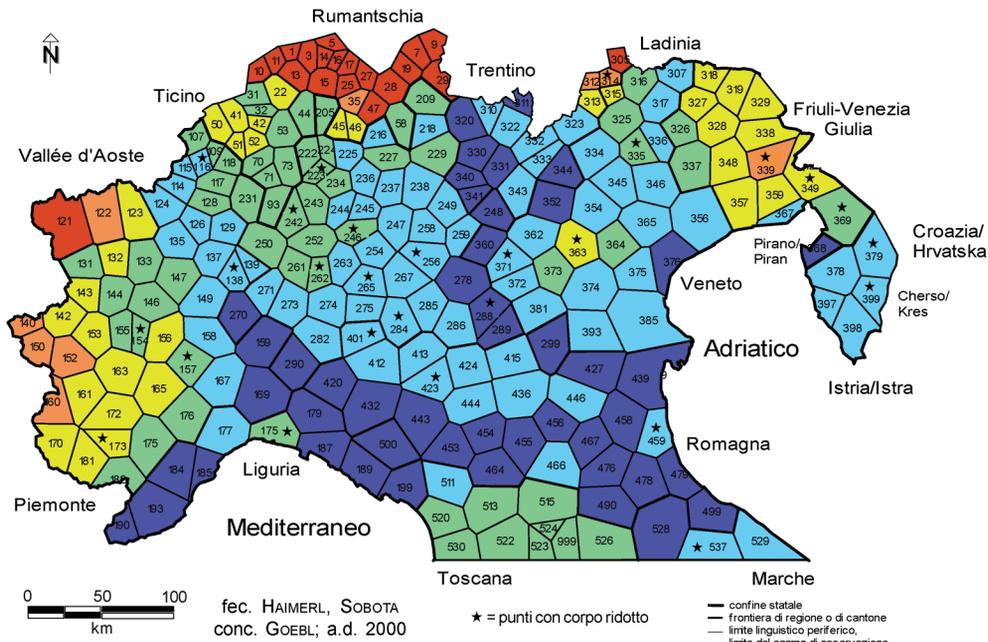
Meßpunkte



R/W – Relativer Identitätswert.
Totalkorpus (1687Arbeitskarten).

Abb. 37.14: Ähnlichkeitskarte zum Prüfbezugspunkt 720 des ALF (Cahors, Département Lot).

AIS



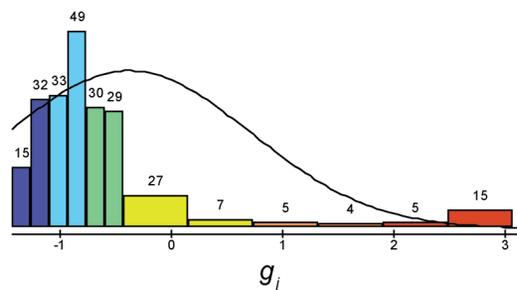
Choroplethenkarte
MINMWMAX 6-fach

Häufigkeitsverteilung
MINMWMAX 12-fach

g_j

1	■	-1,43	-	-1,08	(47)
2	■		-	-0,73	(82)
3	■		-	-0,38	(59)
4	■		-	0,81	(34)
5	■		-	1,99	(9)
6	■		-	3,18	(20)

$\Sigma = 251$
Meßpunkte



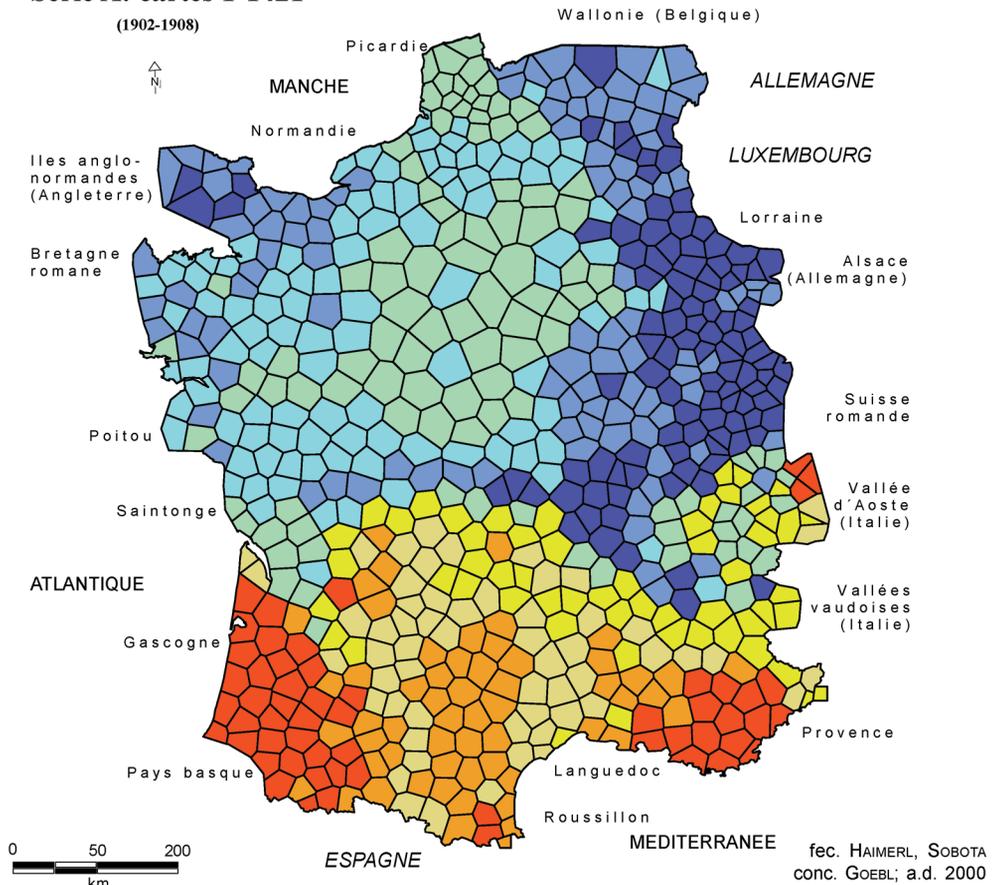
RIW – Relativer Identitätswert.
Totalkorpus (696 Arbeitskarten).

Abb. 37.15: Choroplethenprofil der Schiefen-Werte Oberitaliens und der Südschweiz (nach AIS).

ALF

Série A: cartes 1-1421

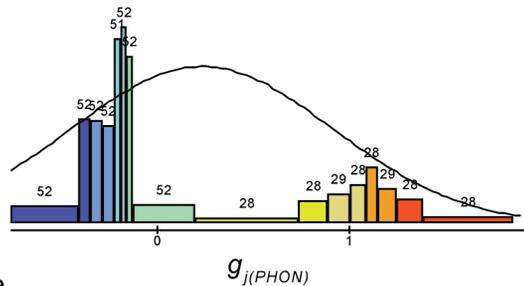
(1902-1908)



Choroplethenkarte
MEDMW 8-fach

Häufigkeitsverteilung
MEDMW 16-fach

	$g_{j(PHON)}$		
1	-0,76	-	-0,34 (104)
2		-	-0,21 (104)
3		-	-0,14 (103)
4		-	0,23 (104)
5		-	0,92 (56)
6		-	1,13 (57)
7		-	1,30 (57)
8		-	1,91 (56)
$\Sigma = 641$ Meßpunkte			



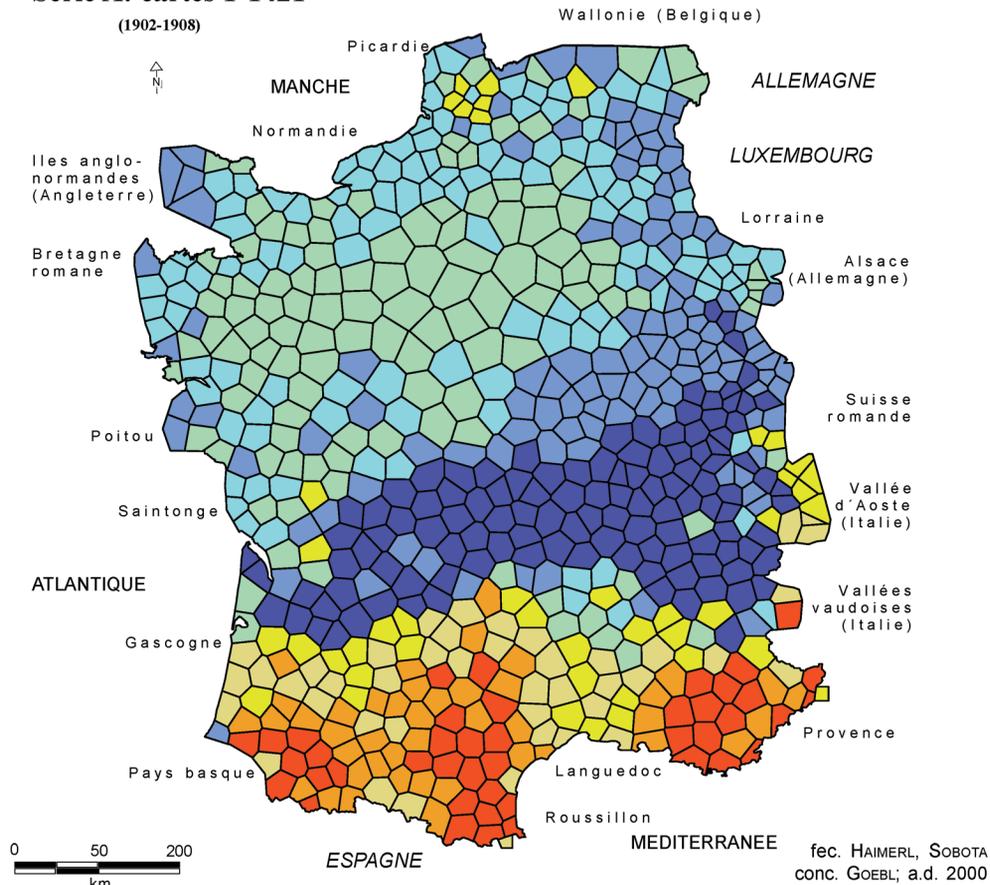
RIW – Relativer Identitätswert.
Teilkorpus: Phonetik (1117 Arbeitskarten).

Abb. 37.16: Choroplethenprofil der Schiefen-Werte der Galloromania (nach ALF).

ALF

Série A: cartes 1-1421

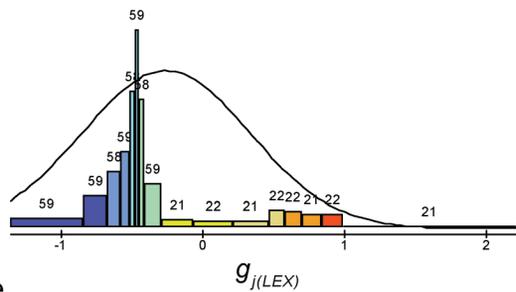
(1902-1908)



Choroplethenkarte
MEDMW 8-fach

Häufigkeitsverteilung
MEDMW 16-fach

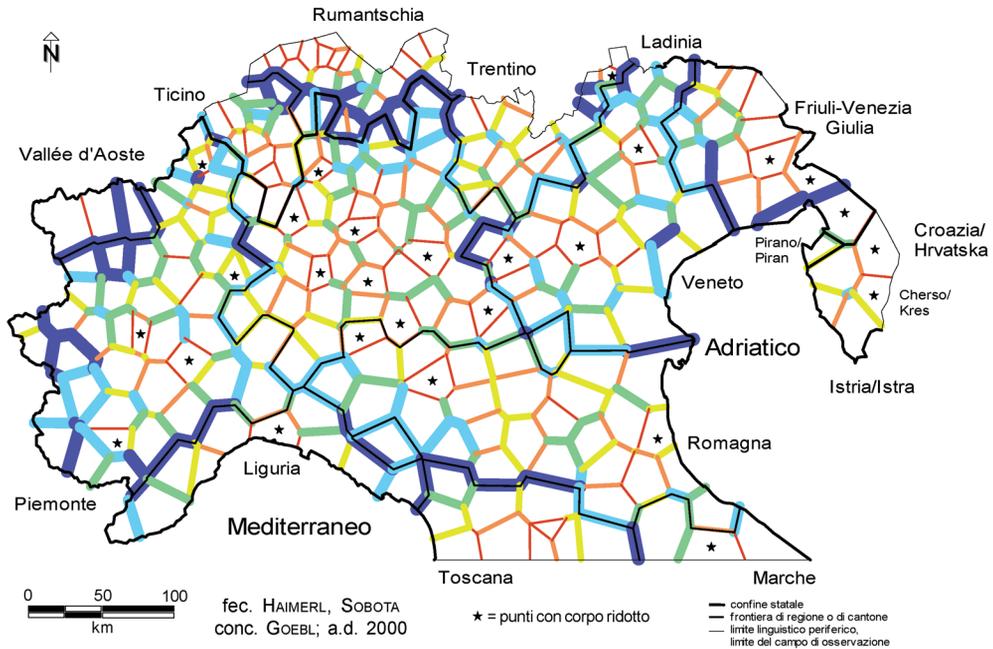
$g_{j(LEX)}$	
1	-1,36 - -0,67 (118)
2	- -0,50 (117)
3	- -0,44 (117)
4	- -0,27 (117)
5	- 0,25 (43)
6	- 0,62 (43)
7	- 0,88 (43)
8	- 2,27 (43)
$\Sigma = 641$ Meßpunkte	



R/W – Relativer Identitätswert.
Teilkorpus: Lexikon (463 Arbeitskarten).

Abb. 37.17: Choroplethenprofil der Schiefen-Werte der Galloromania (nach ALF).

AIS



fec. HAIMERL, SOBOTA
conc. GOEBL; a.d. 2000

* = punti con corpo ridotto

— confine statale
— frontiera di regione o di cantone
— limite linguistico periferico,
limite del campo di osservazione

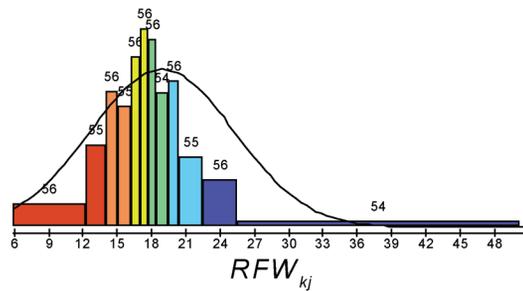
ZP-Karte
MED 6-fach
 R/W_{kj}

Häufigkeitsverteilung
MED 12-fach

1	5,86	-	14,12	(111)
2		-	16,42	(111)
3		-	18,05	(112)
4		-	19,91	(110)
5		-	23,02	(111)
6		-	50,88	(110)

$\Sigma = 665$

Grenzsegmente



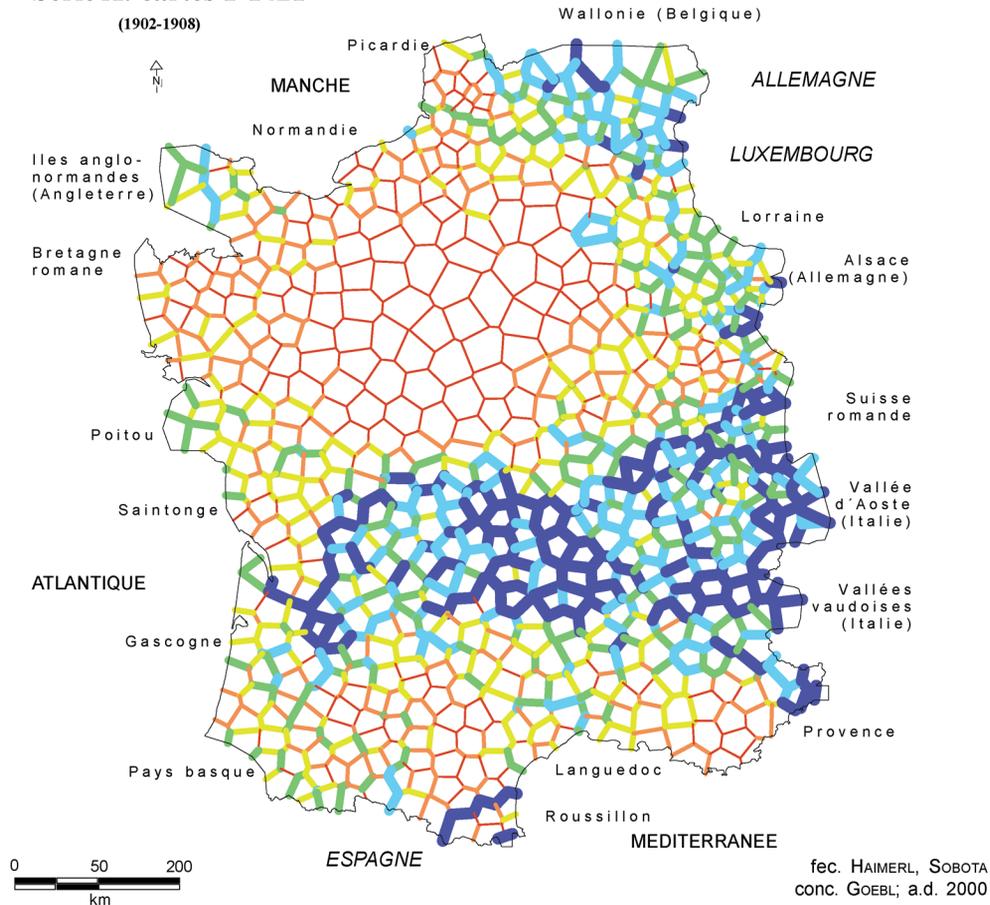
R/W – Relativer Identitätswert.
Totalkorpus (696 Arbeitskarten).

Abb. 37.18: Zwischenpunkt-, Grenzsegment- oder Schottenkarte Oberitaliens und der Südschweiz (nach AIS).

ALF

Série A: cartes 1-1421

(1902-1908)



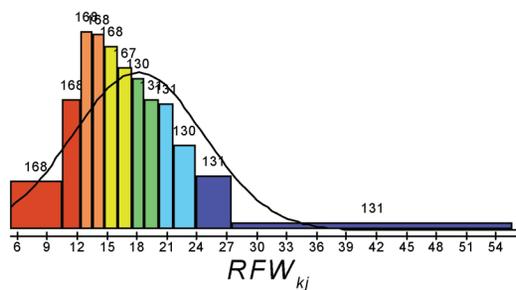
ZP-Karte
MEDMW 6-fach
 RFW_{kj}

Häufigkeitsverteilung
MEDMW 12-fach

1	5,29	-	12,51	(336)
2		-	15,13	(336)
3		-	18,09	(335)
4		-	20,94	(261)
5		-	24,86	(261)
6		-	56,80	(262)

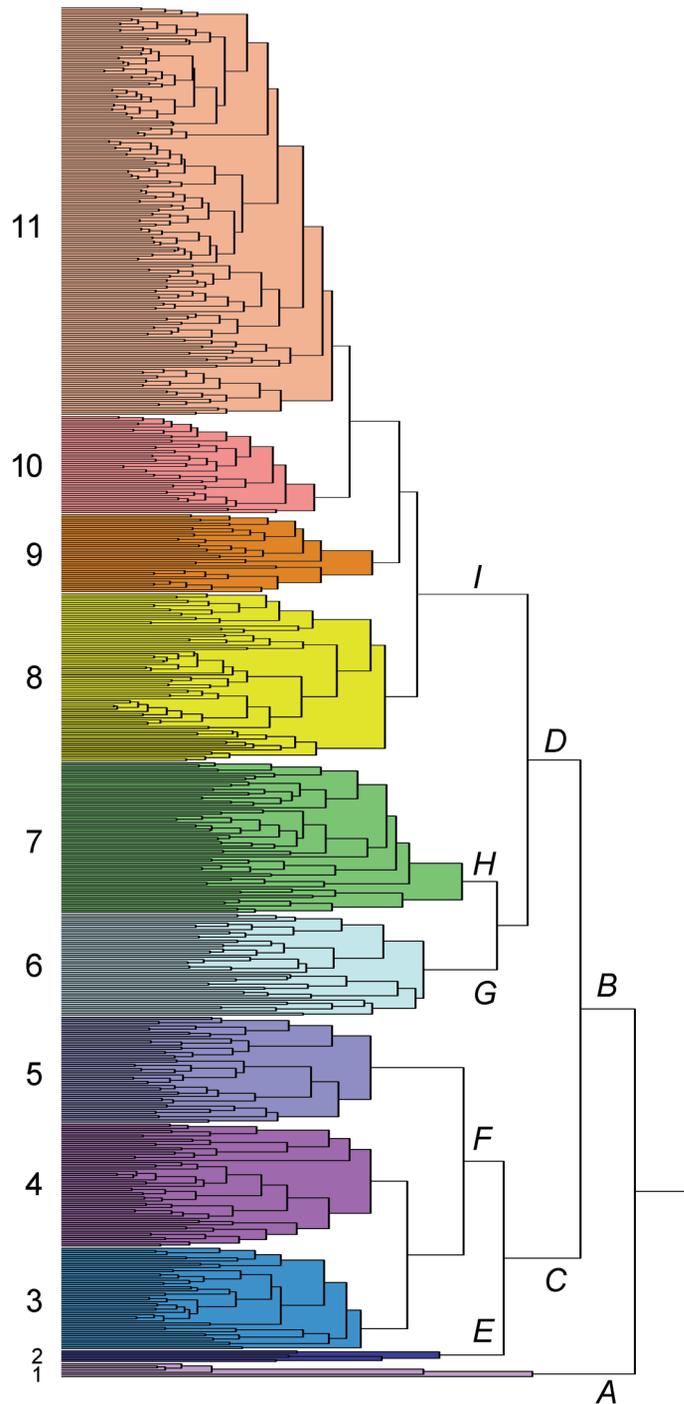
$\Sigma = 1791$

Grenzsegmente



RFW – Relativer Fündigkeitswert.
Totalkorpus (1687 Arbeitskarten).

Abb. 37.19: Zwischenpunkt-, Grenzsegment- oder Schottenkarte der Galloromania (nach ALF).



R/W – Relativer Identitätswert.
 Totalkorpus (1687 Arbeitskarten).
 A ... I: Bifurkationen.

Abb. 37.20: Hierarchisch-agglomerative Klassifikation der ALF-Daten (nach „Complete Linkage“).
 Dendrogramm („Stammbaum“) mit elf (illustrationshalber ausgewählten) Dendremen.

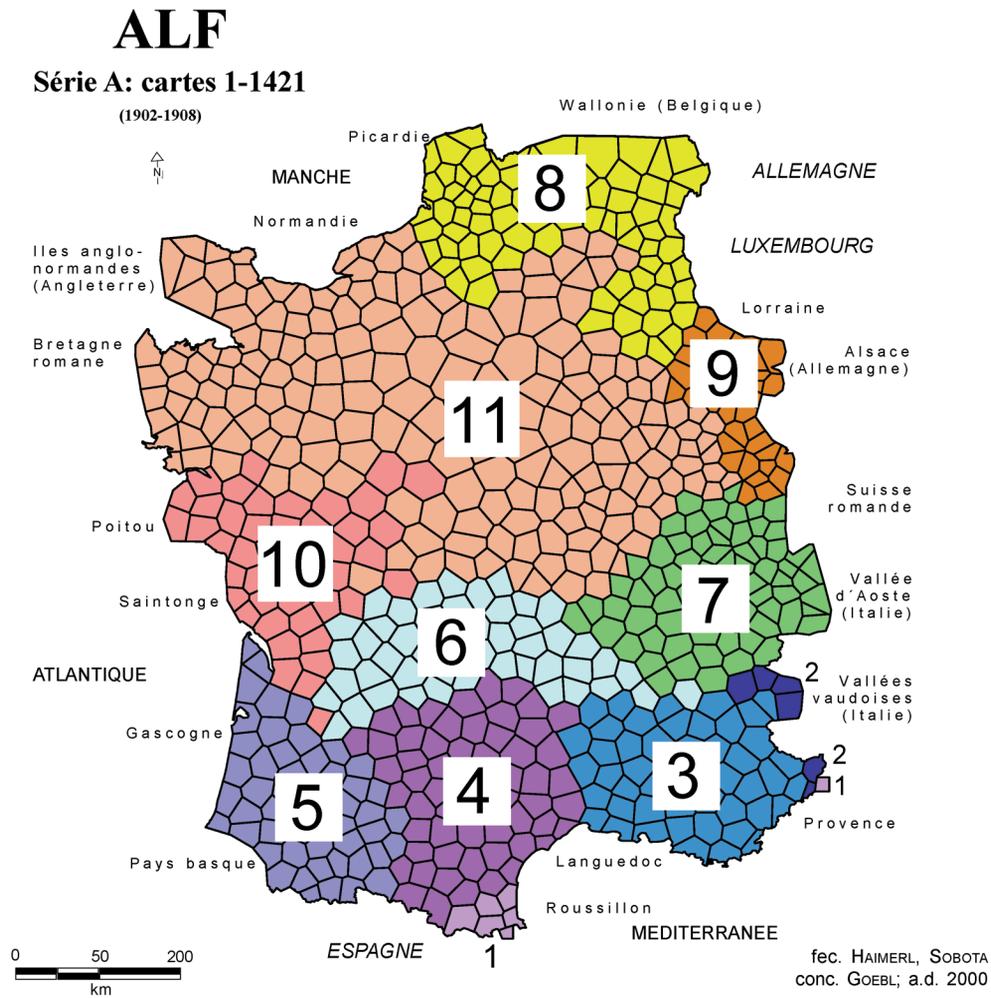


Abb. 37.21: Hierarchisch-agglomerative Klassifikation der ALF-Daten (nach „Complete Linkage“). Räumliche Umlegung (Chorematisierung) der elf Dendreme von Abb. 37.20.