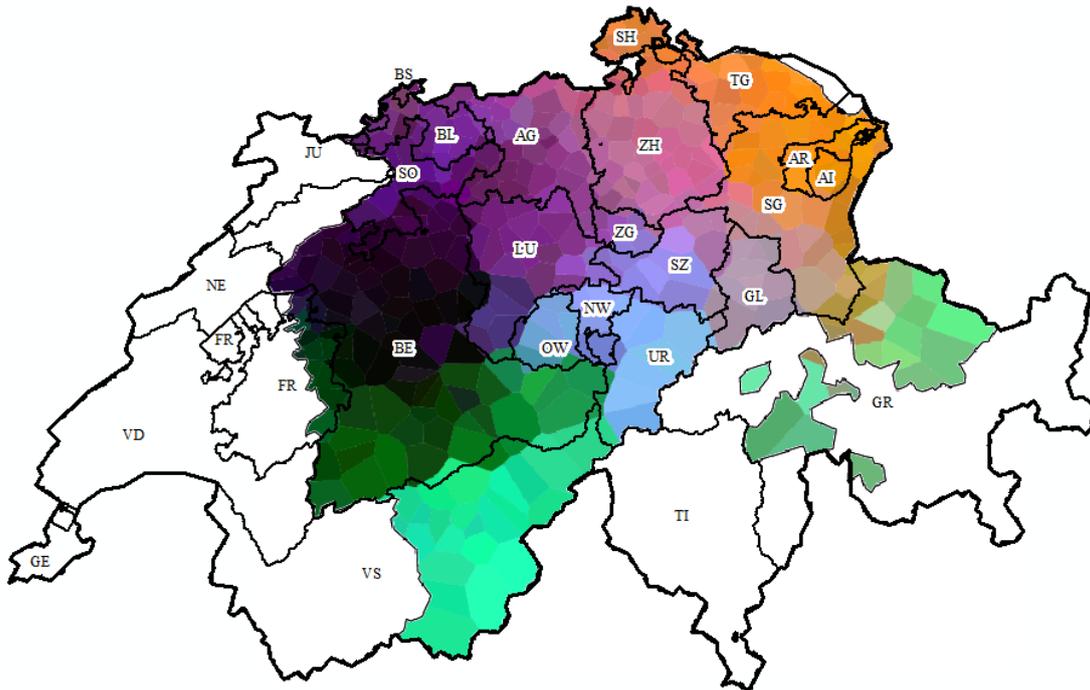


Dialektometrische Analyse und Visualisierung von schweizerdeutschen Dialekten auf verschiedenen linguistischen Ebenen



Masterarbeit GEO 511
Spezialisierung in Geographischer Informationswissenschaft

Sandra Kellerhals
09-115-452

Geographisches Institut, Universität Zürich

Eingereicht am 30. April 2014

Betreuung & Fakultätsvertretung:
Prof. Dr. Robert Weibel

Externe Betreuung:
Prof. Dr. Elvira Glaser, Deutsches Seminar, Universität Zürich
Dr. des. Philipp Stöckle, Deutsches Seminar, Universität Zürich
Dr. Yves Scherrer, LATL, Centre universitaire d'informatique, Universität Genf

Kontakt

Sandra Kellerhals, sandra.kellerhals@gmail.com

Prof. Dr. Robert Weibel
Universität Zürich
Geographisches Institut
Winterthurerstrasse 190
CH-8057 Zürich

Prof. Dr. Elvira Glaser
Universität Zürich
Deutsches Seminar
Schönberggasse 9
CH-8001 Zürich

Vorwort und Danksagung

Schweizerdeutsche Dialekte sind dank ihrer Vielfältigkeit ein interessantes Forschungsobjekt. Die Vorstellung, die aggregierten Unterschiede in den Schweizer Dialekten visuell darstellen zu können, reizte mich und wem immer ich von meinem Vorhaben erzählte, der war begeistert von der Idee. Da die meisten Personen in der Deutschschweiz täglich mit der Unterschiedlichkeit in den Dialekten konfrontiert sind, besteht auch ein allgemeines Interesse an deren Erforschung. Warum aber sollte ich mich als Geographiestudentin mit Dialekten befassen? Es ist eine Weiterführung der interdisziplinären Zusammenarbeit zwischen dem Deutschen Seminar und dem Geographischen Institut der Universität Zürich. Mit Hilfe von Werkzeugen aus der geographischen Informationswissenschaft können die grossen Mengen an linguistischen Daten für die dialektometrischen Analysen aufbereitet werden. Somit wird in dieser Arbeit die geographische Informationswissenschaft in die Disziplin der Linguistik exportiert und es wird am Beispiel der schweizerdeutschen Dialekte gezeigt, dass geographische Informationssysteme auch für nicht primär geographische Forschungsrichtungen wie die Linguistik von Bedeutung sein können.

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mich bei der Erstellung dieser Arbeit unterstützt haben. Ein grosser Dank gilt Prof. Dr. Robert Weibel für die Idee zu diesem Thema und die intensive Unterstützung und Beratung während der ganzen Dauer der Arbeit. Bei Frau Prof. Dr. Elvira Glaser möchte ich mich herzlich für die Ermöglichung des Themas und die Hilfe bei allen linguistischen Fragestellungen bedanken. Ebenso geht ein grosses Dankeschön an Dr. des. Philipp Stöckle für die tatkräftige Unterstützung, u.a. beim digitalisieren der Karten und die Beratung und Hilfe bei den unterschiedlichsten Fragen, insbesondere bei der linguistischen Interpretation der Resultate. Dr. Yves Scherrer danke ich für die Weitergabe seiner dialektometrischen Erfahrungen, der vorhandenen Daten und Python-Skripte sowie für die Ideen, Hinweise und die Beantwortung von technischen Fragen. Weiter bedanke ich mich bei Gabriela Bart für die Hilfe mit den SADS-Daten und bei Ronald Schmidt für die technische Unterstützung und das zeitweise Zurverfügungstellen des Institut-Laptops. Prof. Dr. Hans Goebl und Pavel Smecka von der Universität Salzburg gilt ein besonderer Dank für die Einführung in VDM und das Zurverfügungstellen der Anwendung. Çağrı Çöltekin von der Universität Groningen danke ich herzlich für die technische Unterstützung mit Gabmap. Der letzte Dank gilt allen weiteren Personen deren Wissen, Tipps oder motivierende Worte zum Gelingen dieser Arbeit beigetragen haben.

Zusammenfassung

Sprachliche Phänomene werden in der Linguistik in verschiedene Ebenen gegliedert. Eine davon ist die Lexik, welche den Wortschatz einer Sprache beschreibt. Die syntaktische Ebene erforscht grammatische Strukturen auf Satzebene, die Morphologie auf Wortebene und die Phonetik auf Lautebene.

Die vorliegende Arbeit aggregiert und visualisiert sprachliche Unterschiede zwischen Orten in der Deutschschweiz anhand einer Vielzahl von Phänomenen aus den beschriebenen vier Ebenen. Sie generiert für jede Ebene Dialektlandschaften und untersucht, wie diese sich unterscheiden und welche Zusammenhänge zwischen ihnen bestehen. Weiter interessiert sie sich für die Dialektlandschaft der Deutschschweiz, die entsteht, wenn linguistische Phänomene aus all den beschriebenen Ebenen aggregiert betrachtet werden. Daraus leitet sich eine zusätzliche Fragestellung ab: Welche einzelnen sprachlichen Phänomene determinieren diese gefundene Dialektlandschaft hauptsächlich und aus welcher linguistischen Ebene stammen sie vorwiegend?

Für die Beantwortung dieser Fragen stehen mit dem Sprachatlas der Deutschen Schweiz (SDS) und dem Syntaktischen Atlas der Deutschen Schweiz (SADS) zwei unterschiedliche Datensätze zur Verfügung. Der Datensatz des SDS ist zwischen 1939 und 1958 erhoben worden und beinhaltet Karten zu lexikalischen, morphologischen und phonetischen Phänomenen aus jeweils 565 Orten. Die in den frühen 2000er Jahren in 383 Orten erhobenen Daten des SADS decken die syntaktischen Phänomene ab.

Die beiden Rohdatensätze werden mit Hilfe von Werkzeugen aus der geographischen Informationswissenschaft (ArcMap, Python-Skripts) so aufbereitet, dass sie sich vergleichen und mittels gängiger dialektometrischer Methoden analysieren lassen. Die Unterschiede in den Datensätzen erfordern für deren Vergleich einige Anpassungen. Dazu gehören die Bildung eines einheitlichen Ortspunktenetzes mit 378 Orten und die Anpassung der Antwortstruktur des SADS, der die Antworten mehrerer Gewährspersonen pro Ort beinhaltet, an diejenige des SDS, mit einer älteren Person pro Ort. Dieses Vorgehen ermöglicht erstmals einen direkten Vergleich der sprachlichen Variation der vier linguistischen Ebenen Lexik, Morphologie, Phonetik und Syntax für schweizerdeutsche Dialektdaten. In bereits durchgeführten dialektometrischen Experimenten mit schweizerdeutschen Dialekten sind die syntaktischen Daten des SADS nicht berücksichtigt (Kelle 2001) oder nicht direkt mit den anderen Ebenen verglichen worden (Scherrer 2012).

Die Resultate werden einerseits visuell mittels Parameterkarten, Multidimensionaler Skalierungen und hierarchischer Clusteranalysen präsentiert. Andererseits geben Korrelationsanalysen zwischen den linguistischen Distanzwerten der einzelnen Ebenen Auskunft über deren mögliche Zusammenhänge. Um herauszufinden, welche einzelnen sprachlichen Phänomene die gefundene Dialektlandschaft der Deutschschweiz hauptsächlich determinieren, wird zuerst eine hierarchische Clusteranalyse durchgeführt und anschliessend die bestimmenden Elemente pro Cluster gesucht (Cöltekin et al. 2012).

Die Resultate verdeutlichen, dass es auf syntaktischer Ebene, im Vergleich zu den anderen Ebenen, schwierig ist, klar trennbare Dialektregionen zu finden. Auch zeigen die Korrelationsanalysen zwischen den linguistischen Distanzwerten der einzelnen Ebenen, dass

sich die sprachliche Variation auf der syntaktischen Ebene am stärksten von der jeweiligen Variation der drei anderen Ebenen unterscheidet. Während die Varianz der linguistischen Distanzen in der lexikalischen Ebene jeweils rund zwei Drittel derjenigen der morphologischen bzw. phonetischen Ebene erklärt, ist es bei der syntaktischen Ebene jeweils nur rund ein Drittel der Varianz, welche durch die jeweilige Varianz aus einer der anderen drei Ebenen erklärt wird. Zur Beantwortung der Frage, welche einzelnen Phänomene die gefundene Dialektlandschaft der Deutschschweiz hauptsächlich determinieren, konnte die Methode von Cöltekin et al. (2012) erfolgreich angewandt werden. Das resultierende Set an Variablen enthält, gemessen an der gesamten Anzahl einflussender Variablen pro Ebene, den grössten Anteil aus der lexikalischen und den kleinsten Anteil aus der syntaktischen Ebene.

Abstract

In linguistics, phenomena can be classified into different levels. As one of these levels, Lexis relates to the vocabulary. The syntactical level studies the word order, morphology the word structure and phonetics the sounds of a language. Considering a multitude of linguistic phenomena for each of the four levels described, this thesis aggregates and visualizes the linguistic differences between locations in German-speaking Switzerland. It investigates the differences and correlations of the generated dialectal landscapes for each level. Furthermore, the study examines the landscape of German-speaking dialects that emerges when linguistic phenomena from all the levels are aggregated. This leads to an additional question: Which individual phenomena mainly determine the detected dialectal landscape of German-speaking Switzerland and from which linguistic level do they originate predominantly?

To answer these questions the study relies on two different data sets: The Swiss German Dialect Atlas («Sprachatlas der deutschen Schweiz», SDS) and the yet unreleased Syntactic Atlas of Swiss German Dialects (SADS). The SDS consists of phonetical, morphological and lexical data, collected from 1939 to 1958 in 565 sites. The SADS data, collected in the early 2000s by the University of Zurich in 383 sites, describe the areal distribution of syntactic construction types in German-speaking Switzerland.

To enable dialectometrical analyses by prevalent methods, the two different data sets are preprocessed by means of Geographical Information Science (ArcMap, Python scripts). The methodological challenges of integrating the two different data sources to enable their comparison include creating a common spatial reference frame for the survey sites, as well as reducing the SADS data, which has multiple respondents per site, to the single-respondent structure of the SDS. The procedure enables the first direct comparison of the linguistic variation at the lexical, morphological, phonetical and syntactical level in German-speaking Switzerland. Previous dialectometrical studies in German-speaking Switzerland have not considered the syntactic level at all (Kelle 2001) or have not directly compared it with the other linguistic levels (Scherrer 2012).

The results will be presented visually by parameter maps, as well as maps created by multidimensional scaling and hierarchical clustering. Furthermore, the linguistic distance measures of the four levels will be compared by correlation analyses. To find individual phenomena that determine the dialectal landscape of German-speaking Switzerland, a hierarchical clustering is conducted in order to obtain the cluster determinants (Cöltekin et al. 2012).

The results illustrate that it is difficult to find homogenous dialect clusters on the syntactical level. The correlation coefficients of the linguistic distances for the four linguistic levels also indicate that the syntactic level is least strongly associated with all the other levels. The variance of the linguistic distances on the syntactical level explains in each case only around one third of the variance on the morphological, lexical or phonetical level. By contrast, the variance of the linguistic distances in the lexical level explains around two-thirds of the variance of the morphological or phonetical level.

Regarding the question which individual phenomena mainly determine the detected dialectal landscape of German-speaking Switzerland, the method of Cöltekin et al. (2012) was

successfully applied. The lexical level provides the largest share of its variables to the set of cluster determinants. In contrast, very few variables from the syntactical level can be found among those variables that mainly determine the clusters.

Inhalt

1	Einleitung.....	14
1.1	<i>Motivation.....</i>	14
1.2	<i>Problemstellung.....</i>	14
1.2.1	<i>Ziele.....</i>	14
1.2.2	<i>Forschungsfragen.....</i>	15
1.3	<i>Aufbau der Arbeit.....</i>	16
2	Hintergrund.....	17
2.1	<i>Linguistische Begriffe und Grundlagen.....</i>	17
2.2	<i>Datengrundlage.....</i>	18
2.3	<i>Forschungsstand.....</i>	21
2.4	<i>Forschungslücken.....</i>	23
3	Methodik: Die Dialektometrische Verfahrenskette.....	24
4	Vorgehen.....	27
4.1	<i>Datenauswahl und Datenintegration.....</i>	27
4.1.1	<i>SDS.....</i>	27
4.1.2	<i>SADS.....</i>	29
4.1.3	<i>Abgleich der Gewährspersonen.....</i>	29
4.1.4	<i>Abgleich des Ortsnetzes.....</i>	30
4.1.5	<i>Vergleichbarkeit der beiden Datensätze.....</i>	31
4.2	<i>Datenaufbereitung.....</i>	33
4.2.1	<i>Digitalisierung der SDS-Karten.....</i>	33
4.2.2	<i>Datenexport der SADS-Daten.....</i>	35
4.2.3	<i>Umformatierung der Daten.....</i>	36
4.3	<i>Datenverarbeitung und Visualisierung mit VDM.....</i>	39
4.3.1	<i>Parameterkarten.....</i>	39
4.3.2	<i>Hierarchische Clusteranalysen.....</i>	41
4.4	<i>Datenverarbeitung und Visualisierung mit Gabmap.....</i>	42
4.4.1	<i>Hierarchische Clusteranalysen.....</i>	42
4.4.2	<i>Multidimensionale Skalierung (MDS).....</i>	43
4.5	<i>Analysen.....</i>	45
4.5.1	<i>Korrelationsanalysen.....</i>	45
4.5.2	<i>Bestimmende Faktoren der Cluster (Cluster determinants).....</i>	48
5	Methodenkritik.....	50
5.1	<i>Vergleich von VDM und Gabmap.....</i>	50
5.2	<i>Vergleich der Analysemethoden.....</i>	51

6	Resultate.....	53
6.1	<i>Visueller Vergleich der Dialektlandschaften</i>	53
6.1.1	<i>Parameterkarten</i>	53
6.1.2	<i>Multidimensionale Skalierung</i>	62
6.1.3	<i>Hierarchische Clusteranalyse.....</i>	66
6.2	<i>Numerischer Vergleich: Korrelationsanalyse.....</i>	70
6.3	<i>Bestimmende Faktoren der Cluster.....</i>	73
7	Diskussion	80
7.1	<i>Visuelle Vergleiche: Parameterkarten, Multidimensionale Skalierung, hierarchische Clusteranalyse</i>	80
7.2	<i>Numerischer Vergleich: Korrelationsanalyse.....</i>	88
7.3	<i>Bestimmende Faktoren der Cluster.....</i>	89
7.4	<i>Interpretation der Resultate.....</i>	90
8	Fazit.....	92
8.1	<i>Beantwortung der Forschungsfragen</i>	92
8.2	<i>Grenzen.....</i>	94
8.3	<i>Ausblick.....</i>	96
	Literaturverzeichnis.....	97
	Anhang	101
A.	<i>Liste der miteinbezogenen SDS-Karten</i>	101
B.	<i>Liste mit den verwendeten SADS-Fragen</i>	105
C.	<i>Liste mit der Zuordnung der SADS- zu den SDS-Orten</i>	114
D.	<i>Liste mit SDS-Karten mit mehr als 20 Mehrfachvarianten.....</i>	123
E.	<i>Liste mit je zehn bestimmenden Faktoren pro Cluster.....</i>	125
F.	<i>Inhalt der CD.....</i>	128

Abbildungen

Abbildung 1	Ortspunkte-Netz des SDS-Datensatzes (565 Orte).....	20
Abbildung 2	Ortspunkte-Netz des SADS-Datensatzes (378 verwendete Orte).....	20
Abbildung 3	Übereinstimmungen der beiden Ortspunktenetze.....	30
Abbildung 4	Vergleich SADS 119 «bezahlt hat» mit SDS 3261 «gewesen bin».....	33
Abbildung 5	Phonetische Karte «Bett» aus dem ersten Band des SDS.....	34
Abbildung 6	Digitalisierte Karte «Bett» in ArcMap.....	35
Abbildung 7	Ausschnitt aus der Endtabelle der SDS- (links) und SADS-Daten (rechts) für Gabmap.....	37
Abbildung 8	Aufbau der Access Datenbank im Hintergrund von VDM.....	38
Abbildung 9	Ausschnitt aus der resultierenden Excel-Tabelle vor dem Einlesen in die Access Datenbank (Kartenummer, Kartenmerkmal, VDM_ID, Variantennummer).....	38
Abbildung 10	Clusteranalyse in VDM mit den SDS-Daten, 4 Clustern und dem Ward Algorithmus.....	42
Abbildung 11	Probabilistisches Dendrogramm aus Gabmap.....	43
Abbildung 12	MDS Plot mit Daten des SDS (links) bzw. des SADS (rechts).....	44
Abbildung 13	Fuzzy Cluster Karte mit allen Daten aus dem SDS und dem SADS («UPGMA»-Algorithmus) und den vier Farben blau, grau, rosa und gelb.....	45
Abbildung 14	Mittelwerte der linguistischen Distanzen, Klassenbildung «MINMWMAX».....	55
Abbildung 15	Mittelwerte der linguistischen Distanzen, Klassenbildung «gleiche Intervalle».....	56
Abbildung 16	Minima der linguistischen Distanzen, Klassenbildung «MINMWMAX».....	59
Abbildung 17	Vergleich der Minima der linguistischen Distanzen auf morphologischer Ebene, links mit 378, rechts mit 565 Ortspunkten.....	60
Abbildung 18	Vergleich der Minima der linguistischen Distanzen auf lexikalischer Ebene, links mit 378, rechts mit 565 Ortspunkte.....	60
Abbildung 19	Schiefe der Verteilung der linguistischen Distanzen, Klassenbildung «MINMWMAX»...	61
Abbildung 20	MDS aller SDS- und SADS-Daten (oben gross) und der einzelnen Ebenen (unten klein)..	63
Abbildung 21	MDS-Graphiken (zweidimensionale Abbildungen der linguistischen Distanzen der 378 Orte) der vier linguistischen Ebenen.....	65
Abbildung 22	MDS-Karte, «Fuzzy Cluster»-Karte («UPGMA»-Algorithmus), Clusterkarte Ward-Algorithmus (10 Cluster) VDM und Gabmap.....	67
Abbildung 23	Clusterkarte aller SDS- und SADS-Daten, Ward-Algorithmus (10 Cluster) mit zweidimensionaler MDS-Graphik (Gabmap).....	68
Abbildung 24	Clusterkarte aller SDS- und SADS-Daten, Ward-Algorithmus (10 Cluster) mit Dendrogramm (VDM).....	68
Abbildung 25	Clusteranalyse mit Ward-Algorithmus und zehn Clustern der einzelnen linguistischen Ebenen Lexik, Morphologie, Phonetik, Syntax.....	69
Abbildung 26	Vergleich der MDS- Analysen mit Minimales Set – Zufallsset mit 79, 26 und 9 Variablen.....	78
Abbildung 27	Vergleich der Clusteranalysen mit 79, 26 und 9 ausgewählten und zufälligen Variablen...	79
Abbildung 28	Schiefe der Verteilung der Ähnlichkeitswerte aller SDS-Daten (235 Variablen) der vorliegenden Arbeit, Klassenbildung «MEDMW», 6 Klassen.....	82
Abbildung 29	Schiefe der Verteilung der Ähnlichkeitswerte aller SDS-Daten (216 Variablen) aus Goebel et al. (2013, S. 174), Klassenbildung «MEDMW», 6 Klassen.....	82
Abbildung 30	Vergleich der MDS-Analysen angewandt auf den SDS- Datensatz, oben mit 196 Variablen von Scherrer (2012, S. 263) unten mit den 235 Variablen aus dieser Arbeit.....	83
Abbildung 31	Clusterkarte mit SDS-Daten (170 Variablen, 101 Ortspunkte) aus Kelle (2001, S. 28) «Complete Linkage»-Algorithmus (6 Cluster).....	85

Abbildung 32 Clusterkarte mit SDS-Daten (196 Variablen, 565 Ortspunkte) aus Scherrer (2012, S. 260), «Complete Linkage»-Algorithmus (6 Cluster).....	85
Abbildung 33 Clusterkarte mit SDS-Daten (235 Variablen, 565 Ortspunkte), «Complete Linkage»-Algorithmus (6 Cluster), links mit VDM (1 Variante pro Ort), rechts mit Gabmap (mehrere Varianten pro Ort).....	86
Abbildung 34 Clusterkarten Morphologie (links) und Phonetik (rechts), «Complete Linkage»-Algorithmus (6 Cluster) aus Gabmap mit Brünig-Napf-Reuss-Linie (rot).....	86
Abbildung 35 Clusterkarte mit SDS-Daten links aus Scherrer (2012, S. 261) mit 196 Variablen, rechts mit verfeinertem Datensatz (235 Variablen), «WPGMA»-Algorithmus (10 Cluster).....	87
Abbildung 36 «Fuzzy Cluster»-Karten mit «UPGMA» und «WPGMA»-Algorithmus, rechts aus Scherrer (2012, S. 267) mit 118 Variablen und allen akzeptierten Antworten, links mit 108 Variablen und allen natürlichen Antworten.....	87

Tabellen

Tabelle 1 Übersicht Unterschiede SDS und SADS (Quellen: Scherrer 2011, Bart & Glaser 2011, Bucheli et al. 2012).....	18
Tabelle 2 Beispiel für eine Datenmatrix (Wortstellungskarten aus dem SDS).....	24
Tabelle 3 Linguistische Distanzmatrix	25
Tabelle 4 Im Vergleich zur Arbeit von Scherrer (2012) zusätzlich berücksichtigte Karten.....	28
Tabelle 5 Im Vergleich zur Arbeit von Scherrer (2012) ergänzte und getrennte Varianten.....	28
Tabelle 6 Eigenschaften sowie Vor- und Nachteile von VDM und Gabmap	50
Tabelle 7 Vor- und Nachteile der verwendeten Methoden.....	52
Tabelle 8 Identisch klassierte Orte wie in Clusteranalyse aller Daten pro linguistische Ebene	70
Tabelle 9 Lokale Inkohärenz und Cronbachs Alpha der linguistischen Ebenen	71
Tabelle 10 Standardabweichung der Distanzen pro linguistische Ebene	71
Tabelle 11 Pearson Produkt-Moment-Korrelationskoeffizienten und Bestimmtheitsmasse der linguistischen Ebenen untereinander und mit allen Ebenen aggregiert (Gabmap und VDM). Grün: hohe Werte, rot: tiefe Werte, fett: Extremwerte.	72
Tabelle 12 Vergleich der Bestimmtheitsmasse aus den Distanzen mit 565 Orten	73
Tabelle 13 Cluster bestimmende Variablen und linguistische Ebenen.....	74
Tabelle 14 Cluster bestimmende Variablen mit der grössten Differenz zwischen den z-Werten pro Cluster. (**: Variable 6027 weist in zwei verschiedenen Clustern die grösste Differenz zwischen den z-Werten auf).....	75
Tabelle 15 Vergleich der identisch klassierten Orte wie in Clusteranalyse aller Daten: Minimales Set – Zufallsset, mit 79, 26 und 9 Variablen	76
Tabelle 16 Übersicht über traditionelle dialektologische Dialektgrenzen aus Scherrer (2012) nach Hotzenköcherle et al. (1984) und die Übereinstimmungen mit den MDS-Analysen.	84
Tabelle 17 Vergleich mit den Bestimmtheitsmassen aus Spruit (2008)	88
Tabelle 18 Die Cluster bestimmenden Variablen nach linguistischer Ebene	93

Abkürzungen

GIW	Gewichteter Identitätswert
MDS	Multidimensionale Skalierung
NA	Engl. «not available», fehlende Information
RGB	Rot, Grün, Blau (Farbraum)
RIW	Relativer Identitätswert
SADS	Syntaktischer Atlas der deutschen Schweiz
SDS	Sprachatlas der deutschen Schweiz
VDM	Visual DialectoMetry

Kantone der Schweiz

AG	Aargau
AI	Appenzell I. Rh.
AR	Appenzell A. Rh.
BE	Bern
BL	Basel-Landschaft
BS	Basel-Stadt
FR	Freiburg
GE	Genf
GL	Glarus
GR	Graubünden
JU	Jura
LU	Luzern
NE	Neuenburg
NW	Nidwalden
OW	Obwalden
SG	St. Gallen
SH	Schaffhausen
SO	Solothurn
SZ	Schwyz
TG	Thurgau
TI	Tessin
UR	Uri
VD	Waadt
VS	Wallis
ZG	Zug
ZH	Zürich

1 Einleitung

1.1 Motivation

Bisher wurden schweizerdeutsche Dialektdaten vor allem mit Hilfe von traditionellen dialektologischen Methoden, also anhand von einzelnen sprachlichen Phänomenen untersucht und deren geographische Verteilung studiert. Die einzigen Ausnahmen bilden der dialektometrische Versuch von Kelle (2001) sowie Experimente von Scherrer (2012), in denen mit Hilfe dialektometrischer Methoden grössere Anzahlen sprachlicher Phänomene aggregiert und visualisiert wurden. In Zusammenarbeit mit dem Deutschen Seminar der Universität Zürich sollen die erwähnten dialektometrischen Arbeiten im Rahmen dieser Masterarbeit ergänzt und optimiert werden, insbesondere was die Auswahl der in die Analysen einflussenden Daten und das untersuchte Ortsnetz betrifft. Es soll vor allem auch zum ersten Mal versucht werden, schweizerdeutsche Dialektregionen zu definieren, die sich aus allen linguistischen Ebenen, inklusive der Syntax (Satzbau), ergeben. Ausserdem soll die Arbeit direkte Vergleiche zwischen den Dialektlandschaften der syntaktischen mit den anderen linguistischen Ebenen ermöglichen und so neue Erkenntnisse in Bezug auf die Verteilung sprachlicher Distanzen im Raum liefern.

1.2 Problemstellung

1.2.1 Ziele

Zur Analyse der Schweizerdeutschen Dialekte stehen im Moment zwei Datensätze zur Verfügung. Einerseits der Sprachatlas der deutschen Schweiz (SDS) (Hotzenköcherle und Baumgartner 1962-1997, 2003) und andererseits der Syntaktische Atlas der deutschen Schweiz (SADS) (Bucheli und Glaser 2002). Die Daten des SDS wurden von 1939 bis 1958 in der Deutschschweiz erhoben. In acht Bänden dokumentiert dieses Werk über 1500 Karten zu phonetischen (Lautung), morphologischen (Wortbildung) und lexikalischen (Wortschatz) Phänomenen (Hotzenköcherle und Trüb 1962). Da es schwierig ist, syntaktische Phänomene (Satzbau) auf die selbe Art zu erheben wie Lautung oder Wortschatz (Löffler 2003), verfügt der SDS nur über sehr wenige syntaktische Karten. Diese Lücke wird nun der SADS füllen. Die entsprechenden Daten wurden ab dem Jahr 2000 erhoben (Bucheli 2008).

Die vorliegende Arbeit verfolgt zwei Hauptziele:

Einerseits sollen die zur Verfügung stehenden SDS- und SADS-Daten so aufbereitet werden, dass sie sich mit den gängigen dialektometrischen Methoden analysieren und vergleichen

lassen. Insbesondere soll herausgefunden werden, inwiefern sich die *aggregierten linguistischen Distanzen* zwischen den untersuchten Orten, welche sich aus den *syntaktischen Phänomenen* ergeben, von denjenigen aus den anderen linguistischen Ebenen unterscheiden.

Als zweites Hauptziel soll herausgefunden werden, ob sich ein *minimales Set an Variablen* finden lässt, mit dem sich bei einer Visualisierung dieselben Dialektregionen herauskristallisieren, wie bei einem Einbezug aller vorhandenen Variablen. Anders formuliert soll untersucht werden, ob gewisse sprachliche Phänomene besser trennen als andere. Dafür wird zuerst aus allen Variablen eine Dialektlandschaft mit verschiedenen Dialektregionen erstellt und anschliessend eruiert, welche Variablen hauptsächlich für dieses Resultat verantwortlich sind.

1.2.2 Forschungsfragen

Aus den im vorangehenden Abschnitt definierten Zielen lassen sich zwei Hauptforschungsfragen ableiten. Die erste Frage bezieht sich auf das erste Hauptziel:

- 1) *Welche Unterschiede gibt es in der räumlichen Verteilung von aggregierten Distanzen morphologischer, phonetischer, lexikalischer und syntaktischer Phänomene in der Deutschschweiz?*

Ähneln sich die Dialektregionen- oder Landschaften, welche sich aus den einzelnen linguistischen Ebenen ergeben oder fällt eine bestimmte Ebene von den anderen ab? Welche linguistische Ebene ist hauptsächlich für die Dialektlandschaft der Deutschschweiz verantwortlich? Korrelieren Karten bzw. Distanzmatrizen, in denen phonetische, lexikalische, morphologische und syntaktische Phänomene aggregiert wurden am ehesten mit Karten, in denen a) nur phonetische, b) nur morphologische, c) nur lexikalische oder d) nur syntaktische Phänomene berücksichtigt wurden? Insbesondere soll untersucht werden, wie sich die syntaktische Ebene im Vergleich zu den anderen linguistischen Ebenen verhält.

Die zweite Forschungsfrage bezieht sich auf das zweite Hauptziel:

- 2) *Welche linguistischen Phänomene trennen besonders gut?*

Lässt sich ein minimales Set an Variablen finden, das besonders stark mit den aggregierten Phänomenen aller Ebenen korreliert? Falls sich ein solches finden lässt, stammen diese Variablen vorwiegend aus einer bestimmten linguistischen Ebene?

Zusammengefasst sollen die Antworten zu den gestellten Fragen einerseits Aufschluss über den Zusammenhang zwischen den verschiedenen linguistischen Ebenen, in Bezug auf die räumliche Verteilung der aggregierten Distanzen geben und andererseits den Einfluss jeder Ebene sowie einzelner Variablen auf die gesamte Dialektlandschaft der Schweiz aufzeigen.

1.3 Aufbau der Arbeit

Die Arbeit besteht aus acht Kapiteln. Das aktuelle, einleitende Kapitel 1 enthält die wissenschaftliche Motivation, die Formulierung der Ziele sowie die Forschungsfragen.

Das Kapitel 2 führt in die Thematik ein und geht auf den Hintergrund der Arbeit ein. Es enthält Definitionen linguistischer Begriffe, Informationen über die zur Verfügung stehenden Daten und zum Forschungsstand. Im letzten Abschnitt des Kapitels wird auf die Forschungslücken hingewiesen, welche mit dieser Arbeit geschlossen werden sollen.

Das Kapitel 3 gibt zuerst einen Überblick über die Methodik, während Kapitel 4 anschliessend die Vorgehensweise von der Datenauswahl über die Datenaufbereitung bis zur Visualisierung und den verschiedenen Analysemethoden Schritt für Schritt beschreibt. Die Methodenkritik in Kapitel 5 diskutiert die Vor- und Nachteile der verwendeten Dialektometrie-Programme sowie der einzelnen Analysemethoden.

Die Kapitel 6 und 7 zeigen und diskutieren die Resultate aus den im vorangehenden Kapitel beschriebenen Analysen.

Das letzte Kapitel 8 fasst die Forschungsfragen und Antworten nochmals zusammen, zeigt die Grenzen der Arbeit auf und weist auf mögliche zukünftige Forschung hin.

2 Hintergrund

Im Abschnitt 2.1 werden wichtige linguistische Begriffe definiert, während der Teil 2.2 die der Arbeit zugrundeliegenden Daten beschreibt. Abschnitt 2.3 gibt einen kurzen Überblick über die Geschichte der Dialektologie und die dialektometrische Forschung, die sich daraus entwickelt hat. Der letzte Abschnitt 2.4 weist schliesslich auf Forschungslücken hin.

2.1 Linguistische Begriffe und Grundlagen

Bei dieser Studie handelt es sich um eine interdisziplinäre Arbeit, an der Schnittstelle zwischen Geographie und Linguistik. Da sie aus geographischer Sicht verfasst ist, werden hier die wichtigsten linguistischen Begriffe und ihre Verwendung in der Arbeit kurz erläutert.

Gewährsperson

Dies ist eine Person, die für die Datenerhebung der Dialektdaten befragt wurde. Sie ist in der Regel in zweiter Generation ortsfest (Bucheli 2005). Daher wird sie als repräsentativ für den im entsprechenden Ort vorherrschenden Dialekt angesehen.

Linguistische Variable

Eine linguistische Variable ist ein sprachliches Merkmal, das unterschiedliche Ausprägungen (Varianten) annehmen kann. Hier wird der Begriff auch mit einer Arbeitskarte gleichgesetzt. Jede solche Karte beschreibt eine linguistische Variable (Goebel 1982a).

Linguistisches Phänomen

Ein linguistisches Phänomen kann durch mehrere linguistischen Variablen beschrieben werden. Beispielsweise wird in der vorliegenden Arbeit das Phänomen «Relativsatz» durch die linguistischen Variablen «der Mann, dem ich», «die Frau, der ich», «ein Auto, das ich» beschrieben.

Linguistische Ebenen

In der Linguistik wird die Sprache in verschiedene Ebenen gegliedert (Pfister und Kaufmann 2008). In der vorliegenden Arbeit werden die folgenden vier linguistischen Ebenen unterschieden.

Syntax

Die *Syntax* erforscht grammatische Strukturen auf Satzebene (Pfister und Kaufmann 2008).

Morphologie

Die Morphologie beschäftigt sich mit grammatischen Strukturen auf Wortebene. Sie wird auch als Wortgrammatik bezeichnet (Fleischer et al. 2001), im Gegensatz zur Syntax, die als Satzgrammatik zu verstehen ist. In dieser Arbeit schliesst der Begriff Morphologie sowohl die Nominal- als auch die Verbalmorphologie mit ein (Goebel et al. 2013).

Phonetik / Phonologie¹

Die Phonetik untersucht grammatische Strukturen auf Lautebene. Diese Ebene beinhaltet sowohl Vokalquantität (Vokallänge) als auch Vokalqualität, welche die Unterscheidung der qualitativen Eigenschaften von Vokalen beschreibt (z.B. offen/ geschlossen, gerundet/ ungerundet) (Bussmann und Gerstner-Link 2002). Auch der Konsonantismus, welcher das System der Konsonanten einer Sprache beschreibt (Fleischer et al. 2001), ist darin enthalten.

Lexik

Die Lexik beschreibt den Wortschatz einer Sprache, also welche Worte zur Beschreibung eines Begriffes verwendet werden (Fleischer et al. 2001).

2.2 Datengrundlage

Im vorliegenden Kapitel werden die beiden in der Arbeit verwendeten Datensätze, der Sprachatlas der deutschen Schweiz (nachfolgend SDS) und der Syntaktische Atlas der Deutschen Schweiz (nachfolgend SADS) sowie deren Unterschiede vorgestellt.

	SDS	SADS
<i>Vorliegen der Daten</i>	Als Atlaskarten auf Papier, davon 216 Karten digitalisiert (Scherrer 2012)	Filemaker-Datenbank
<i>Erhebungszeitraum</i>	1939-1958	2000-2002
<i>Erhebungsmethode</i>	Mündliche Befragung durch Exploratoren vor Ort mit Fragebuch, über Tage hinweg	4 schriftliche Fragebogen, Übersetzungs-, Ergänzungs- und Ankreuzfragen (Auswahl aller akzeptierten und der natürlichsten Variante)
<i>Informantenauswahl</i>	Ältere Männer mit hand-werklich-landwirtschaftlichem Beruf, 1-4 pro Ort	Männer und Frauen ab 12 Jahren, alle Berufe, total 3185 Gewährspersonen, 3-26 pro Ort (Durchschnitt = 7)
<i>Anzahl Ortspunkte</i>	565	383 (342 davon sind auch SDS-Orte)
<i>Anzahl einbezogene Variablen</i>	235 Karten	108 Fragen zu 52 syntaktischen Phänomenen
<i>Linguistische Ebenen</i>	Phonologie, Morphologie, Lexik	(Morpho-)Syntax

Tabelle 1 Übersicht Unterschiede SDS und SADS (Quellen: Scherrer 2011, Bart & Glaser 2011, Bucheli et al. 2012)

Wie in Tabelle 1 ersichtlich ist, liegen die SDS-Daten als Atlaskarten auf Papier vor. Der gesamte Datensatz umfasst 1548 Sprachkarten, die in acht Atlasbänden von Rudolf Hotzenköcherle und Rudolf Trüb in den Jahren 1962 bis 1997 veröffentlicht wurden (Trüb

¹ Während mit Phonetik die Formseite der Sprachlaute gemeint ist, bezeichnet Phonologie deren Funktion im Sprachsystem (Staffeldt 2010). Obwohl in dieser Arbeit beide Seiten vertreten sind, wird fortan der Einfachheit halber nur noch von «Phonetik» gesprochen.

2003). Die dazugehörigen Daten sind zwischen 1939 und 1958 durch mündliche Befragungen von über 1500 Personen an 565 Orten in der Schweiz erhoben worden (s. Abbildung 1). Die Gewährspersonen waren meist in der Landwirtschaft tätig und beantworteten jeweils rund 2500 Fragen (Christen et al. 2011). Im Normalfall wurde eine Person pro Ort befragt. Konnte diese jedoch nicht über den gesamten Befragungszeitraum zur Verfügung stehen, wurde sie durch eine andere Gewährsperson ersetzt.

216 Arbeitskarten² sind bereits von Yves Scherrer für seine Dissertation (Scherrer 2012) digitalisiert worden. Davon gehören 36 Karten zur lexikalischen Ebene, 65 zur phonetischen und 115 zur morphologischen, wobei eine davon als Wortstellungskarte auch als syntaktische Karte angesehen werden kann. Diese 216 Karten wurden aufgrund ihrer Relevanz für ein automatisches Übersetzungsprogramm ins Schweizerdeutsche ausgewählt, das von Scherrer (2012) entwickelt wurde.

Für diese Studie sollen nun die bereits digitalisierten Karten noch mit, aus linguistischer Sicht, relevanten Karten ergänzt werden. In Zusammenarbeit mit den Linguistinnen und Linguisten des SADS-Teams am Deutschen Seminar der Universität Zürich sind dafür zusätzlich 19 Karten ausgewählt worden. Zehn davon gehören in die phonetische Ebene, die restlichen neun in die morphologische, wobei drei davon Wortstellungskarten sind, die auch der syntaktischen Ebene zugeordnet werden können. Somit fließen 235 SDS Karten in die Analyse ein: 36 lexikalische, 75 phonetische, 120 morphologische und 4 (morpho-) syntaktische Karten.

Die Daten des zurzeit noch nicht veröffentlichten SADS sind ab dem Jahr 2000 mittels 4 schriftlicher Fragebögen erhoben worden. Insgesamt haben 3185 Gewährspersonen aus 383 Orten (s. Abbildung 2) zu 118 Fragen Auskunft gegeben (Bucheli et al. 2012). Einige Fragen waren Übersetzungs- bzw. Ergänzungsfragen. Bei den meisten gestellten Fragen handelte es sich jedoch um sogenannte Ankreuzfragen, bei denen die Personen zwischen verschiedenen Antwortmöglichkeiten auswählen mussten. Dabei sollten sie einerseits die in ihrem Dialekt akzeptierten Varianten auswählen und andererseits sich für eine natürlichste bzw. präferierte Variante entscheiden. Von den meisten Orten liegen von fünf oder mehr Personen vier vollständig ausgefüllte Fragebogen vor, von vereinzelten Orten sind es weniger, mindestens jedoch zwei vollständig ausgefüllte Fragebogen-Serien. Im Gegensatz zur Erhebung des SDS wurden nicht nur männliche, handwerklich tätige Gewährspersonen befragt, sondern Personen unterschiedlichen Geschlechts und aus verschiedenen Berufs- und Altersgruppen (12-103 Jahre) (Bart und Glaser 2011). Ansonsten wäre es nicht möglich gewesen, genügend ortsfeste Gewährspersonen zu finden (Bucheli 2008).

² Diese Arbeitskarten, welche je eine Variable beschreiben, sind aus 193 originalen SDS-Atlaskarten entstanden.

2 Hintergrund

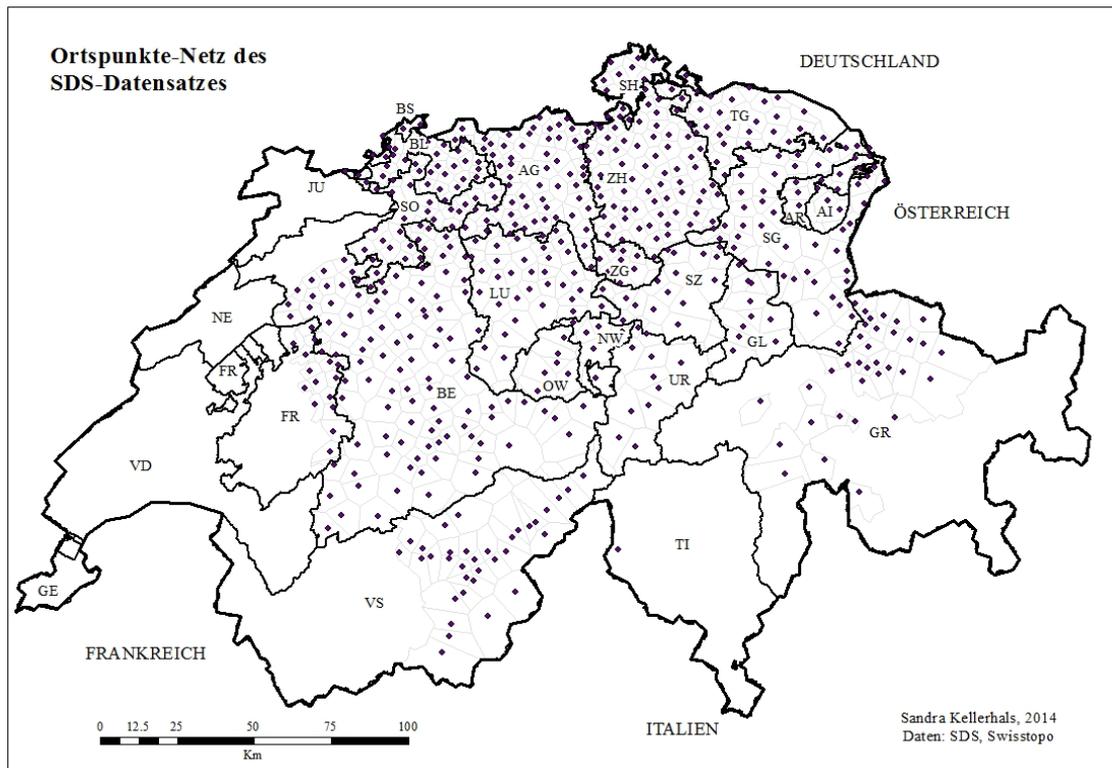


Abbildung 1 Ortspunkte-Netz des SDS-Datensatzes (565 Orte)

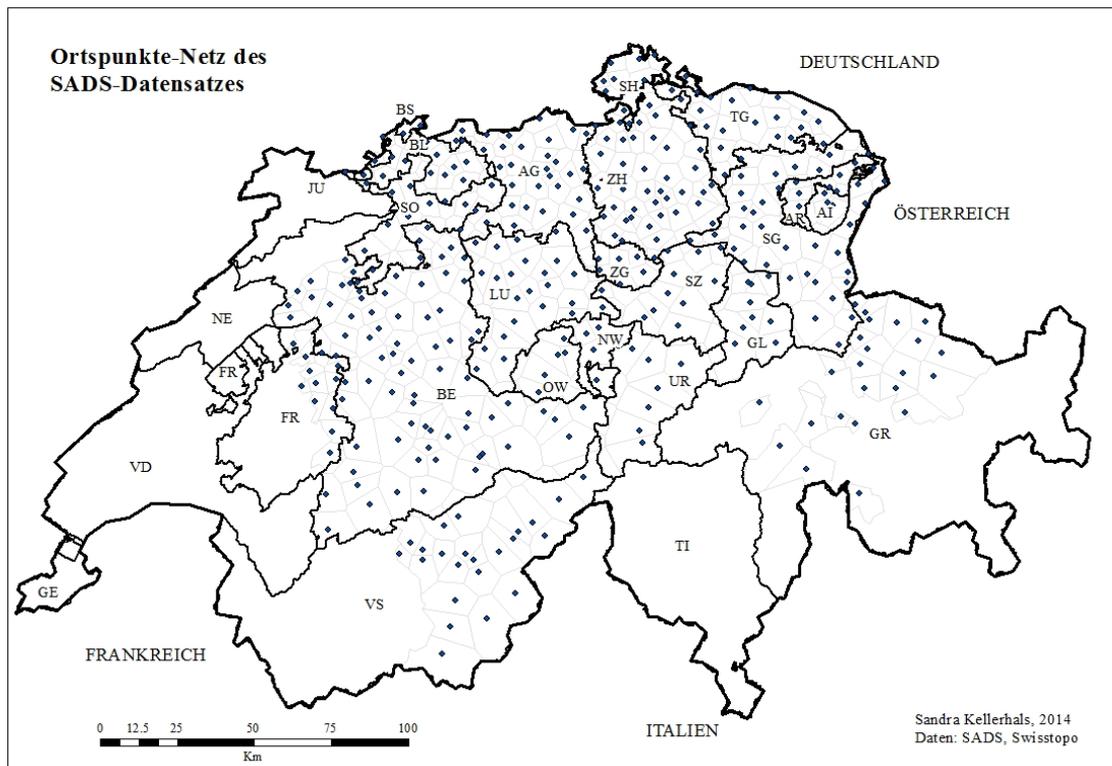


Abbildung 2 Ortspunkte-Netz des SADS-Datensatzes (378 verwendete Orte)

2.3 Forschungsstand

In dieser Arbeit werden dialektometrische Methoden auf schweizerdeutsche Dialekte angewandt, weshalb hier eine kurze Übersicht zum Stand der Forschung auf diesem Gebiet gegeben werden soll. Ausserdem wird zur Hintergrundinformation kurz auf die Entwicklung der Dialektforschung eingegangen.

Bevor Computer zur Verfügung standen, beschränkte sich die Dialektforschung auf einfache dialektologische Methoden, welche die sprachlichen Phänomene und deren geographische Verteilung einzeln untersuchen. Die Idee eine Karte zu zeichnen, welche die Eigenschaften der Sprache innerhalb eines Sprachraumes darstellt, kam im 18. Jahrhundert von Adalbert von Keller (Schrambke 2010). Sein Doktorand Georg Wenker erstellte anschliessend die ersten handgezeichneten Karten, die er im ersten Sprachatlas, dem «Sprach-Atlas der Rheinprovinz nördlich der Mosel sowie des Kreises Siegen» und später im erweiterten «Sprachatlas von Nord- und Mitteldeutschland» (Wenker 1881) veröffentlichte. Als Basis dazu dienten ihm die sogenannten Wenkersätze. Dies sind 40 Sätze, hauptsächlich zu phonologischen und zum Teil morphologischen Phänomenen, welche er an Lehrer in verschiedenen Orten verteilte, um sie von deren Schülern übersetzen zu lassen (Schrambke 2010). Die durch unterschiedliche Merkmale ausfindig gemachten Sprachregionen grenzte Wenker mit handgezeichneten Linien, sogenannten Isoglossen, ab (Lanwermeyer und Womelsdorf 2009). Die Marburger dialektologische Schule, welche auf Wenkers Ansatz zurückging, bildete bis ins 20. Jahrhundert ein wichtiges Forschungsparadigma der deutschen Dialektologie (Lanwermeyer und Womelsdorf 2009).

Karl Haag (1860-1946) und die Württemberger Schule bevorzugte später die direkte, mündliche Befragung zur Untersuchung der sprachlichen Mikrovariation. Ausserdem war er der erste, der die verschiedenen sprachlichen Merkmale, in denen sich die Dialektregionen unterschieden, bündelte beziehungsweise gewichtete und die Stärke der Unterschiede mittels unterschiedlich breiter Isoglossen-Linien darstellte (Schrambke 2010).

Seitdem man mit dem Aufkommen von leistungsfähigen Computern die Möglichkeit hat, relativ einfach grössere Berechnungen durchzuführen, sind neue Analysemethoden für die Untersuchung von Dialektregionen entstanden. Zum einen haben sich zur traditionellen Dialektologie alternative quantitative Ansätze entwickelt. Grieve et al. (2011) stellen beispielsweise einen Zugang vor, welcher die Vorgehensweise der traditionellen Dialektologie als Basis nimmt und jeden Schritt durch einen statistischen Ansatz ersetzt. Dies macht die Methode gegenüber der traditionellen Dialektologie objektiver und replizierbar. Rumpf et al. (2009) präsentieren einen Ansatz, der Auftretensintensitäten von sprachlichen Varianten anhand von Kerndichteschätzungen (KDE) flächenhaft visualisiert. Die Struktur der Karten kann anschliessend in geostatistischen Analysen auf die Homogenität, Komplexität und Kompaktheit der Varianten untersucht werden. Während Rumpf et al. (2009) für ihre Untersuchungen die Dialektdaten des Sprachatlas von Bayerisch-Schwaben (SBS) nutzen, wendet Siblinger (2011) die Methode auf einzelne Phänomene aus dem Syntaktischen Atlas der Deutschen Schweiz (SADS) an. Er erstellt dabei dialekt syntaktische Flächenkarten zu den Phänomenen Finalanschluss,

Komparativ und Artikelverdoppelung und untersucht die räumliche Verteilung der Phänomene anschliessend mit geostatistischen Methoden.

Im Gegensatz zur Untersuchung von einzelnen sprachlichen Phänomenen, seien sie qualitativ oder quantitativ, steht die Methode der Dialektometrie, welche die aggregierten Differenzen oder Ähnlichkeiten zwischen Dialekten, aus einer Vielzahl von Phänomenen berechnet, analysiert und visualisiert. Während in der Dialektologie häufig Sprachregionen auf lokaler bis regionaler Skala untersucht werden, bezieht sich die dialektometrische Forschung in der Regel auf grössere, regionale bis transnationale Skalen. Sie beschäftigt sich mit Fragen zur Verteilung von sprachlichen Variation in einer Sprachregion: Sind die sprachlichen Änderungsraten kontinuierlich oder gibt es scharfe Grenzen? Gibt es in einer Sprachregion relativ homogene Unterregionen, die sich nicht überlappen und durch erkennbare, sprachliche Grenzen abtrennen? Ausserdem kann auch die Frage nach den entscheidenden einzelnen Phänomenen angegangen werden, indem man versucht herauszufinden, welche einzelnen Phänomene am besten mit den Resultaten aus der aggregierten Analyse korrelieren (Nerbonne 2010). Jean Séguy (Séguy 1971), der Begründer der Methode, kam auf die Idee, übereinstimmende Merkmale zwischen zwei oder mehreren Orten zu zählen und so den Dialektunterschied zu quantifizieren. Diese Methode kann auf Phänomene aus unterschiedlichen linguistischen Ebenen angewandt werden (Nerbonne und Kretzschmar 2003). Vor allem in Europa (Grieve et al. 2011) werden dialektometrische Methoden für die Analyse von Dialekten der verschiedensten Sprachen verwendet (z.B. für Italienisch: Goebel 1982a, Goebel 1984, Goebel 2006a, für Rätoromanisch: Goebel 1984, für Französisch: Goebel 2006b, Goebel 2010, für Deutsch: Nerbonne 2009, Nerbonne 2010, für Niederländisch: Heeringa 2004, Spruit 2008). Aber auch zu amerikanischen Dialekten gibt es Studien, z.B. Nerbonne und Kleiweg (2007).

In der Dialektometrie haben sich hauptsächlich zwei Strömungen entwickelt: Die Salzburger und die Groninger Dialektometrie.

Die *Salzburger Schule*, begründet durch Hans Goebel, zeigte das Potential der Dialektometrie systematisch auf. Goebel hat den Gewichteten Identitätswert (GIW) eingeführt (Goebel 1984), der im Gegensatz zum Relativen Identitätswert (RIW) die selten und daher weniger wahrscheinlich vorkommenden Varianten stärker gewichtet als die häufigeren (Nerbonne und Kretzschmar 2003). Ein weiteres Charakteristikum der Salzburger Schule um Goebel ist die Analyse und Visualisierung der Dialektähnlichkeiten mittels Referenzpunkt- (bzw. Ähnlichkeits-) oder Parameterkarten. Ausgehend von bestimmten Orten kann auf die Dialektähnlichkeit zu allen anderen Orten geschlossen werden. Oder es können beispielsweise die Mittelwerte der Distanzen jedes Ortes zu allen anderen Orten dargestellt werden und so auf die sprachliche Isoliertheit der Orte geschlossen werden (Goebel 2010). Ausserdem schlägt Goebel zur Analyse der Dialektdaten die hierarchische Clusteranalyse vor (Goebel 1982a). Auf diese Analysemethoden wird im methodischen Teil dieser Arbeit weiter eingegangen.

Die *Groninger Dialektometrie* um John Nerbonne verwendete erstmals nicht nur die Hamming-Distanz, also das simple Zählen der Anzahl unterschiedlicher Merkmalsausprägungen zwischen den Orten, sondern benutzte für phonetische Merkmale die Levenshtein-Distanz zur Errechnung der Differenz zwischen zwei Varianten. Dieses Distanzmass zählt die Anzahl Änderungen (Einfügen, Löschen, Substituieren), die benötigt werden um ein Wort in ein anderes

überzuführen und vermag daher die Unterschiede differenzierter zu erfassen (Nerbonne et al. 1996). Da sich die Methode jedoch nicht für alle linguistischen Ebenen eignet, wird sie in dieser Arbeit nicht verwendet. Hingegen wird in den Analysen auf die ebenfalls auf die Groninger Dialektometrie zurückgehende Datenreduktionstechnik der Multidimensionalen Skalierung zurückgegriffen, auf die im methodischen Teil noch genauer eingegangen wird.

Während sehr viele dialektometrische Studien sich vorwiegend auf Differenzen auf der phonetischen, lexikalischen oder morphologischen Ebene beziehen (Girnth 2010), gibt es inzwischen auch einige Autoren, welche die syntaktische Ebene ebenfalls untersuchen und Vergleiche mit dieser und den anderen linguistischen Ebenen anstellen. Spruit (2008) tut dies beispielsweise für die niederländischen, Uiboed et al. (2013) für die estnischen Dialekte.

2.4 Forschungslücken

Wie schon im Einführungskapitel erwähnt, wurden dialektometrische Methoden auf schweizerdeutsche Dialekte nur sehr begrenzt angewandt. Kelle (2001) hat auf diesem Gebiet Pionierarbeit geleistet und aus den ersten drei Bänden (Phonetik und Morphologie) des SDS jede zweite Karte für seine Analysen erfasst. Dabei berücksichtigte er aus den 600 Orten ein reduziertes Ortsnetz von 101 Orten. Mittels hierarchischer Clusteranalysen hat er so versucht Dialektgebiete abzugrenzen. Weitere dialektometrische Experimente hat Scherrer (2012) im Rahmen seiner Dissertation durchgeführt. Diese Experimente sollen nun in dieser Arbeit verfeinert werden, so dass Aussagen über Zusammenhänge zwischen den verschiedenen linguistischen Ebenen möglich werden.

Kelle (2001) hat in seine Analysen weder lexikalische noch syntaktische Daten miteinbezogen. Scherrer (2012) verwendete neben den Daten aus dem SDS auch syntaktische Daten aus dem SADS. Jedoch wurden die verschiedenen linguistischen Ebenen separat analysiert. Der Grund dafür liegt unter anderem darin, dass die unterschiedliche Anzahl Ortspunkte sowie die unterschiedliche Informantenauswahl einen Vergleich zwischen den beiden Datensätzen schwierig macht. Durch einen neuen Ansatz in dieser Arbeit sollen diese beiden Probleme eliminiert und ein Vergleich der sprachlichen Unterschiede zwischen allen vier Ebenen ermöglicht werden.

3 Methodik: Die Dialektometrische Verfahrenskette

Dieses Kapitel soll einen allgemeinen Überblick über die verwendeten Methoden geben, während das Kapitel 4 detaillierter auf die einzelnen Schritte eingeht und die konkrete Umsetzung zeigt.

Die dialektometrische Verfahrenskette beschreibt die Überführung der linguistischen Daten von ihrem Ausgangszustand in eine Form, die dialektometrische Analysen zulässt (Goebel 1995). In einem ersten Schritt muss für jede linguistische Variable festgelegt werden, welche Antwortmöglichkeiten bzw. Varianten berücksichtigt werden sollen – und ob sie als einzelne Antwort oder zusammengefasst in eine Klasse bzw. als ein Typ in die Analyse einfließen sollen. Für die Ergänzungsfrage I.4 aus dem SADS «Die Nachbarin ist...(einkaufen)» wurden beispielsweise die Antworten «go poschte gange» und «gange go poschte» zu einem Typen zusammengefasst, während «go poschte» einen weiteren Typ bildet. Antworten wie «goge poschte gange» und «gange goge poschte» wurden gar nicht berücksichtigt. Sind diese Entscheidungen gefällt, kann aus den Daten eine Datenmatrix gebildet werden (s. Tabelle 2). Die Zeilen der Matrix enthalten die Orte, während die Spalten die Variablen enthalten.

Datenmatrix	3259 du es / es du: Wortstellung	3260 ihr es / es ihr: Wortstellung	3261 Syntax: [gewesen bin] f.	3262 gehen lassen: Wortstellung
Aarau (AG)	du_es	ihr_es	pp_aux	la_ge
Aarberg (BE)	du_es	ihr_es	pp_aux	la_ge
Aarburg (AG)	du_es	ihr_es	pp_aux	la_ge
Abländschen (BE)	du_es	ihr_es	aux_pp	la_ge
Adelboden (BE)	du_es	ihr_es	aux_pp	la_ge
Aesch (BA)	NA	ihr_es / es_ihr	pp_aux	la_ge
Aesch (LU)	es_du	ihr_es / es_ihr	pp_aux	la_ge
Aeschi (SO)	du_es	ihr_es / es_ihr	pp_aux	la_ge
Agarn (WS)	es_du	es_ihr	pp_aux	la_ge
Alpnach (UW)	es_du	ihr_es / es_ihr	NA	la_ge

Tabelle 2 Beispiel für eine Datenmatrix (Wortstellungskarten aus dem SDS)

Aus einer Datenmatrix kann anschliessend über die Wahl eines geeigneten linguistischen Distanzmasses eine Distanzmatrix generiert werden. Anstelle von geographischen Distanzen zwischen Orten enthält eine solche Matrix die linguistischen Distanzen zwischen den Orten, die auf verschiedene Arten berechnet werden können. Die in Tabelle 3 aufgeführten Distanzen sind mit dem Distanzmass der Hamming-Distanz bzw. des Relativen Identitätswerts berechnet (RIW) (Goebel 1982a). Dieser Wert stellt den prozentualen Anteil der identischen Merkmale im Verhältnis zur gesamten Anzahl von Merkmalen zwischen zwei Orten dar. Die Werte bewegen

sich zwischen 0 und 1, wobei 1 die maximale Übereinstimmung zwischen zwei Orten bedeutet. Um nun eine Distanzmatrix zu erhalten, müssen diese Ähnlichkeitswerte nur noch von eins subtrahiert werden. Im vorliegenden Beispiel stimmen beim Orts-Paar Aarburg (AG) und Abländschen (BE) von vier Variablen drei überein und eine nicht, was zu einer linguistischen Distanz von 0.25 zwischen diesen beiden Orten führt. Beim Orts-Paar Aarau (AG) und Aeschi (SO) stimmen ebenfalls drei von vier Variablen überein. Zusätzlich stimmt die Variable 3260 zur Hälfte überein, da in Aeschi beide Variablen vorkommen und in Aarau nur eine davon. Deswegen beträgt die Distanz zwischen den Orten Aarau (AG) und Aeschi (SO) 0.125.

Distanzmatrix	Aarau_(AG)	Aarberg_(BE)	Aarburg_(AG)	Abländschen_(BE)	Adelboden_(BE)	Aesch_(BA)	Aesch_(LU)	Aeschi_(SO)	Agarn_(WS)	Alpnach_(UW)
Aarau_(AG)	0	0	0	0.25	0.25	0.375	0.375	0.125	0.5	0.625
Aarberg_(BE)	0	0	0	0.25	0.25	0.375	0.375	0.125	0.5	0.625
Aarburg_(AG)	0	0	0	0.25	0.25	0.375	0.375	0.125	0.5	0.625
Abländschen_(BE)	0.25	0.25	0.25	0	0	0.625	0.625	0.375	0.75	0.625
Adelboden_(BE)	0.25	0.25	0.25	0	0	0.625	0.625	0.375	0.75	0.625
Aesch_(BA)	0.375	0.375	0.375	0.625	0.625	0	0.25	0.25	0.375	0.5
Aesch_(LU)	0.375	0.375	0.375	0.625	0.625	0.25	0	0.25	0.125	0.25
Aeschi_(SO)	0.125	0.125	0.125	0.375	0.375	0.25	0.25	0	0.375	0.5
Agarn_(WS)	0.5	0.5	0.5	0.75	0.75	0.375	0.125	0.375	0	0.375
Alpnach_(UW)	0.625	0.625	0.625	0.625	0.625	0.5	0.25	0.5	0.375	0

Tabelle 3 Linguistische Distanzmatrix

Der RIW eignet sich als Distanzmass für Variablen aus allen linguistischen Ebenen (Zastrow 2011). Die Distanztabelle dient nun als Grundlage für die dialektometrischen Analysen. Es wurden in der Literatur verschiedene Analysemöglichkeiten und –werkzeuge entwickelt. Die in dieser Arbeit verwendeten sollen im Folgenden kurz erläutert werden.

Parameterkarten

Aus den Distanzen eines Ortes zu allen anderen Orten lassen sich verschiedene statistische Kennzahlen, wie zum Beispiel der Mittelwert oder die Schiefe berechnen. Diese Werte können sodann auf einer sogenannten Parameterkarte farblich differenziert dargestellt werden. Beispielsweise können Orte, für welche positive Werte für die berechneten Schiefen resultieren, warme Farben verwendet werden. Dies deutet auf Regionen mit grossem Sprachausgleich hin, während kältere Farben für negative Schiefe-Werte stehen und für eine Residuallage der Gegend mit geringem Sprachausgleich stehen (Goebl 2004).

Hierarchische Clusteranalysen

Bei der hierarchischen Clusteranalyse wird das Dialektgebiet in eine bestimmte Anzahl Untergebiete eingeteilt, in dem schrittweise die zwei ähnlichsten Gruppen zusammengefasst werden. Für die Clusterbildung existieren verschiedene Algorithmen (Goebl 1982a).

Multidimensionale Skalierung

Multidimensionale Skalierung reduziert die vieldimensionalen Distanzmatrizen in eine kleine Zahl von n Dimensionen. In der Dialektometrie wird üblicherweise $n = 2$ oder $n = 3$ gewählt und zur Visualisierung verwendet. Dies ermöglicht eine angemessene Visualisierung der Distanzen, ohne dass diese sich zu stark von den ursprünglichen unterscheiden (Heeringa 2004).

Um die dialektometrischen Analysen durchzuführen, bieten sich zwei verschiedene Software-Pakete an: Einerseits das von den Salzburger Dialektologen entwickelte Visual DialectoMetry (VDM) (Goebl 2004) und andererseits das online verfügbare Analysewerkzeug Gabmap (Nerbonne et al. 2011) der Dialektologen der Universität Groningen.

Während nur VDM die Möglichkeit bietet Parameterkarten zu generieren, hat Gabmap den Vorteil, dass Multidimensionale Skalierungen durchgeführt werden können. Ausserdem lässt sich in VDM nur eine Variante pro Variable und Ort einbeziehen, während Gabmap auch Mehrfachantworten zulässt. Für diese Arbeit sind Analysen mit beiden Software-Paketen durchgeführt worden.

Das nächste Kapitel erklärt das Vorgehen entlang der dialektometrischen Verfahrenskette schrittweise, stellt die verschiedenen Analysemöglichkeiten vor und geht detaillierter auf die verwendeten Verfahren ein.

4 Vorgehen

4.1 Datenauswahl und Datenintegration

Um das Ziel eines möglichst sinnvollen Vergleichs zwischen den beiden Datensätzen SDS (lexikalische, morphologische und phonetische Phänomene) und SADS (syntaktische Phänomene) zu erreichen, soll aus dem SADS ein Datensatz gebildet werden, dessen Eigenschaften möglichst denjenigen des SDS entsprechen. Wie die dafür verwendeten Daten ausgewählt werden, wie sie vergleichbar sind und welche Schwierigkeiten dabei zu meistern sind, wird in den nächsten Abschnitten besprochen.

4.1.1 SDS

Im Anhang A befindet sich eine Liste mit allen für diese Studie verwendeten Karten des SDS. Es sollten möglichst viele Karten in die Analyse mit einfließen und die aus linguistischer Sicht relevanten Phänomene nicht fehlen. Durch die Arbeit von Scherrer (2012) war eine, für die vorliegende Arbeit nicht repräsentative, Anzahl an SDS-Karten bereits digitalisiert worden. Diese sollten nun mit den aus linguistischer Sicht ebenfalls interessanten Phänomenen ergänzt werden. So wurden zusätzlich die in Tabelle 4 aufgeführten phonetischen, morphologischen und morpho-syntaktischen Karten digitalisiert und mitberücksichtigt.

SDS-Band	Ling. Ebene	Karten-nr.	Phänomen	Beispiel
1	Phonetik	1.015	mhd. e (Primärumlaut)	Bett
1	Phonetik	1.021	mhd. ë (germ. e)	Speck
1	Phonetik	1.041	mhd. o	Gotte
1	Phonetik	1.054 (55)	Umlaut/ Nichtumlaut von mhd. u vor ck/ gg	Rücken, drücken, Brücke
1	Phonetik	1.067	mhd. â	Samen
1	Phonetik	1.095	mhd. ê	Schnee (Brechung)
2	Phonetik	2.015	Dehnung von alter Kürze in offener Silbe – mhd. a	Magen
2	Phonetik	2.027	Dehnung von alter Kürze in offener Silbe – mhd. ë	Besen
3	Morphologie	3.028	Endungen des Normalverbs – 3. Pers. Sg. Ind. Präs.	(es) kostet
3	Morphologie	3.050	1. 2. 3. Pers. Pl. Ind. Präs.	sein
3	Morphologie	3.051	1. 3. Pers. Sg. Konj. Präs.	sein
3	Morphologie	3.055	Gerundium	tun
3	Morphologie	3.205	Personalpronomen	ihm
3	Syntax	3.259	Wortstellung -du es/ -es du	Hast du's gern?

4 Vorgehen

SDS-Band	Ling. Ebene	Karten-nr.	Phänomen	Beispiel
3	Syntax	3.260	Wortstellung -ihr es/ -es ihr	Könnt ihr's auch?/ Sagt ihr's?
3	Syntax	3.262	Wortstellung	(er hat ihn) gehen lassen
3	Morphologie	3.263	Erstartete Infinitiv-Partikeln	Lass ihn gehen
3	Morphologie	3.265	Erstartete Infinitiv-Partikeln	Es kommt regnen
3	Morphologie	3.266	Umschreibung der inchoativen Funktion	(sonst) wirst / kommst (du krank)

Tabelle 4 Im Vergleich zur Arbeit von Scherrer (2012) zusätzlich berücksichtigte Karten

Zudem wurden bei einigen von Scherrer (2012) bereits digitalisierten Karten zusammengefasste Varianten wieder getrennt oder zusätzliche Varianten ergänzt (Tabelle 5). Dies geschah ausschliesslich bei Karten aus dem dritten Band des SDS (Morphologie).

Karten-nr.	Phänomen	Bereits vorhandene Varianten (Anzahl)	Ergänzte bzw. getrennte Variante(n) (Anzahl)	Beispiel
3.050	Infinitiv/ Part. Perfekt (Plural)	si, se, sō (3)	sei (+1)	sein
3.057	2./3. Pers. Sg. Ind. Präs.	gaa/ goo, gei, gee/ gi, gāi (4)	goo, gou (+2)	gehen
3.064	1. Pers. Sg. Ind. Präs.	laa(n)/ loo(n)/ luu(n), loss (2)	loo(n), luu(n), lou(n), lää(n), laasse (+5)	lassen
3.070	1. Pers. Sg. Ind. Präs.	schlaa(n)/ schloo(n)/ schluu(n), schlag, schlee/ schlää/ schlei (3)	schloo(n), schluu(n), schlou(n), schlää(n) (+4)	schlagen
3.071	2./ 3.Pers. Sg. Ind. Präs.	schlaa(n)/ schloo(n)/ schlou(n), schlee(n)/ schlää(n)/ schlei(n) (2)	schloo(n), schlou(n), schlei(n) (+3)	schlagen
3.074	1. Pers. Sg. Ind. Präs.	faa(n)/ foo(n), fang(e), fee(n)/ fää(n) (3)	foo(n), fää(n), fou(n), fuu(n) (+4)	(an)fangen
3.075	2./ 3. Pers. Sg. Ind. Präs.	faa(n)/ foo(n), fang(e), fee(n)/ fää(h), foch/ foh (4)	foo(n), fou(n), föö(n), fie(n) (+4)	(an)fangen
3.087	Indikativ Plural	müe, müesse, mue (3)	muesse, müi (+2)	müssen
3.091	1.-3.Person Plural	näme/ neme/ nime, nä/ ne/ ni, nāi	nämd (+1)	nehmen
3.092	2./ 3.Person Singular	nim, nin, nüm, nen, nii, nee, nün (7)	nem (+1)	nehmen
3.094	Vokalqualität	-ä, -e, -i (3)	-ea (+1)	geben/ nehmen
3.095	Partizip Perfekt	gnoo, gnuu, gnaa, gnoe (4)	gnou (+1)	nehmen
3.096	1./ 2./ 3. Person Singular	(g)see, (g)sie (2)	(g)sech, (g)siech (+2)	sehen
3.108	3. Person Singular	töörfe, tööre (2)	terf, töör, täär (+3)	dürfen

Tabelle 5 Im Vergleich zur Arbeit von Scherrer (2012) ergänzte und getrennte Varianten³

³ Durch Schrägstriche getrennte Wörter bilden eine Variante. Verschiedene Varianten sind durch Kommas getrennt.

4.1.2 SADS

Von den zur Verfügung stehenden SADS-Daten, welche in einer Filemaker-Datenbank vorliegen, sind von 118 Fragen 108 in die Analysen miteinbezogen worden. Gründe für das Weglassen einiger Fragen sind unter anderem keine Variation (in allen Orten ist dieselbe Variante gewählt worden) oder zu viele Nullwerte. Dabei ist keine der zur Verfügung stehenden Varianten von der Gewährsperson genannt worden oder die Antwort ist ungültig. Die Liste im Anhang B gibt Auskunft über die in die Analyse mit einbezogenen Fragen.

Im SADS und in dieser Arbeit werden bewusst nicht alle linguistischen Phänomene durch dieselbe Anzahl von Fragen bzw. Variablen beschrieben. Nach den Linguistinnen und Linguisten des SADS-Teams ist für die Behandlung von komplexeren Phänomenen eine höhere Anzahl von Fragen erforderlich, was auch ein höheres Gewicht dieser Phänomene im Vergleich zu den einfacheren rechtfertigt. Um die Anzahl der Varianten pro Variable etwas zu begrenzen sind linguistisch ähnliche Antworten zu einem Antworttyp bzw. einer Klasse zusammengefasst worden. Aus der erwähnten Liste im Anhang B sind auch diese Typenbildungen ersichtlich. Die Auswahl der miteinbezogenen Fragen sowie die Typenbildung der Antworten sind durch die Linguistinnen und Linguisten des SADS-Teams erarbeitet worden.

4.1.3 Abgleich der Gewährspersonen

Da beim SADS-Datensatz die Antworten mehrerer Personen pro Ort zur Verfügung stehen, beim SDS jedoch im Normalfall nur eine Person pro Ort Auskunft gegeben hat, musste eine Methode gefunden werden, welche die zum Teil verschiedenen Antworten des SADS pro Ort auf eine Antwort pro Ort reduziert. In Annäherung an die Methode des SDS ist deshalb pro Ort eine Gewährsperson ausgewählt worden, deren Charakteristika möglichst denjenigen der SDS-Gewährspersonen entsprechen. So sollte die Person bei der Befragung möglichst 60 Jahre oder älter gewesen sein und insgesamt möglichst viele brauchbare Antworten gegeben haben. Hatten mehrere ältere Personen bei allen Fragen brauchbare Antworten gegeben, sind diejenigen bevorzugt worden, welche nicht länger als ein Jahr oder höchstens vor langer Zeit, für eine begrenzte Periode, in einem anderen deutschsprachigen Kanton gelebt haben, bei denen höchstens ein Elternteil nicht aus dem untersuchten Ort stammt und die eher einen landwirtschaftlich-handwerklichen Beruf ausüben. Da alle der für den SADS befragten Personen so oder so als «ortsfest» gelten, das heisst, dass sie als Repräsentanten des vorherrschenden Dialektes geeignet sind, ist diese sekundäre Auswahl vor allem getroffen worden, um eine Vorgehensweise nach einheitlichen Kriterien zu gewährleisten und die Auswahl so nachvollziehbarer zu machen. Insgesamt konnte in 16 Orten (4,2 %) keine Person mit Jahrgang 1940 oder älter gefunden werden (das heisst älter als 60 Jahre zum Erhebungszeitpunkt des SADS). So sind in 14 Orten Personen bis Jahrgang 1944 miteinbezogen worden. Zwei Orte, Solothurn (SO) und Niederweningen (ZH), sind weggelassen worden, da die ältesten Personen Jahrgang 1949 bis 1952 haben und somit für diese Auswahl zu jung sind. Ebenfalls in Anlehnung an die Methode des SDS werden weiter nur die Varianten berücksichtigt, die von der jeweils ausgewählten Gewährsperson bei den Ankreuzfragen als natürlichste genannt worden sind. Es werden also nicht alle «akzeptierten» Varianten

berücksichtigt, sondern es fließt in den meisten Fällen, wie beim SDS, nur eine Variante pro Ort in die Analyse mit ein. Mehrfachnennungen sind allerdings auch hier möglich.

4.1.4 Abgleich des Ortsnetzes

Der SDS-Datensatz verfügt über 565 Ortspunkte, während beim SADS Daten aus 383 Orten zur Verfügung stehen. Um die Dialekt-Unterschiede in den beiden Datensätzen SDS und SADS auch mittels statistischer Methoden vergleichen zu können, muss jedoch ein für beide Datensätze identisches Ortsnetz gebildet werden. Für die Kartierung der Dialektunterschiede werden um die einzelnen Orte sogenannte Thiessen-Polygone (oder Voronoi-Polygone) gebildet, deren Fläche dann jeweils mit der entsprechenden Farbe eingefärbt wird. Die Orte sind die jeweiligen Zentren dieser Polygone. Da die Auswahl der SADS-Orte in Anlehnung an die SDS-Orte geschah, stimmen 342 Ortspunkte in beiden Datensätzen überein (Abbildung 3).

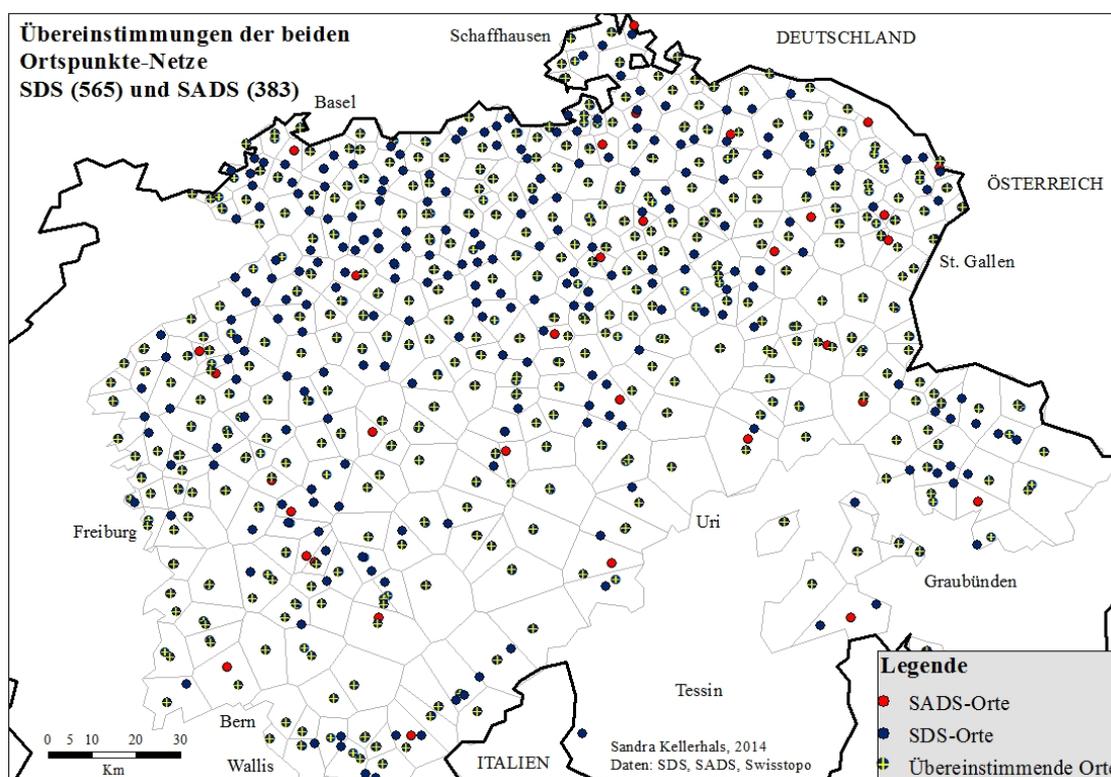


Abbildung 3 Übereinstimmungen der beiden Ortspunktenetze (mit Thiessen-Polygonen um die SADS Orte)

Die Übereinstimmung lässt sich mittels Abgleich der BFS-Nummern⁴ überprüfen. Damit jedoch möglichst viele SADS-Ortspunkte berücksichtigt werden können, ist jedem der restlichen 41 Ortspunkte aus dem SADS (rot in Abbildung 3), jeweils der geographisch nächstliegende «freie» SDS-Ort (blau in Abbildung 3) zugeordnet worden. Häufig handelt es sich bei den jeweils zugewiesenen Orten um Nachbargemeinden oder um Gemeinden, die sich im selben, aus dem SADS-Ortsnetz gebildeten Thiessen-Polygon befinden. Die Zuweisung geschah in Absprache mit den Linguistinnen und Linguisten des SADS-Teams. Im SADS-Ortsnetz sind

⁴ Die BFS-Nummer ist eine durch das Bundesamt für Statistik (BFS) festgelegte eindeutige Nummer zur Identifizierung der politischen Gemeinden in der Schweiz.

einige Gemeinden, die über mehrere Dorfkerne verfügen, mehrmals vorhanden. Die BFS-Nummern unterscheiden sich dann nur in der letzten Ziffer (z.B. Davos: 3851_1 und 3851_2). Bei drei dieser Gemeinden (Davos GR, Pfäfers SG und Ermatingen TG) konnte nicht für alle Ortspunkte eine sinnvolle Entsprechung im SDS-Datensatz gefunden werden, weshalb jeweils ein SADS-Ortspunkt weggelassen werden musste. Alle diese drei Gemeinden sind jedoch im Datensatz mindestens ein weiteres Mal vorhanden. Zusammen mit den beiden eliminierten Ortspunkten aufgrund des zu jungen Alters der Gewährspersonen blieben somit 378 Orte für die Analyse beider Datensätze übrig. Im Anhang C befindet sich eine Liste mit den 378 verwendeten Orten, wo auch die Zuordnung der SDS- zu den entsprechenden SADS-Ortspunkten ersichtlich ist. Auch die fünf weggelassenen SADS-Ortspunkte sind in einer Tabelle im Anhang C aufgeführt.

4.1.5 Vergleichbarkeit der beiden Datensätze

Für den Vergleich der SADS mit den SDS-Daten musste bei beiden Datensätzen eine Reduktion vorgenommen werden. Beim SDS können, wie im vorangehenden Kapitel erklärt, nicht alle 565 Ortspunkte berücksichtigt werden, wodurch Information verloren geht. Quantitative Vergleiche wären aber mit zwei unterschiedlichen Ortsnetzen nicht möglich. Um die Dimensionen des Informationsverlustes abschätzen zu können, werden die Analysen der aggregierten Dialektunterschiede wo möglich auch mit dem gesamten Ortsnetz des SDS durchgeführt und Karten daraus generiert. Diese können dann mit den aus 378 Orts-Punkten generierten Karten verglichen werden.

Beim SADS stünden eigentlich mehrere Datensätze pro Ort zur Verfügung, und mit Berücksichtigung aller akzeptierten Varianten, jeweils auch mehrere Varianten pro Gewährsperson. In der vorliegenden Arbeit wird allerdings jeweils nur eine Gewährsperson und deren natürlichste Antwort berücksichtigt. Dies hat den Vorteil, dass die Struktur der Antworten eher derjenigen des SDS entspricht. Andererseits kann die syntaktische Variation eines Ortes möglicherweise nicht in ihrer Gesamtheit erfasst werden. Es ist zudem möglich, dass die ausgewählte Gewährsperson nicht die in dem Ort dominante, also meistgenannte Variante als natürlichste gewählt hat, sondern zur Minderheit gehört. Da das Ziel dieser Arbeit jedoch nicht ist, die Dialektsyntax in all ihren Variationen zu erfassen, sondern eher einen möglichst sinnvollen Vergleich mit dem SDS-Datensatz zu erreichen, wird dieser allfällige Verlust bewusst in Kauf genommen.

Trotz aller Bemühungen die beiden Datensätze möglichst vergleichbar zu machen, gibt es weiterhin Unterschiede, die nicht ausgeglichen werden können. Dazu gehört die unterschiedliche Anzahl der verwendeten Variablen pro linguistische Ebene. Diese im Rahmen dieser Arbeit auszugleichen wäre jedoch wenig sinnvoll, da mit der vorliegenden Anzahl gemäss den Linguistinnen und Linguisten des SADS-Teams die wichtigsten Phänomene des jeweiligen Bereichs abgedeckt sind und nicht jeder linguistische Bereich die selbe Anzahl Variablen benötigt, um bezüglich Sprachvariation angemessen repräsentiert zu sein. Vor allem bei den relativ spärlich vorhandenen lexikalischen Karten (36) wäre der Aufwand, weitere solche Phänomene zu digitalisieren, sehr gross, da in vielen Fällen eine grosse Anzahl Varianten vorhanden ist.

4 Vorgehen

Auch variiert die durchschnittliche Anzahl Varianten pro Variable in den unterschiedlichen linguistischen Bereichen (von 6.9 in der Lexik bis 2.9 in der Phonetik; Syntax: 3.9, Morphologie: 3.6). Dies könnte im lexikalischen Bereich zu eher grösseren linguistischen Distanzen zwischen den Orten führen, als beispielsweise im phonetischen Bereich, da häufig nicht nur zwei, sondern mehr Ausprägungen pro Variable in Frage kommen. Da jedoch nicht jede linguistische Variable mit derselben Anzahl an Varianten erfasst werden kann, ist es nicht sinnvoll, diese Unterschiede ausgleichen zu wollen.

Weiter könnten die unterschiedliche Befragungsmethode und der unterschiedliche Erhebungszeitraum zwischen den Datensätzen ein Problem sein. Die SDS-Daten wurden 40-60 Jahre vor den SADS-Daten mündlich erhoben und die Erhebung erstreckte sich über einen viel grösseren Zeitraum als bei den SADS-Daten, welche schriftlich erhoben wurden. Somit ist nicht klar, welche Unterschiede durch die zeitliche Komponente oder durch die unterschiedliche Befragungsmethode bedingt sind und welche rein durch die linguistische Ebene. Zur Überprüfung dieser Frage ist das Phänomen der Position der Verbalteile im Nebensatz (Perfekt) aus dem SADS-Datensatz (Karte 119 «bezahlt hat»), mit der entsprechenden Karte aus dem SDS (3261 «gewesen bin») verglichen worden. Die Übereinstimmung der beiden Karten liegt bei hohen 79 %. Abbildung 4 zeigt den Vergleich grafisch. Die Punktsymbole entsprechen den Varianten der SADS-Karte 119, die Flächensymbole denjenigen der SDS-Karte 3261. Vor allem in den Kantonen Wallis und Bern, wo im SDS die Variante «bin gewesen» noch fast 80 Mal vorkommt, wird «hat bezahlt» in der SADS-Befragung nur noch neun Mal genannt (zwei Mal zusammen mit «bezahlt hat») und acht Mal davon in den selben Orten wie im SDS. In beiden Befragungen dominierte aber die Variante mit dem nachgestellten Hilfsverb («bezahlt hat» bzw. «gewesen bin») klar in Freiburg, der Nord-, Ost- und Zentralschweiz. Das Resultat zeigt, dass es zwar Unterschiede gibt, welche auf die unterschiedliche Befragungsmethode oder auf die zeitliche Veränderung zurückgeführt werden könnten, dass die Übereinstimmung jedoch immer noch relativ hoch ist und Vergleiche zwischen den beiden Datensätzen trotzdem Sinn machen. Leider sind dies die einzigen Karten aus den beiden Datensätzen, welche direkt verglichen werden können. Der Grund dafür ist die genannte geringe Zahl von Karten zu syntaktischen Phänomenen im SDS.

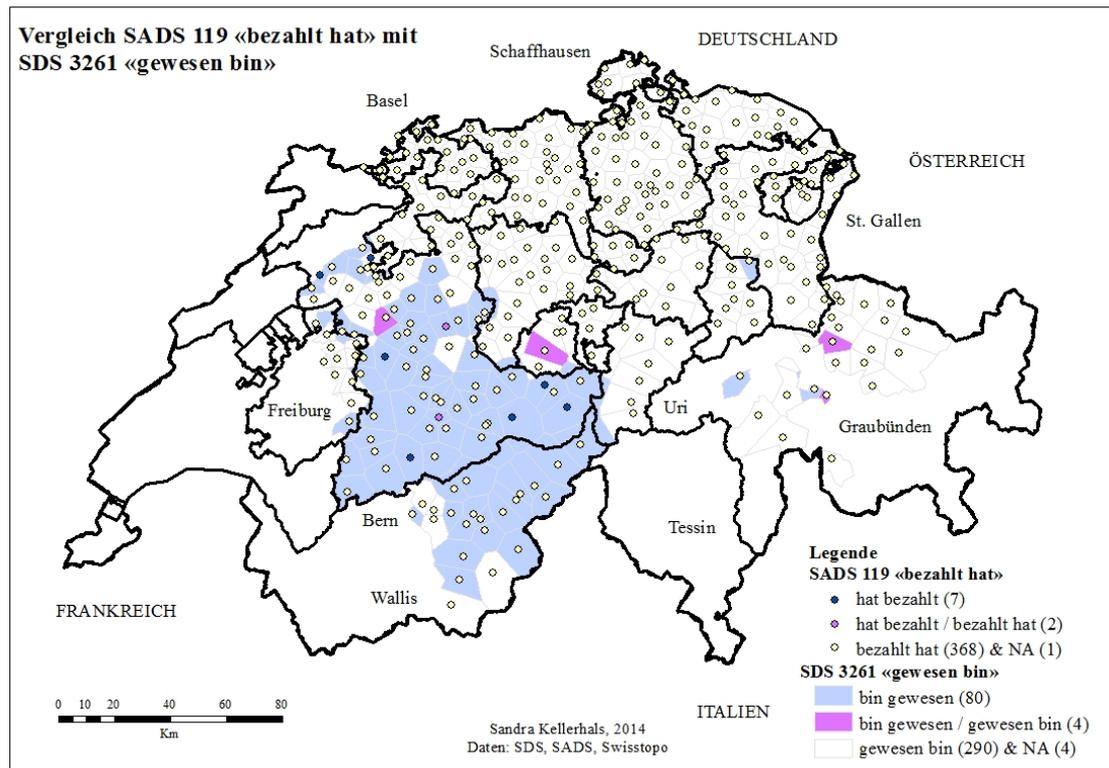


Abbildung 4 Vergleich SADS 119 «bezahlt hat» mit SDS 3261 «gewesen bin»

4.2 Datenaufbereitung

4.2.1 Digitalisierung der SDS-Karten

Wie in Abschnitt 4.1.1. beschrieben sollten zusätzlich zur Arbeit von Scherrer (2012) weitere sprachlich relevante Karten aus dem SDS digitalisiert und in der Analyse mitberücksichtigt werden. Im Folgenden wird der Ablauf der Digitalisierung an einem Beispiel kurz beschrieben. Abbildung 5 zeigt die phonetische Karte «Bett» aus dem ersten Band des SDS. Hier wird unterschieden zwischen einem «geschlossenen e» und einem «offenen e». Die dritte Variante «überoffenes e» kommt nur in zwei Orten zusammen mit dem «offenen e» vor und wird deshalb weggelassen. Es sind nur Varianten berücksichtigt worden, die mindestens einmal alleine an einem Ort auftreten oder mindestens dreimal zusammen mit einer anderen Variante an einem Ort⁵.

⁵ Diese Kriterien sind strenger, als diejenigen aus Scherrer (2012).

4 Vorgehen

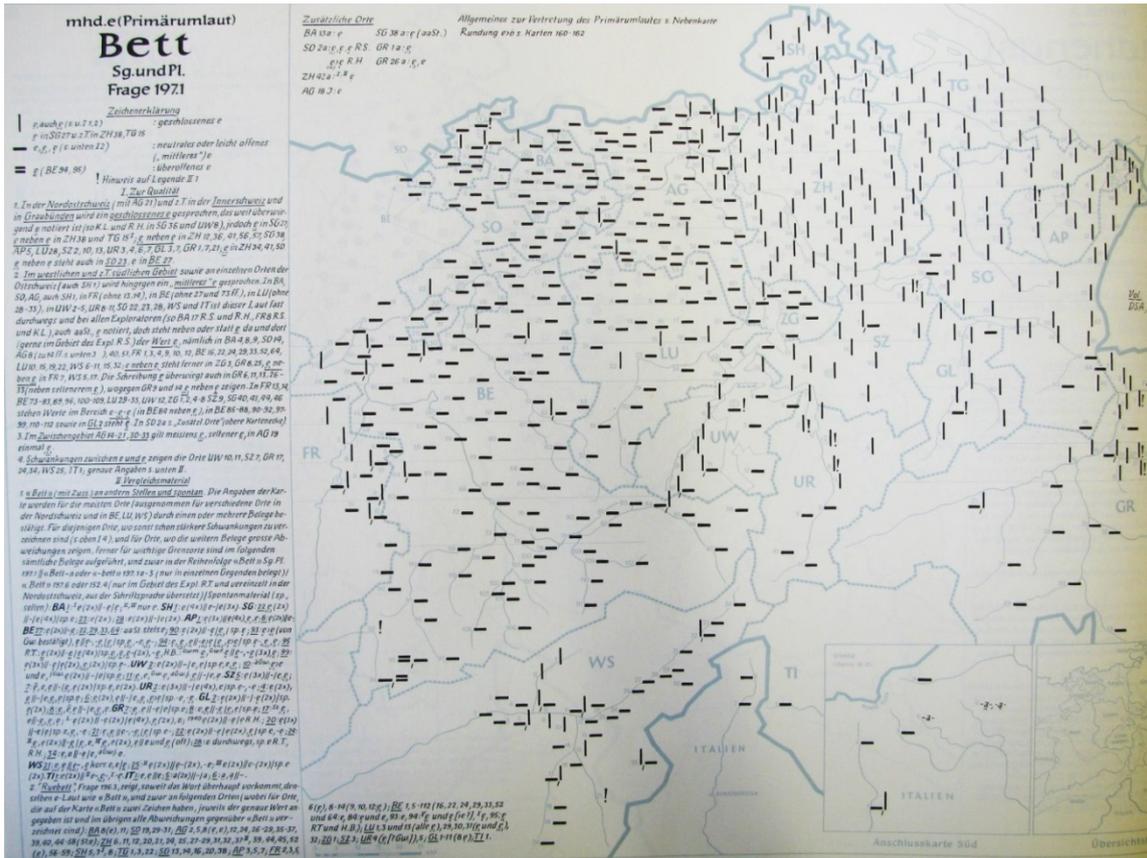


Abbildung 5 Phonetische Karte «Bett» aus dem ersten Band des SDS

Zur Digitalisierung wird das Bild der SDS-Karte in das geographische Informationssystem ArcGIS geladen. Zudem wird eine bereits digitalisierte Karte aus der Arbeit von Scherrer, die als Shapefile vorliegt kopiert und unter neuem Namen ebenfalls in das ArcGIS (bzw. ArcMap) geladen. Das Shapefile enthält alle SDS-Ortspunkte mit den dazugehörigen Koordinaten und Attributen, so dass in ArcMap daraus eine Karte erstellt werden kann.

Als erstes muss das Bild der SDS-Karte georeferenziert werden, das heisst den Ortspunkten auf dem Bild müssen dieselben Koordinaten zugewiesen werden wie denjenigen im Shapefile. Das Bild wird also so in die Karte eingepasst, dass die entsprechenden Ortspunkte auf dem Bild und im Shapefile jeweils untereinander liegen. Für diese Aufgabe gibt es in ArcMap verschiedene Georeferenzierungs-Funktionen, mit Hilfe derer markante Punkte auf dem Bild als Passpunkte mit den entsprechenden Punkten im Shapefile verbunden werden können. Für die besten Resultate wählt man vorzugsweise mindestens vier möglichst weit auseinanderliegende Punkte, im vorliegenden Fall z.B. Basel, Bern, St. Gallen und Chur, und verbindet diese mit den entsprechenden Passpunkten im Shapefile.

Nun muss in der Attributtabelle des Shapefiles für jede Variante, hier also für «e geschlossen» und für «e offen» eine Spalte gebildet werden. Da nun jeder Punkt auf dem entsprechenden Bildpunkt liegt, ist für jeden Punkt ersichtlich, welche Variante diesem zugeordnet werden muss. So können alle Punkte einer bestimmten Variante ausgewählt und in der entsprechenden Attributspalte eine 1 eingefügt werden. Dieser Vorgang wird nun für jede Spalte wiederholt und am Ende kann man mittels einer «Total»-Spalte kontrollieren, ob jedem Punkt eine Variante zugeordnet wurde und ob an den Orten, wo beide Varianten vorkommen, auch beide erfasst

4 Vorgehen

worden sind. Die Abbildung 6 zeigt die fertig digitalisierte Karte inklusive zugehöriger Attributtabelle.

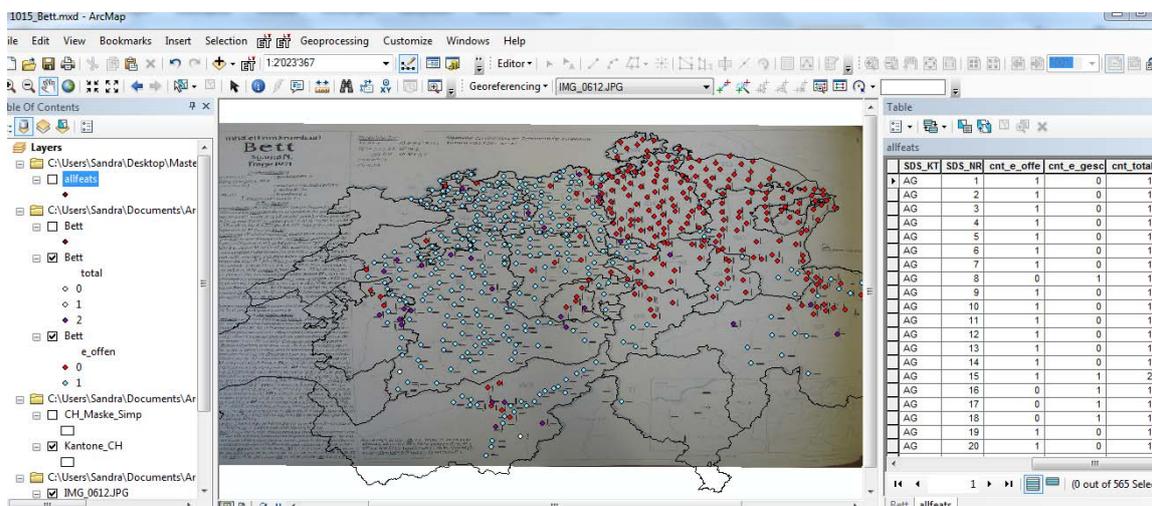


Abbildung 6 Digitalisierte Karte «Bett» in ArcMap

4.2.2 Datenexport der SADS-Daten

Die SADS-Daten liegen in einer Filemaker-Datenbank vor. Jede Frage ist, sortiert nach den vier Fragebogen, in einem sogenannten «Layout» abgespeichert. Ein solches enthält eine Tabelle, in der jede Zeile (Tupel) einer Gewährsperson entspricht. Die Spalten enthalten neben den Antworten auch weitere Attribute wie Alter, Wohnort und pro Gewährsperson eine Nummer (Primärschlüssel). Für jede Antwortmöglichkeit existiert eine Spalte mit der entsprechenden Antwortnummer, die eine Eins enthält, wenn die Antwort von der Gewährsperson so übersetzt oder ergänzt wurde (Übersetzungs- bzw. Ergänzungsfrage) oder als akzeptiert genannt wurde (Ankreuzfrage) und eine Null, wenn dies nicht der Fall war. Bei den Ankreuzfragen existiert zusätzlich eine Spalte «nat», welche die Ziffer der als am natürlichsten angesehenen Antwort enthält. Ausserdem gibt eine Spalte «OK» darüber Auskunft, ob die Gewährsperson für die entsprechende Frage eine brauchbare Antwort geliefert hat. Enthält die Spalte eine Eins, ist dies der Fall. Bei einer Null hat die Person die Frage beispielsweise nicht richtig verstanden, in Schriftsprache übersetzt oder beim Ausfüllen Fremdhilfe angenommen. Weitere Layouts enthalten Tabellen mit ausführlicheren Angaben zu den Gewährspersonen, wie zum Beispiel Informationen über die Herkunft der Eltern oder über Aufenthalte in anderen Dialektregionen, aber auch die zu den Erhebungsorten gehörenden BFS-Nummern. Mittels einer Eingabemaske können pro Tabelle die für den Export gewünschten Spalten ausgewählt werden. Im vorliegenden Fall wird jeweils die Nummer der Gewährsperson zusammen mit den jeweils benötigten Antwortspalten exportiert. Bei den Ankreuzfragen wird nur die Spalte mit den natürlichsten Antworten ausgewählt, da die Analyse nur mit diesen erfolgen soll. Vor dem Export können mit Hilfe einer Suchfunktion und des Felds «OK» nur diejenigen Gewährspersonen ausgewählt werden, welche für die entsprechende Frage eine brauchbare Antwort gegeben haben. Die einzelnen exportierten Tabellen werden nun unter der Nummer der Frage im «.dbf-Format» abgespeichert.

Das Ziel ist, pro Frage ein Shapefile zu erhalten, das pro SADS-Ortschaft nur die Antworten der in Abschnitt 4.1.3 beschriebenen ausgewählten Gewährspersonen enthält, analog zu den SDS-Daten. Zudem sollte pro Antworttyp nur eine Spalte existieren. Die Antworten sollten also gemäss den Angaben in der Liste im Anhang B zusammengefasst werden.

Dafür wird in ArcMap zuerst das Shapefile, welches alle SADS-Orte enthält, über die BFS-Nummer mit der Tabelle verknüpft («Join»), welche die von Hand selektierten Gewährspersonen, inklusive deren eindeutiger Gewährsperson-Nummer, enthält. Das resultierende Shapefile kann nun über die Gewährsperson-Nummer mit jedem einzelnen Antwortfile verknüpft und jeweils in einem Ordner mit der Fragenummer neu abgespeichert werden. Gleichzeitig müssen neue Antwortspalten gebildet werden, welche die Antworten gemäss der Liste im Anhang B in Typen zusammengefasst enthalten. Mit der Funktion «Add Field» werden in ArcMap neue Spalten in die Tabelle eingefügt. Anschliessend können mit dem «Field Calculator» die Werte für diese neuen Spalten berechnet werden, so dass beispielsweise bei Frage 1.01 die neue Spalte des Typs «für (zu)» bei allen Orten eine Eins erhält, welche entweder bei Antwort 1 («für zu») oder bei Antwort 2 («für») eine Eins enthalten. Bei den Fragen, welche nur eine Spalte «nat» enthalten, wird entsprechend für jede einbezogene Antwort eine neue Spalte eingefügt. Mit der Funktion «Select by Attributes» können nun alle Orte ausgewählt werden, welche im «nat»-Feld eine bestimmte Antwort-Ziffer enthalten und die entsprechende neue Spalte kann daraufhin an diesen Orten mit einer Eins gefüllt werden.

Da für jede Frage andere Spalten zusammengefügt werden müssen, ist diese Arbeit von Hand verrichtet worden. Für die weitere Verarbeitung der Daten, von den Shapefiles zur für den Import in die beiden verwendeten dialektometrischen Softwarepaketen geeigneten Form, konnten Python-Skripts erstellt und verwendet werden. Diese Schritte werden im nächsten Kapitel beschrieben.

4.2.3 Umformatierung der Daten

Für die Aufbereitung der beiden Datensätze SDS und SADS für die beiden Programme VDM und Gabmap mussten in Python vier verschiedene Methoden erstellt werden. Jede ruft eine Reihe von Skripts auf, welche die Daten von der Ursprungsform in die jeweils gewünschte Form überführen. Als wichtigste Änderung dabei werden alle Antworten pro Variable, als nominale Variante, in eine Spalte gebracht. Weiter wird dafür gesorgt, dass für jedes Orte-Variablen-Paar ein Eintrag vorhanden ist. Fehlt ein Wert, muss im entsprechenden Feld «Null» beziehungsweise «NA» stehen. Die Skripts inklusive Beschreibung und einer Übersichtstabelle über alle Methoden befinden sich auf der zur Arbeit gehörenden CD. Die folgenden Abschnitte beschreiben die Datenaufbereitung für die beiden Programme Gabmap und VDM.

Datenaufbereitung für Gabmap

Für das Laden in Gabmap müssen alle Daten, welche zusammen analysiert werden sollen in einer «.txt» (Unicode) Tabelle vorliegen, deren Zeilen die Ortsnamen und deren Spalten die Variablen enthalten (Abbildung 7). In den Zellen stehen dann die jeweiligen Varianten in Textform. Es handelt sich dabei um die Spaltennamen der Ausgangs-Shapefiles, die für den entsprechenden Ort eine Eins enthalten haben. Kommen in einem Ort mehrere Varianten vor,

4 Vorgehen

müssen diese mit einem « / » getrennt voneinander in derselben Zelle stehen. Fehlt ein Wert in einer Zelle, muss diese mit «NA» versehen werden. Diese Textdatei wird dann zusammen mit einer «.kml-Datei» in ein Projekt in Gabmap reingeladen. Die «.kml Datei» enthält alle Orte mit Koordinaten und ermöglicht so die Visualisierung der Daten auf einer Karte. Sie kann aus einem Shapefile erstellt werden, das zuerst in ArcGIS in eine «.kmz-Datei» umgewandelt, in GoogleEarth exportiert und dort als .kml-Datei abgespeichert wird. Wichtig ist, dass die Schreibweise der Orte in der Textdatei und in der .kml-Datei exakt übereinstimmen. Für die Benennung der Orte wurde die Schreibweise aus dem SDS-Datensatz übernommen. Gabmap, das seinerseits mit Python-Skripten im Hintergrund arbeitet, ermöglicht so eine Vielzahl von Analysen, auf die in Abschnitt 4.4 noch genauer eingegangen wird. Für die Analysen der vorliegenden Arbeit wurde für jede linguistische Ebene sowie für alle Ebenen zusammen jeweils ein Projekt mit 378 Orten erstellt (für die Vergleiche zwischen den SADS- und den SDS-Daten) sowie zusätzlich für die Ebenen aus dem SDS-Datensatz je ein Projekt mit allen 565 SDS-Orten.

1015 [e] [tt: mhd.e (Primärumlaut)]			1021 Sp[e] [ck: mhd. ë (germanisches e)]			1035 [e] [ng:mhd.e v.Nas+Kon]			101 Finalanschluss: Kleingeld für zum		102 Interrogativpronomen: wem	
Aarau (AG)	e_offe	e_uebe			ae	Aarau (AG)		zum		wem		
Aarberg (BE)	e_offe	e_uebe			ae	Aarberg (BE)		fuer		wem		
Aarburg (AG)	e_offe	e_uebe			ae	Aarburg (AG)		fuer		wem		
Abländschen (BE)	e_offe	e_uebe			e	Abländschen (BE)		fuer		wem		
Adelboden (BE)	e_offe	e_uebe			e	Adelboden (BE)		fuer		wem		
Aesch (BA)	e_offe	e_uebe			ae	Aesch (BA)		um		wem		
Aesch (LU)	e_offe	e_uebe			ae	Aesch (LU)		um		wem		
Aeschi (SO)	e_offe	e_uebe			ae	Aeschi (SO)		fuer		wem		
Agarn (WS)	e_offe	e_uebe			ae	Agarn (WS)		fuer		wem		
Alpnach (UW)	e_gesc	e_uebe			ae	Alpnach (UW)		fuer		xwem		
Alpthal (SZ)	e_gesc	e_uebe			ae	Alpthal (SZ)		fuer		wem		
Altdorf (UR)	e_gesc	e_uebe			ae	Altdorf (UR)		fuer		wem		
Altstätten (SG)	e_gesc	e_neut			e	Altstätten (SG)		zum		wem		
Amden (SG)	e_gesc	e_neut			e	Amden (SG)		um		wem		
Amriswil (TG)	e_gesc	e_neut			e	Amriswil (TG)		zum		wem		
Andwil (SG)	e_gesc	e_neut			e	Andwil (SG)		zum		wem		
Appenzell (AP)	e_gesc	e_neut			NA	Appenzell (AP)		zum		wem		
Arlesheim (BA)	e_offe / e_gesc	e_uebe			ae	Arlesheim (BA)		fuer		wem		

Abbildung 7 Ausschnitt aus der Endtabelle der SDS- (links) und SADS-Daten (rechts) für Gabmap

Datenaufbereitung für VDM

Hinter VDM steht eine Access-Datenbank. Abbildung 8 zeigt deren Schema. Bevor die Daten in die Tabelle «_data» eingefügt werden können, müssen alle anderen mit ihr in Beziehung stehenden Tabellen gefüllt werden. Diese enthalten einerseits Informationen über die Kartentitel, die dazugehörigen Varianten und die linguistischen Bereiche, denen sie angehören sowie andererseits Informationen über die Lage der Orte und die umgebenden Polygone. Letztere sind in der Tabelle «_locations» enthalten, welche separat mittels eines Export-Tools für das Polygonnetz aus ArcGIS geschrieben wird (Castellazzi 2010). Für jedes Ortsnetz muss ein separates VDM-Projekt erstellt werden. Im vorliegenden Fall wird ein Projekt erstellt, das alle SDS-Daten von 565 Orten enthält und ein Projekt, das alle SDS- und alle SADS-Daten von 378 Orten enthält.

4 Vorgehen

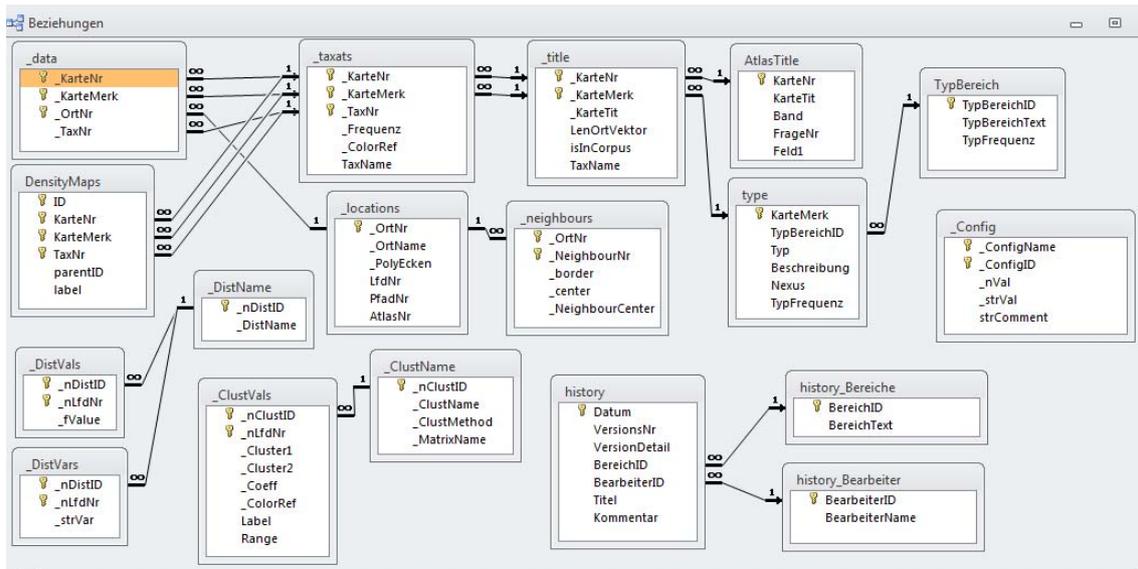


Abbildung 8 Aufbau der Access Datenbank im Hintergrund von VDM

Aus den erstellten Skripten geht für jede Karte eine Excel-Tabelle hervor, welche in der ersten Spalte eine jedem Ort neu zugeteilte ID enthält (VDM_ID 1-378 bzw. 1-565) und in der zweiten Spalte die Nummer der entsprechenden Variante, welche in diesem Ort vorkommt. Diese Nummer entspricht der Position der entsprechenden Varianten-Spalte in der Attributtabelle des Ausgangs-Shapefiles. Bevor diese Daten nun in VDM geladen werden können, muss jede Zeile noch mit der Kartennummer und der Nummer des Kartenmerkmals versehen werden (Abbildung 9). Letzteres ist eine Nummer, welche zuvor in der Tabelle «type» für jedes Phänomen definiert wurde. Alle Tabellen können nun mittels Kopieren und Einfügen aus der Excel-Tabelle untereinander in die Tabelle «_data» in Access eingefügt werden.

	A	B	C	D
1	1015	115	1	1
2	1015	115	2	1
3	1015	115	3	1
4	1015	115	4	1
5	1015	115	5	1
6	1015	115	6	1
7	1015	115	7	2
8	1015	115	8	1

Abbildung 9 Ausschnitt aus der resultierenden Excel-Tabelle vor dem Einlesen in die Access Datenbank (Kartennummer, Kartenmerkmal, VDM_ID, Variantennummer)

VDM kann nur mit einer Antwort pro Ort arbeiten. Deshalb muss für Orte, in denen mehrere Antworten vorkommen, eine Lösung gefunden werden, die Antworten sinnvoll auf eine zu reduzieren. Für die SADS-Daten, bei denen ja eigentlich Antworten von mehreren Gewährspersonen vorliegen, besteht die Möglichkeit jeweils diejenige der zwei oder drei Varianten auszuwählen, die von allen Gewährspersonen in diesem Ort am häufigsten genannt wurde. Diese Möglichkeit besteht beim SDS-Datensatz nicht. Dort kann jedoch bei den meisten Karten, in denen Orte mit Mehrfachnennungen vorkommen, mit einer gewissen Berechtigung die auf der Karte links eingezeichnete als die wichtigere bzw. häufiger verwendete betrachtet werden (Trüb 2003). Da die Reihenfolge der Varianten bei der Digitalisierung nicht mit berücksichtigt werden konnte, sind die SDS-Daten für die Verwendung in VDM vor der Umwandlung durch die Python-Skripte angepasst worden. Ein erstelltes Skript hat dabei alle

Karten ausgewählt, welche in mehr als 20 Orten Mehrfachnennungen aufweisen. Dies ist bei 70 Karten, also rund einem Drittel der Fall⁶. Bei diesen ist dann dafür gesorgt worden, dass immer die linke Variante pro Ort übernommen wird, indem die andere Variante aus dem entsprechenden Ort gelöscht wurde. Karten, auf denen bei allen Orten die selbe Variante links stand und diese in den Antwortspalten ebenfalls ganz links aufgeführt war, konnten so belassen werden, da beim Durchlaufen durch die Umwandlungsprogramme für VDM immer nur die Antwort der ersten, linksten Spalte übernommen wird, wenn Mehrfachnennungen auftreten. Bei einigen Karten (z.B. 3-044) sind die Mehrfachvarianten nicht mit einer links und einer rechts stehenden Variante gekennzeichnet, sondern durch Mischtypen oder Übergangsgebiete dargestellt. Im Falle von Mischtypen ist zufällig eine Variante ausgewählt worden. Bei in Übergangsregionen liegenden Orten, ist diesen zufällig mal die Variante des östlichen und mal diejenige des westlichen Gebietes zugeteilt worden. Die Liste mit allen für VDM angepassten SDS-Karten inklusive Kommentaren mit Entscheidungen, falls die Auswahl der zu übernehmenden Variante nicht eindeutig war, befindet sich im Anhang D.

4.3 Datenverarbeitung und Visualisierung mit VDM

4.3.1 Parameterkarten

VDM bietet die Möglichkeit, die aggregierten Dialekt-Ähnlichkeiten oder Distanzen zwischen den Orten mittels Parameterkarten darzustellen. Da die andere Software Gabmap nur mit Distanzen umgehen kann (Distanzwert = 1 – Ähnlichkeitswert) sollen hier auch in den Analysen mit VDM die Distanzen und nicht die Ähnlichkeiten betrachtet werden. Es steht eine Vielzahl solcher charakteristischer Kennwerte der Distanzen zur Verfügung. Für die vorliegende Arbeit sollen die folgenden drei betrachtet werden, da sie eine direkte linguistische Deutung aufweisen: Mittelwert, Minimum und Schiefe. (Goebel 1984)

Mittelwerte

Die Parameterkarte der Mittelwerte stellt für jeden Ort den Mittelwert der Distanzen zu allen anderen Orten dar. Tiefe Mittelwerte der Distanzen deuten auf eine hohe Kommunikationslage, also auf eine gute Vernetzung hin, während hohe Mittelwerte für eine isolierte Lage des Ortes in Bezug auf die sprachlichen Merkmale zu den anderen Orten stehen (Goebel 1984).

⁶ Im Durchschnitt weisen 64 von 565 Orten (11 %) Mehrfachnennungen auf, bei einer Standardabweichung von 75. Circa 16 % der phonetischen Karten, 33 % der morphologischen und 50 % der lexikalischen Karten verfügen in mehr als 20 Orten über Mehrfachnennungen. Die Mittelwerte der Anzahl Orte mit Mehrfachnennungen bewegen sich von 39 (Phonetik) über 58 (Lexik) bis zu 75 (Morphologie), bei einer Standardabweichung von 16 (Phonetik), 92 (Morphologie) und 52 (Lexik).

Minima

Die Parameterkarte der Minima zeigt für jeden Ort den minimalen Wert aus der Verteilung der Distanzen zu allen anderen Orten an. Auf der Karte werden so Regionen sichtbar, in welchen Orte besonders kleine Distanzwerte zu benachbarten Orten aufweisen. Somit können «Dialektkerne» ausgemacht werden (Goebel 1984).

Schiefen

Die Parameterkarte der Schiefe zeigt für jeden Ort die Symmetrie der Häufigkeitsverteilung der linguistischen Distanzwerte zu allen anderen Orten an. Der Wert Null bedeutet eine perfekte Symmetrie, also dass gleich viele Distanzwerte unterhalb wie oberhalb des arithmetischen Mittelwertes liegen. Rechtsschief heisst, dass die Verteilung auf der rechten Seite flacher abfällt als auf der linken und daher relativ wenige hohe Werte aufweist. Die Werte der Schiefe sind dann positiv. In Bezug auf linguistische Distanzen bedeutet dies, dass der betreffende Ort eine gute kommunikative Lage, also einen hohen Sprachausgleich aufweist. Umgekehrt stehen linksschiefe Verteilungen, also negative Schiefe-Werte für eine schlechte kommunikative Lage und einen geringen Sprachausgleich der betreffenden Orte (Goebel 1984).

Für die Darstellung der Dialektlandschaften mittels der beschriebenen Parameter existieren in VDM drei unterschiedliche Klassenbildungsalgorithmen. Durch die Auswahl eines bestimmten Algorithmus verkleinert oder erhöht sich der Bildkontrast.

MINMWMAX (Minimum – Mittelwert – Maximum)

Die Klassen ober- und unterhalb des theoretischen Mittelwertes⁷ sind gleich breit. Bei sechs Klassen werden also die Differenzen zwischen Mittelwert und Maximum und Mittelwert und Minimum jeweils gedrittelt. Mit diesem Intervallalgorithmus entsteht in der Regel ein ruhiger Flächeneindruck mit einem hohen Bildkontrast (Goebel 1982b). Er wird hier daher bevorzugt.

MEDMW (Median – Mittelwert)

Die Klassen ober- und unterhalb des effektiven Mittelwertes⁸ enthalten möglichst gleich viele Werte. Der Bildkontrast ist nicht so ausgeprägt wie bei MINMWMAX (Goebel 1984).

MED (Median)

Jede Klasse enthält dieselbe Anzahl an Werten (Quantilklassierung). Hier ist der Bildkontrast am geringsten, der entstehende Flächeneindruck ist unruhig (Goebel 1984).

Um einen Vergleich über verschiedene Karten hinweg zu ermöglichen, zum Beispiel von verschiedenen linguistischen Ebenen, können die Werte auch exportiert und beispielsweise in ArcMap neu klassiert und visualisiert werden.

⁷ Mit dem theoretischen Mittelwert ist hier der errechnete Mittelwert aus der Verteilung gemeint, der jedoch nicht zwingend in den Daten vertreten sein muss.

⁸ Der effektive Mittelwert ist derjenige Wert, der in den Daten vorkommt und am nächsten beim theoretischen Mittelwert liegt.

4.3.2 Hierarchische Clusteranalysen

Eine verbreitete Methode für die Visualisierung der linguistischen Distanzen zwischen den Dialektorten ist die hierarchische Clusteranalyse. Hierbei werden die Dialektorte in eine bestimmte Anzahl Cluster eingeteilt, wobei schrittweise die beiden Gruppen mit den geringsten linguistischen Distanzen zusammengefasst werden. Für die Berechnung der Distanzen zwischen den Elementen eines neu gebildeten Clusters zu allen anderen Elementen existieren verschiedene Algorithmen. In VDM stehen sechs verschiedene Algorithmen zur Verfügung: «Single Linkage», «Complete Linkage», «UPGMA», «WPGMA», «Centroid» und «Ward». Für diese Analysen wird vorwiegend mit dem Ward Algorithmus gearbeitet (Ward 1963). Dieser Algorithmus führt immer diejenigen Cluster oder Orte zusammen, bei deren Vereinigung die Zunahme der Varianz innerhalb eines Clusters am geringsten ist (Wilks 2006). Dieses Verfahren scheint sinnvoll und bildet Cluster, die jeweils ungefähr gleich gross sind. Der Nachteil anderer Algorithmen, bei denen Cluster entstehen können, die nur ein Element enthalten, besteht somit nicht. Ungefähr gleich grosse Cluster müssen in der Dialektometrie jedoch nicht immer vorteilhaft sein (Heeringa 2004). Deswegen kann es sinnvoll sein, zum Vergleich auch Analysen mit anderen Algorithmen beizuziehen. Beispielsweise könnte dies der «UPGMA – Unweighted Pair Group Method using Arithmetic averages» Algorithmus oder kurz genannt «Group Average» sein. Wird aus zwei Clustern A und B ein neues Cluster C gebildet, berücksichtigt dieser Algorithmus bei der Berechnung der Distanz dieses neuen Clusters C, zu einem anderen Cluster D, den Durchschnitt der Distanzen zwischen allen Orte-Paaren in A, in B und zwischen A und B. Dies hat den Vorteil, dass jeder Ort dasselbe Gewicht erhält, egal in welchem Cluster er sich befindet. Bei der «WPGMA»-Methode («Weighted Pair Group Method using Arithmetic averages») wird die Distanz aus dem arithmetischen Mittel der durchschnittlichen Distanz aus A und derjenigen aus B gebildet. Die beiden Cluster sind somit gleich gewichtet, was jedoch dazu führt, dass die Orte in kleineren Clustern ein höheres Gewicht erhalten als diejenigen in grösseren Clustern (Heeringa 2004). Im Vergleich zum «Single-Linkage»- und «Complete-Linkage»-Algorithmus haben die Algorithmen, die auf dem arithmetischen Mittel beruhen und auch der Ward-Algorithmus den Vorteil, dass alle Element-Paare zwischen den beiden Clustern berücksichtigt werden und nicht nur dasjenige Paar mit der geringsten, beziehungsweise grössten Distanz. Die «Single-Linkage»- und «Complete-Linkage»-Algorithmen reagieren so besonders sensibel auf Messfehler und Ausreisser (Wagner 2003). Da es beim «Centroid»-Algorithmus manchmal vorkommen kann, dass die Distanzen zwischen zwei Orten in zwei verschiedenen Clustern kleiner sind als zwischen zwei Orten innerhalb eines Clusters (Heeringa 2004), wird dieser hier nicht verwendet. Für eine detaillierte Beschreibung der Algorithmen sei auf Heeringa (2004) verwiesen.

4 Vorgehen

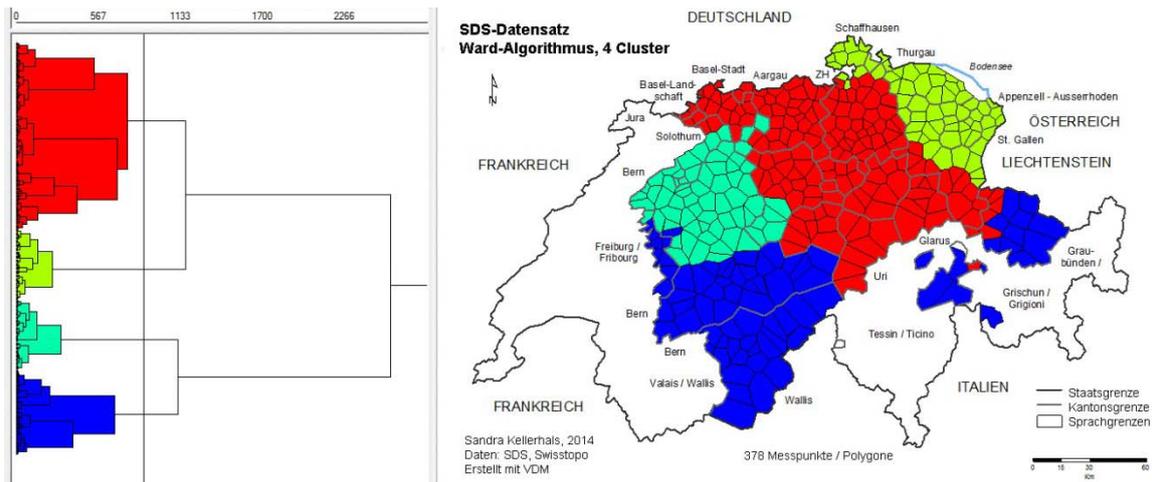


Abbildung 10 Clusteranalyse in VDM mit den SDS-Daten, 4 Clustern und dem Ward Algorithmus

Die Resultate der Clusteranalyse werden in einer Baumstruktur, einem Dendrogramm, dargestellt. Jede Verzweigung entspricht einem Cluster, das aus zwei Subclustern besteht. Die Position der Verzweigung auf der X-Achse entspricht der Distanz zwischen den beiden Subclustern. Diese Distanz wird auch als «cophenetic distance» bezeichnet.

Ein Vorteil der Clusteranalyse mit VDM besteht darin, dass die Schnittpunktlinie, welche die Anzahl Cluster bestimmt, flexibel gewählt werden kann. So kann die Anzahl der dargestellten Cluster nach Einsicht in die Baumstruktur festgelegt werden und die Schnittpunktlinie so gewählt werden, dass sie sich möglichst weit von Verzweigungen entfernt befindet.

4.4 Datenverarbeitung und Visualisierung mit Gabmap

4.4.1 Hierarchische Clusteranalysen

Während in VDM sechs verschiedene Cluster-Algorithmen zur Verfügung stehen, sind es in Gabmap deren vier (ohne Single-Linkage und Centroid). Neben der Berechnung der Cluster bietet Gabmap auch die Möglichkeit, die Cluster mit den Resultaten aus der Analyse der Multidimensionalen Skalierung (siehe nächsten Abschnitt) zu vergleichen und so zu validieren. Ausserdem kann mit Hilfe der Funktion «probabilistic dendrogram» mittels eines sogenannten «Fuzzy Clusterings» überprüft werden, welche Cluster sich bei einer mehrmaligen leichten Veränderung der Daten mit mehrmaligen Durchläufen immer wieder bilden (Leinonen 2013). Die Zahlen im daraus erstellten Dendrogramm (Abbildung 11) geben an, in wie vielen Prozent der Fälle ein Cluster wieder aufgetreten ist. Die Cluster, welche in Abbildung 11 die Zahl 100 aufweisen, sind in jedem Durchgang gebildet worden und können somit als stabil betrachtet werden (Nerbonne 2010). Diese Funktion kann mit zwei verschiedenen Clusteralgorithmen (Group Average UPGMA & Weighted Average WPGMA) oder einer Kombination aus beiden durchgeführt werden (Nerbonne et al. 2011). Die Resultate lassen sich in einer «Fuzzy Cluster»-Karte darstellen (siehe Ende folgender Abschnitt).

4 Vorgehen

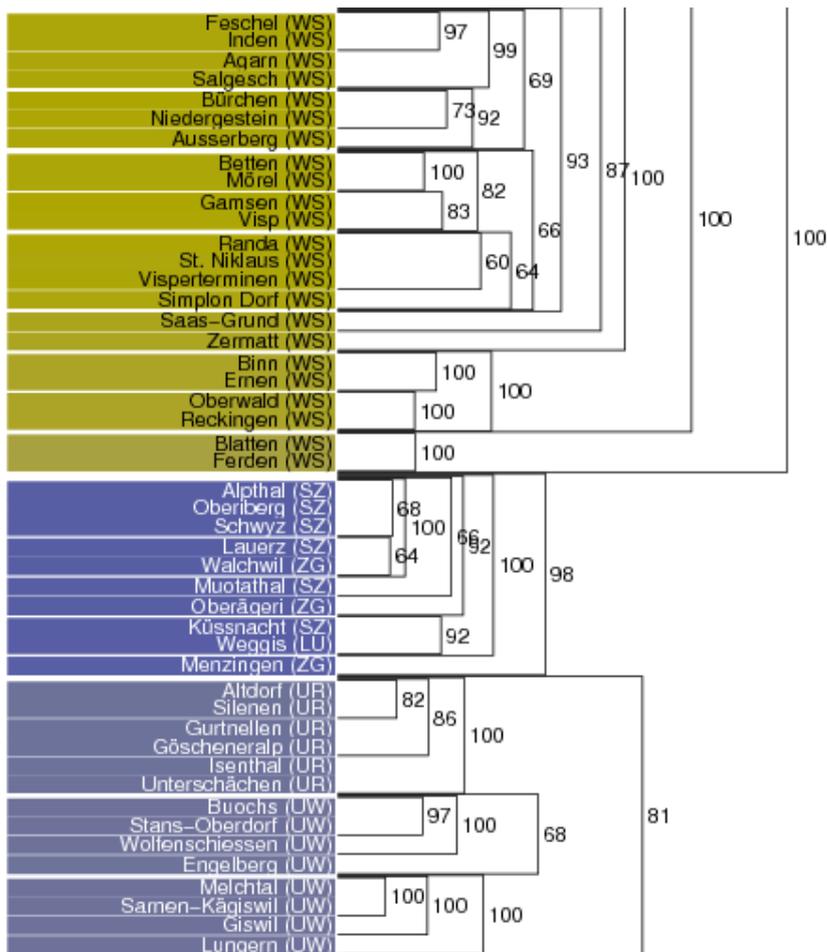


Abbildung 11 Probabilistisches Dendrogramm aus Gabmap

4.4.2 Multidimensionale Skalierung (MDS)

Die Methode der Multidimensionalen Skalierung (MDS) geht auf Torgerson (1952) zurück und wurde von Embleton (1993) erstmals in der Dialektometrie verwendet. Für eine detaillierte Beschreibung der Methode sei auf Heeringa (2004) verwiesen. Kurz erklärt reduziert das Verfahren die vieldimensionale Matrix, welche die linguistischen Distanzen zwischen allen n Ortspunkten in $n-1$ Dimensionen darstellt, auf zwei oder drei Dimensionen, wobei die damit einhergehende Verzerrung der Distanzen möglichst klein gehalten werden soll.

Die Resultate lassen sich in Gabmap auf in zwei oder drei Dimensionen darstellen. In zwei Dimensionen können linguistisch weiter entfernte Orte auch im zweidimensionalen Graphen weiterentfernt dargestellt werden und umgekehrt (Abbildung 12). So lässt sich erkennen, ob die sich durch ihre reduzierte linguistische Distanz unterscheidenden Orte sich gruppieren lassen oder ob die Unterschiede eher kontinuierlich sind und daher eine Clusteranalyse weniger sinnvoll ist. In den beiden Beispielen in Abbildung 12 lässt sich vor allem ein Cluster im unteren Bereich der linken Graphik gut abgrenzen (roter Kreis). Bei den anderen Ortspunkten scheint dies in zwei Dimensionen schwieriger zu sein.

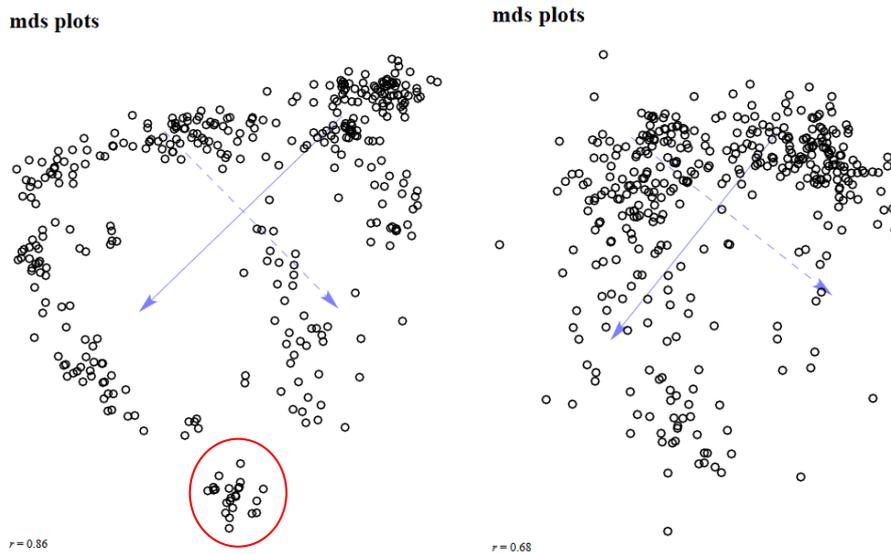


Abbildung 12 MDS Plot mit Daten des SDS (links) bzw. des SADS (rechts)

Eine weitere Visualisierungsmöglichkeit, die Heeringa (2004) eingeführt hat, stellt die Distanzen in drei Dimensionen dar, in dem jeder Dimension ein Farbwert aus dem RGB⁹-Farbraum zugeteilt wird. In den entstehenden Karten erhalten so Ortspaare mit einer geringen linguistischen Distanz ähnelnere Farben als solche mit einer grösseren Distanz. So können Regionen mit kontinuierlich ändernden Dialekten von solchen unterschieden werden, die eher abrupte Grenzen aufweisen.

Die «Fuzzy cluster»-Karte stellt eine Kombination aus MDS und Clusteranalyse dar. MDS wird hier auf die Länge der Äste im Dendrogramm angewandt («cophenetic distance») und nicht auf die ursprünglichen Distanzen. Orte in stabilen Clustern können so farblich von solchen unterschieden werden, die mit einer niedrigeren Wahrscheinlichkeit bei mehrmaligen Durchläufen demselben Cluster zugeteilt werden. In Gabmap wird eine solche Karte immer mit vier Hauptfarben dargestellt. Es ist daher empfohlen, vorher zu eruieren, ob für die verwendeten Daten eine Darstellung mit vier Hauptfarben sinnvoll ist (Leinonen 2013). Abbildung 13 zeigt ein Beispiel für eine solche Karte. Je klarer die Farben blau, grau, rosa und gelb, umso höher ist die Wahrscheinlichkeit, dass die entsprechenden Orte zum jeweiligen Cluster gehören. Sind die Farben durchmischt, wurden die Orte bei verschiedenen Durchläufen verschiedenen Clustern zugeteilt.

⁹ Additiver Farbraum (Rot, Grün, Blau)

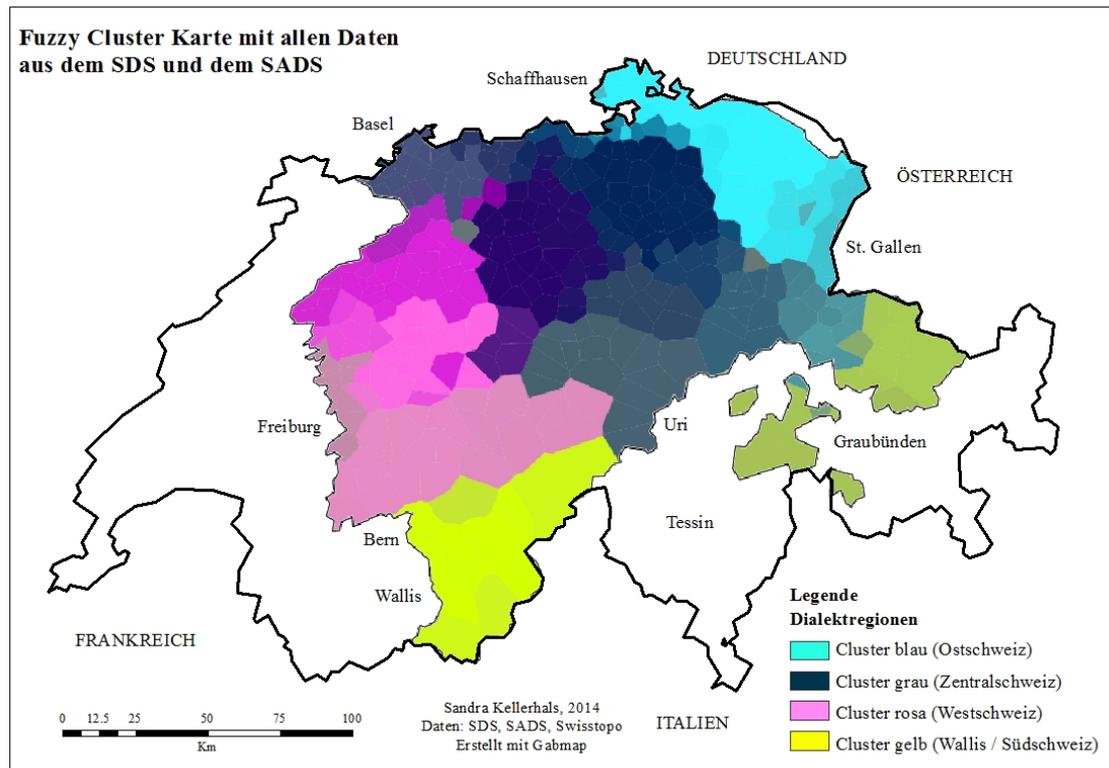


Abbildung 13 Fuzzy Cluster Karte mit allen Daten aus dem SDS und dem SADS («UPGMA»-Algorithmus) und den vier Farben blau, grau, rosa und gelb

4.5 Analysen

4.5.1 Korrelationsanalysen

Aus beiden Systemen VDM und Gabmap können linguistische Distanzmatrizen exportiert und damit statistische Analysen durchgeführt werden. Da VDM nur eine Variante pro Ort berücksichtigen kann, Gabmap aber mehrere, werden die Analysen wenn möglich jeweils mit den aus beiden Systemen exportierten Distanzmatrizen durchgeführt und verglichen.

Aus dem Datensatz mit den 378 Ortspunkten sollen einerseits die Distanzmatrizen der linguistischen Ebenen Morphologie (120 Variablen), Phonetik (75 Variablen), Lexik (36 Variablen) und Syntax (108 Variablen SADS + 4 Variablen SDS) untereinander korreliert werden und andererseits soll die Ähnlichkeit jeder Ebene mit der aggregierten Matrix aller Ebenen bestimmt werden. Es werden jeweils die Pearson Produkt-Moment-Korrelationskoeffizienten sowie Spearmans Rangkorrelationskoeffizienten berechnet. Letztere Methode ist nicht-parametrisch und eignet sich daher besser für nicht normalverteilte Datensätze. Verschiedenen statistischen Tests auf Normalverteilung (Kolmogorov-Smirnov, Lilliefors und Jarque-Bera) zufolge, kann die Nullhypothese, dass die vorliegenden Distanzmatrizen der verschiedenen linguistischen Ebenen normalverteilt sind, verworfen werden. Jedoch sagt der zentrale Grenzwertsatz, dass bei genügend grossem n (hier 71'253

Ortspaare) die Verteilung der arithmetischen Mittelwerte der Zufallsvariablen als annähernd normal betrachtet werden kann (Hogg et al. 2013), was auch Pearsons Korrelationskoeffizienten zulässt.

Für den Vergleich der Distanzen der einzelnen linguistischen Ebenen mit der Distanzmatrix aller Daten stellt sich das Problem, dass nicht allen linguistischen Ebenen dieselbe Anzahl Variablen unterliegt. Somit könnte diejenige Ebene, aus welcher die grösste Anzahl Variablen in die gesamte Matrix einfließt, eventuell deswegen auch den höchsten Korrelationskoeffizienten mit der Gesamtmatrix aufweisen. Dasselbe gilt für die linguistische Ebene, deren Distanzen die höchste Varianz aufweisen. Zur Lösung dieser Probleme werden die Distanzmatrizen der einzelnen linguistischen Ebenen mit einer standardisierten Gesamtmatrix verglichen. Diese wird gebildet, indem die Werte der Matrizen der einzelnen Ebenen standardisiert¹⁰ und die resultierenden z-Werte anschliessend aufsummiert werden. So ist jede linguistische Ebene in der aggregierten Matrix gleich gewichtet und alle vorhandenen Daten können miteinbezogen werden. Zusammen mit den Vergleichen der linguistischen Ebenen untereinander sollte dies ausreichend Aufschluss über deren mögliche Zusammenhänge geben.

Um die Verlässlichkeit der ermittelten sprachlichen Distanzen in den einzelnen linguistischen Ebenen vergleichen zu können, werden für jede Ebenen die Lokale Inkohärenz sowie Cronbachs Alpha ermittelt. Um die Signifikanz der Korrelationen zu bestimmen wird der Mantel-Test verwendet und eine Fishers z-Transformation hilft bei der Überprüfung, ob sich die ermittelten Korrelationskoeffizienten signifikant voneinander unterscheiden. Die folgenden Abschnitte liefern jeweils eine kurze Beschreibung dieser verwendeten Grössen.

Räumliche Autokorrelation: Lokale Inkohärenz

Bei räumlichen Analysen gilt allgemein als Tendenz, dass geographisch nahe beieinander liegende Punkte ähnlicher sind als solche, die weiter entfernt liegen (Burrough und McDonnell 1998). Aus diesem Grundsatz, der auch als räumliche Autokorrelation bezeichnet wird, haben Nerbonne und Kleiweg (2007) ein Mass abgeleitet, das Aussagen über die Verlässlichkeit der Dialektdaten ermöglichen soll. Korrelieren die linguistischen Distanzen eines bestimmten Datensatzes besser mit den geographischen Distanzen (z.B. gemessen in Kilometer Luftlinie) als dies bei einem anderen Datensatz der Fall ist, werden erstere als verlässlicher angesehen. Die Lokale Inkohärenz geht davon aus, dass dieser Zusammenhang vorwiegend lokal und nicht zwingend über grössere Distanzen gilt. Konkret werden für jeden Ortspunkt die acht Orte ermittelt, welche die geringste linguistische Distanz zum untersuchten Ort aufweisen. Anschliessend werden die linguistischen Distanzwerte mit den geographischen Distanzen zwischen diesen Orts-Paaren verglichen. Die Differenzen werden mit der geographischen Distanz normiert und aufsummiert. Höhere Werte der Lokalen Inkohärenz stehen für weniger verlässliche Distanzmessungen als tiefere.

Man könnte argumentieren, dass die geographische Distanz in Kilometern Luftlinie für die Schweiz kein geeignetes Mass ist, da die Topographie nicht mitberücksichtigt wird. Da es hier

¹⁰

$Z = \frac{x-\mu}{\sigma}$, mit x = Ursprungswert, μ = Mittelwert der Population,
 σ = Standardabweichung der Population

jedoch rein um einen Vergleich der Verlässlichkeit zwischen den verschiedenen linguistischen Ebenen geht, wobei die geographischen Distanzen bei allen Ebenen gleich sind, und nicht um die Werte der Korrelation der geographischen mit den linguistischen Distanzen an sich, kann das Mass als hinreichend angesehen werden. Die Lokale Inkohärenz lässt sich mittels einer Funktion in Gabmap ermitteln. Sie wird hier für die einzelnen linguistischen Ebenen berechnet und zwar mit den Daten, die Mehrfachantworten beinhalten.

Verlässlichkeit der Daten: Cronbachs Alpha

Ein weiteres Mass zur Bestimmung der Verlässlichkeit der linguistischen Distanzen in den Dialektdaten pro linguistische Ebene ist Cronbachs Alpha. Cronbach (1951) ist der Begründer der Grösse, welche die interne Konsistenz eines Datensatzes misst und Werte zwischen 0 und 1 annimmt (Tavakol und Dennick 2011). Cronbachs Alpha ist eine Funktion der Anzahl linguistischer Variablen und dem durchschnittlichen Pearson Produkt-Moment-Korrelationskoeffizienten aller Variablen untereinander. Je höher der Wert, umso grösser ist die interne Konsistenz der Daten und somit deren Verlässlichkeit. Als Faustregel haben Nunnally und Bernstein (1994) 0.7 als Grenzwert definiert, ab welchem ein Datensatz für sozialwissenschaftliche Analysen als ausreichend verlässlich angesehen werden kann. Wie auch die Lokale Inkohärenz kann Cronbachs Alpha mittels Gabmap für jede linguistische Ebene mit den Daten, die Mehrfachantworten beinhalten, berechnet werden.

Signifikanz der Korrelationen: Mantel-Test

Um sicher zu stellen, dass die ermittelten Korrelationskoeffizienten zwischen den verschiedenen Distanzmatrizen signifikant sind, wird der Mantel-Test verwendet (Mantel 1967). Er eignet sich besonders gut für Distanzdaten und geht von der Nullhypothese aus, dass zwischen den beiden untersuchten symmetrischen Distanzmatrizen keine Korrelation besteht. Dies würde bedeuten, dass eine Veränderung der linguistischen Distanzen in der einen Matrix keinen Einfluss auf die Distanzen in der anderen Matrix hätte. Die Hypothese wird getestet, indem die Elemente in der ersten Matrix wiederholt neu angeordnet werden, wobei jedes Mal wieder der Korrelationskoeffizient mit der zweiten Matrix berechnet wird. Bei einem Signifikanzniveau von $\alpha = 0.05$ sind circa 1000 Durchläufe nötig (Manly 2007). Würde die Nullhypothese zutreffen, wäre der Korrelationskoeffizient ungefähr gleich oft grösser wie kleiner als der ursprünglich berechnete Korrelationskoeffizient zwischen den beiden unveränderten Matrizen. Sind jedoch in weniger als 5 % der Fälle die Korrelationskoeffizienten grösser als der ursprünglich berechnete, wird die Nullhypothese verworfen und der Koeffizient als signifikant betrachtet (Spruit 2008). Der Mantel-Test wird hier mit Hilfe der Software Matlab durchgeführt.

Signifikanz der Unterschiede der Korrelationskoeffizienten: Fishers z-Transformation

Um zu überprüfen, ob sich zwei ermittelte Korrelationskoeffizienten r_1 und r_2 signifikant unterscheiden, muss für jeden Koeffizienten zuerst eine Fishers z-Transformation durchgeführt werden, da die Korrelationen nicht intervallskaliert sind. Mit Hilfe der beiden so erhaltenen Z^{11} -Werte Z_1 und Z_2 und der Stichprobengröße n kann nun ein z-Wert¹² berechnet werden, welcher mit einem kritischen z-Wert aus der Standardnormalverteilungsfunktion verglichen wird. Ist der ermittelte z-Wert grösser als der kritische z-Wert, so kann der Unterschied zwischen den beiden Korrelationskoeffizienten als signifikant angenommen werden (Bortz 1993). Diese Berechnungen werden in Excel durchgeführt.

4.5.2 Bestimmende Faktoren der Cluster (Cluster determinants)

Für die Beantwortung der zweiten Forschungsfrage, welche linguistischen Phänomene speziell gut trennen, werden bestimmende Elemente pro Cluster gesucht. Dafür muss die Distanzmatrix, welche aus allen vorhandenen SDS- und SADS-Variablen erstellt wird, zuerst mit einem bevorzugten Algorithmus in eine geeignete Anzahl Cluster unterteilt werden, die dann als Referenz für diese Analyse gilt. Dazu wird auf die gesamte Distanzmatrix zuerst eine MDS-Analyse durchgeführt. Anschliessend wird durch Ausprobieren verschiedener Clusteralgorithmen und Anzahl Cluster diejenige Kombination gesucht, welche mittels eines visuellen Vergleichs am besten mit dem Bild aus der MDS-Analyse übereinstimmt. Gleichzeitig wird auch darauf geachtet, dass die Anzahl Cluster, für die durch den gewählten Algorithmus entstandene Baumstruktur, ideal ist. Ideal bedeutet hier, dass sich die Trennung der Cluster möglichst weit entfernt von Knotenpunkten befindet. Nachdem so Dialektregionen (Cluster) gefunden worden sind, können die Variablen gesucht werden, welche diese Cluster determinieren.

In Gabmap gibt es dafür die Funktion «Cluster Determinants» (Cöltekin et al. 2012). Diese berechnet pro Cluster und Variable aus den Distanzen die standardisierten z-Werte, einerseits für die Distanzen innerhalb eines Clusters und andererseits für diejenigen zwischen den Clustern. Ein negativer z-Wert, z.B. -2.5, bedeutet, dass die Distanzen für diese bestimmte Variable 2.5 Standardabweichungen unter dem Erwartungswert liegen. Bei einem positiven z-Wert liegen die Distanzen der Variable über dem Erwartungswert. Nun wird nach denjenigen Variablen gesucht, welche innerhalb des Clusters einen möglichst hohen negativen z-Wert aufweisen und zwischen den Clustern einen möglichst hohen positiven z-Wert. Variablen also, bei welchen in den Orten, die innerhalb eines Clusters liegen möglichst immer dieselbe Ausprägung auftritt und gleichzeitig möglichst eine andere Ausprägung, als in den Orten der restlichen Cluster. Konkret werden für jedes Cluster bis zu zehn Variablen gesucht, welche zwischen dem negativen z-Wert innerhalb und dem positiven z-Wert zwischen den Clustern, die grösste Differenz aufweisen.

¹¹ Der Z-Wert errechnet sich wie folgt: $Z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$

¹² $Z = \frac{Z_1 - Z_2}{\sigma_{Z_1 - Z_2}}$ mit $\sigma_{Z_1 - Z_2} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$

4 Vorgehen

Anschliessend wird mit diesen ausgewählten Variablen nochmals eine MDS- und Clusteranalyse durchgeführt und visuell mit dem Bild der Analyse aller Variablen verglichen. Nun sollten ungefähr dieselben Dialektregionen erkennbar sein. Zum Vergleich kann die MDS- und Clusteranalyse noch mit einer identischen Anzahl zufällig ausgewählter Variablen durchgeführt werden. Dieses Bild sollte sich nun mehr vom Originalen unterscheiden als dasjenige mit den durch die «Cluster Determinants»-Funktion bestimmten Variablen. Zur numerischen Überprüfung kann jeweils eruiert werden, welcher prozentuale Anteil der Orte bei der Clusteranalyse mit den berechneten Variablen und den zufälligen Variablen wieder demselben Cluster zugeteilt worden ist, wie bei der Clusteranalyse mit allen Variablen.

5 Methodenkritik

5.1 Vergleich von VDM und Gabmap

Einige Unterschiede sowie Vor- und Nachteile der beiden dialektometrischen Softwarepakete VDM und Gabmap wurden bereits im Verlaufe der vorangehenden Kapitel erwähnt. Hier sollen diese Eigenschaften nochmals zusammenfassend verglichen und diskutiert werden. Die Tabelle 6 erhebt keinen Anspruch auf Vollständigkeit und bezieht sich vor allem auf die für diese Arbeit verwendeten Funktionen.

	<i>VDM</i>	<i>Gabmap</i>
<i>Basis</i>	Access Datenbank	Web
<i>Erhalt</i>	Auf CD	Frei zugänglich auf dem Web
<i>Projektdatei</i>	Pro Ortsnetz ein Projekt, Möglichkeit der Bildung von Untergruppen der Variablen	Ein Projekt pro Ortsnetz und Gruppe von Variablen, die zusammen analysiert werden sollen
<i>Datenimport</i>	- Relativ aufwändig	+ Relativ einfach (keine Nummerierung der Orte und Varianten nötig)
<i>Varianten pro Variable und Ort</i>	- Nur eine möglich (nummeriert)	+ Mehrere möglich (als Strings)
<i>Übersicht/ Wechsel zwischen den Karten</i>	+ Gute Übersicht über alle Karten + Rasche Wechsel möglich	- Wechsel zwischen den Karten relativ umständlich
<i>Parameterkarten</i>	+ Generierung möglich	- Generierung nicht möglich
<i>Clusteranalyse</i>	+ Flexible Wahl der Anzahl Cluster nach Einsicht des Dendrogramms + Eigene Farbzuteilung möglich + Anzeige der Anzahl und Nummer enthaltener Orte pro Cluster	- Anzahl Cluster muss im Voraus bestimmt werden - Keine eigene Farbzuteilung möglich + Möglichkeit der Clustervalidierung mittels probabilistischen Dendrogramms + Anzeige der Namen enthaltener Orte pro Cluster
<i>Multidimensionale Skalierung</i>	- Nicht möglich	+ Möglich
<i>Kartenrandangaben</i>	+ Fixer Kartenhintergrund mit Kartenrandangaben	- Jede exportierte Karte muss noch mit Kartenrandangaben versehen werden
<i>Weitere nützliche Funktionen</i>	+ Möglichkeit Korrelationen zu visualisieren	+ Möglichkeit zur Bestimmung der «Cluster Determinants» + Automatische Berechnung der Lokalen Inkohärenz und von Cronbachs Alpha

Tabelle 6 Eigenschaften sowie Vor- und Nachteile von VDM und Gabmap

Die Vor- und Nachteile der beiden Systeme in Bezug auf die verwendeten Funktionen halten sich in etwa die Waage. Die beiden Programme ergänzen sich gut. Ein wichtiger Vorteil von Gabmap gegenüber VDM ist der einfachere Datenimport ins System. Einerseits können mehrere Varianten pro Variable berücksichtigt werden, wodurch das Problem der Auswahl einer Hauptvariante wegfällt. Andererseits werden nicht nur numerische Daten akzeptiert. So können die Varianten pro Variable auch als String eingegeben werden und müssen nicht vorher nummeriert werden. Auch ist der Import des Ortsnetzes bei Gabmap, über eine .kml-Datei via GoogleEarth, einfacher als bei VDM, wo ein spezielles Tool für den Export des Ortsnetzes aus ArcMap nötig ist, das bisher nur bis ArcGIS-Version 9.3 implementiert ist (Castellazzi 2010). Dafür kann in VDM eine .wmf-Datei als Metadatei hinterlegt werden, in welcher der Kartenhintergrund mit allen Randangaben gespeichert ist und der hinter jeder Karte eingeblendet wird. So müssen diese Angaben nicht bei jedem Export hinzugefügt werden. Weitere Vorteile von Gabmap gegenüber VDM sind die Visualisierung der linguistischen Distanzen mittels Multidimensionaler Skalierung, die Clustervalidierung mittels eines probabilistischen Dendrogramms sowie die Bestimmung der «Cluster Determinants» und die automatische Berechnung der Lokalen Inkohärenz und von Cronbachs Alpha. Die Hauptvorteile von VDM gegenüber Gabmap beinhalten die Möglichkeit der Analyse durch Parameterkarten und die Vorteile, welche die Clusteranalyse von VDM bietet. Dies sind die flexible Wahl der Anzahl Cluster nach Einsicht der Baumstruktur und die Anzeige der Anzahl im Cluster enthaltener Orte.

5.2 Vergleich der Analysemethoden

Das vergangene Kapitel hat verschiedene Methoden aufgezeigt, mittels derer aus linguistischen Distanzen Dialektlandschaften auf einer Karte visualisiert werden können. Jede Methode vermag einen anderen Aspekt der linguistischen Unterschiede hervorzuheben, weshalb die Analyse hier mittels einer Kombination aus allen beschriebenen Methoden durchgeführt wird. In diesem Abschnitt sollen die Vor- und Nachteile der einzelnen Methoden reflektiert und erklärt werden, weshalb welche Methode verwendet wird.

Tabelle 7 gibt einen Überblick über die Vor- und Nachteile der Methoden. Die Interpretation der Kartenbilder, welche durch eine MDS- oder Clusteranalyse erstellt worden sind, scheint auf den ersten Blick einfacher zu sein, als diejenige der Parameterkarten. Bei den Parameterkarten kann nicht der simple Schluss gezogen werden, dass Orte mit ähnlichen Farben auch ähnliche Dialekte aufweisen. Sie zeigen vielmehr auf, wie sich ein Ort sprachlich im Verhältnis zu allen anderen Orten verhält. Bei der Interpretation solcher Karten muss genau überlegt werden, was diese aussagt und wie das Resultat in einen Gesamtkontext gebracht werden kann. Auch kann der Vergleich solcher Karten zwischen den einzelnen linguistischen Ebenen schwierig sein, da sich für die Darstellung der unterschiedlichen Werte nicht immer dieselbe Skala oder derselbe Klassierungsalgorithmus eignet. Auf der anderen Seite kann bei den Parameterkarten jedem Ort anhand seiner Farbe ein bestimmter Wert bzw. eine bestimmte Werteklasse zugeordnet werden. Bei der Multidimensionalen Skalierung lässt sich hingegen einzig am r-Wert ablesen, wie genau die Darstellung die ursprünglichen Distanzen abbildet. Denn bei dieser Analysemethode werden

die Distanzen zwischen den Orten verzerrt dargestellt. Dafür ist sie gegenüber Änderungen in den Daten stabil und Dialektgrenzen können realitätsnaher auch fließend dargestellt werden, wenn sich keine klaren Grenzen ziehen lassen. Diese beiden Punkte entsprechen genau den Nachteilen der Clusteranalyse. Einerseits können klare und diffuse Grenzen nicht differenziert dargestellt werden, wobei die Clusterbildung in Regionen zu Dialektgrenzen führen kann, wo gar keine wären. Andererseits können geringfügige Änderungen in den Daten oder im Ortsnetz zu grossen Veränderungen in der Clusterbildung führen (Scherrer 2012). Auch hat der gewählte Clusterbildungsalgorithmus sowie die Anzahl Cluster einen grossen Einfluss auf das Ergebnis. Die Resultate der Clusteranalyse sind also für sich alleine schwierig einzuordnen. Deshalb ist es wichtig die Cluster mit den Resultaten der MDS- Analyse zu vergleichen und mittels eines «Fuzzy Clusterings» zu validieren. Der Vorteil der Clusteranalyse ist hingegen, dass Clusterveränderungen, z.B. je nach linguistischer Ebene, auch quantifizierbar sind, da jeweils bekannt ist, welche Elemente bzw. Orte sich in welchem Cluster befinden. So kann nach einer Veränderung festgestellt werden, welcher Anteil der Orte wieder demselben, bzw. einem vergleichbaren Cluster zugeteilt worden ist.

Methode	Vorteile	Nachteile
<i>Parameter-karten</i>	<ul style="list-style-type: none"> • Darstellung der kommunikativen Lage jedes Ortes im Vergleich zu den anderen Orten 	<ul style="list-style-type: none"> • Linguistische Distanzen zwischen den Orten werden nicht direkt dargestellt
<i>Cluster-analyse</i>	<ul style="list-style-type: none"> • Numerische Vergleiche zwischen den resultierenden Karten möglich (Anteil identischer Orte pro Cluster) • Erkennung der bestimmenden Variablen pro Cluster 	<ul style="list-style-type: none"> • Instabil • Erzeugt Illusion von klaren Dialektgrenzen
<i>MDS</i>	<ul style="list-style-type: none"> • Stabil • Dialektgrenzen können scharf oder kontinuierlich dargestellt werden 	<ul style="list-style-type: none"> • Linguistische Distanzen verzerrt • Vergleiche zwischen den resultierenden Karten nur visuell möglich

Tabelle 7 Vor- und Nachteile der verwendeten Methoden

6 Resultate

Das erste Hauptziel der Arbeit besteht darin, die Dialektregionen bzw. die Verteilungen der aggregierten linguistischen Distanzen aus den einzelnen Ebenen zu vergleichen. Die ersten beiden Teile dieses Kapitels zeigen nun die Resultate dieses Vergleichs, zuerst anhand von Parameterkarten, Multidimensionaler Skalierung und hierarchischer Clusteranalyse visuell und anschliessend mittels Korrelationsanalysen numerisch. Im letzten Teil des Kapitels werden die Resultate zur zweiten Hauptforschungsfrage präsentiert, die herauszufinden versucht, ob es linguistische Variablen gibt, welche Dialektregionen speziell gut trennen.

6.1 Visueller Vergleich der Dialektlandschaften

6.1.1 Parameterkarten

Mittelwerte

Abbildung 14 zeigt für die vier Ebenen Lexik, Morphologie, Phonetik und Syntax jeweils eine Karte mit den Mittelwerten der linguistischen Distanzen pro Ort zu allen anderen Orten. Die tiefen Mittelwerte pro Ebene sind jeweils in blau dargestellt, die hohen Werte in rot. Auf den ersten Blick fällt auf, dass die Randgebiete, vor allem im südlichen Teil der Schweiz, vorwiegend hohe Mittelwerte der Distanzen aufweisen. Dies deutet darauf hin, dass diese Regionen sprachlich eher isoliert dastehen. Die zentraler gelegenen Regionen, insbesondere Aargau und Luzern, weisen hingegen tiefere Mittelwerte auf und sind somit sprachlich besser mit den anderen Regionen vernetzt. Am wenigsten deutlich kommt dieser Effekt auf der syntaktischen Karte zum Ausdruck. Auch die isolierte Lage des Wallis, die in allen linguistischen Ebenen auffällt, ist in der syntaktischen Ebene am wenigsten ausgeprägt. Hingegen fällt hier, wie auch auf der morphologischen Karte, die sprachliche Sonderstellung der Region Freiburg besonders auf, während dies auf der phonetischen Karte kaum der Fall ist. Hingegen tritt auf phonetischer Ebene die Isoliertheit der Region Basel leicht in Erscheinung, wenn auch nicht so deutlich, wie auf der lexikalischen Ebene. Die Nordostschweiz ist einzig auf der morphologischen Ebene isoliert. Grundsätzlich kann festgehalten werden, dass die Darstellung der Mittelwerte in der lexikalischen, morphologischen und phonetischen Karte relativ strukturiert erscheint, während sie in der syntaktischen Karte stark fragmentiert ist.

Schaut man sich die Legenden der einzelnen Karten an, fällt auf, dass die Mittelwerte in der Lexik viel höher sind als in den anderen drei Ebenen. Ein Grund dafür könnte die hohe durchschnittliche Anzahl Varianten pro Variable auf der lexikalischen Ebene sein (6.9), die eher zu grösseren linguistischen Distanzen führt.

Die syntaktische Ebene fällt hier mit relativ tiefen Mittelwerten ab. Dies obwohl sie durchschnittlich über mehr Varianten pro Variable verfügt (3.9) als die morphologische (3.6) und phonetische Ebene (2.9). Es ist jedoch denkbar, dass auf der syntaktischen Ebene jeweils eine oder zwei Varianten pro Variable stark dominieren bzw. grossflächig auftreten und die restlichen eher selten vorkommen. Dies würde dann die eher tiefen Distanzen zwischen vielen Orten erklären. In der syntaktischen Ebene weisen deutlich mehr Orte Werte in den untersten drei Klassen auf, als dies in den anderen drei Ebenen der Fall ist. Da sich die Mittelwerte in den vier Karten so stark unterscheiden, ist mit der Darstellung in Abbildung 14 kein Vergleich über alle vier Ebenen hinweg möglich. Die vier Karten in Abbildung 15 zeigen die Mittelwerte anhand einer Klassierung mit gleich grossen Intervallen über alle Werte der vier Ebenen. Die Klassen und deren Farben sind somit über alle vier Karten identisch und ermöglichen einen Vergleich, wobei nicht alle Werteklassen in allen Karten vertreten sind. Die Werteklassen der lexikalischen und syntaktischen Karte überlappen sich sogar überhaupt nicht. Die Farbskala reicht, wie bei den Karten in Abbildung 14, von blau für tiefe Mittelwerte über grün und gelb zu rot für hohe Mittelwerte, wobei extremere Werte dunkler dargestellt sind. Nun kommen die deutlich höheren Mittelwerte der Lexik und die eher tiefen Werte der Syntax zum Ausdruck. Auch hier fällt die Fragmentierung des nördlichen Teils der Schweiz auf der syntaktischen Ebene auf. Die Karten deuten darauf hin, dass sich die Orte auf der syntaktischen Ebene insgesamt weniger gut sprachlich von einander unterscheiden lassen, als in den anderen Ebenen.

6 Resultate

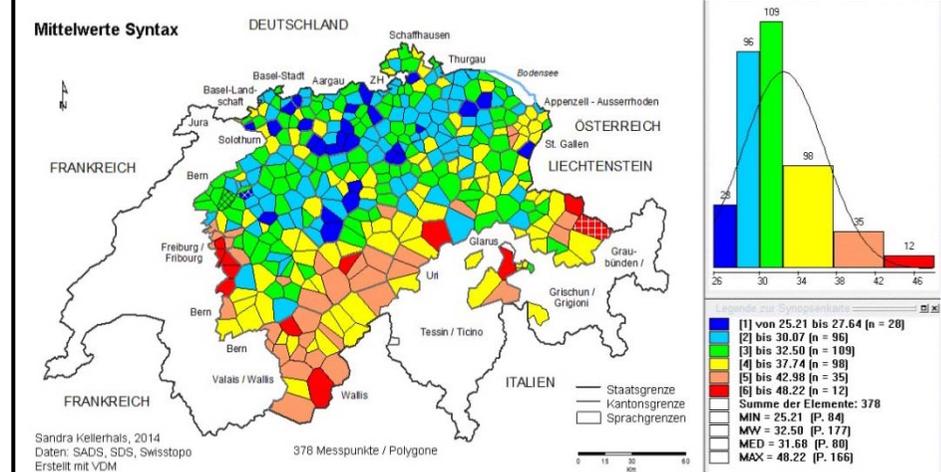
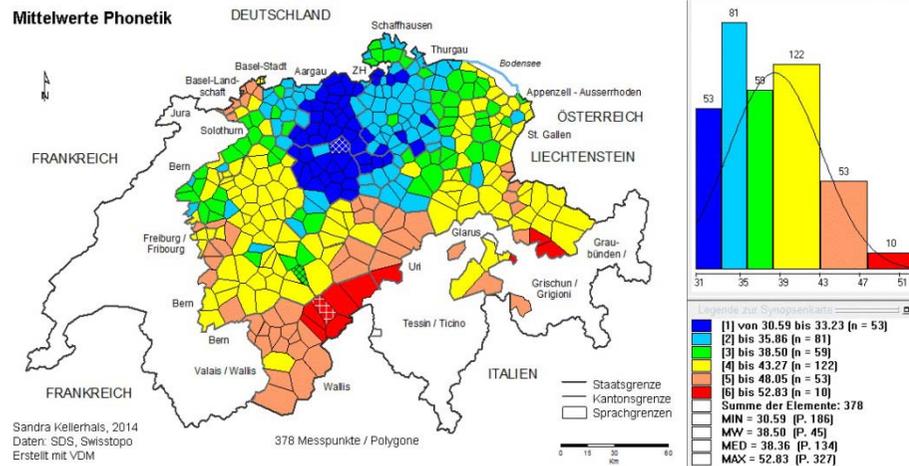
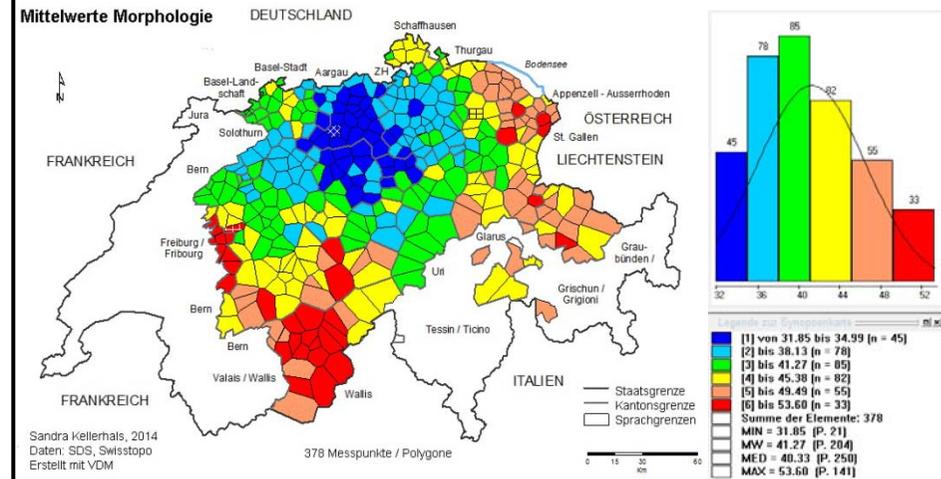
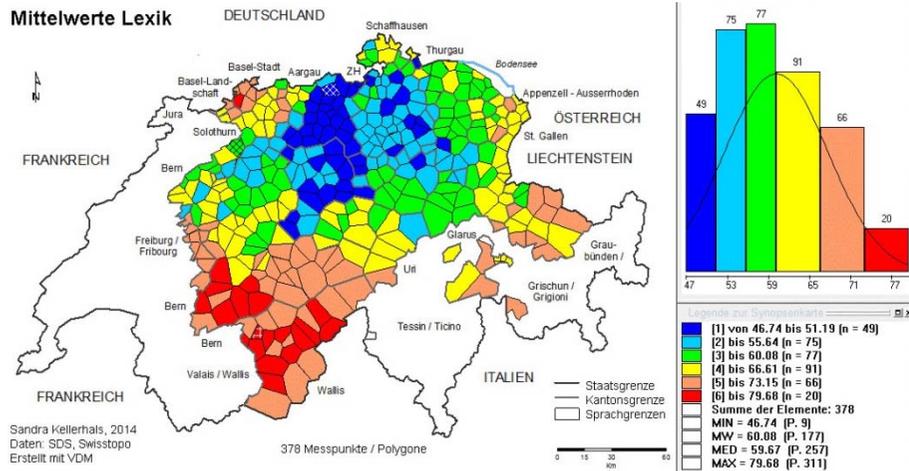


Abbildung 14 Mittelwerte der linguistischen Distanzen, Klassenbildung «MINMWMAX», Lexik, Morphologie, Phonetik und Syntax

6 Resultate

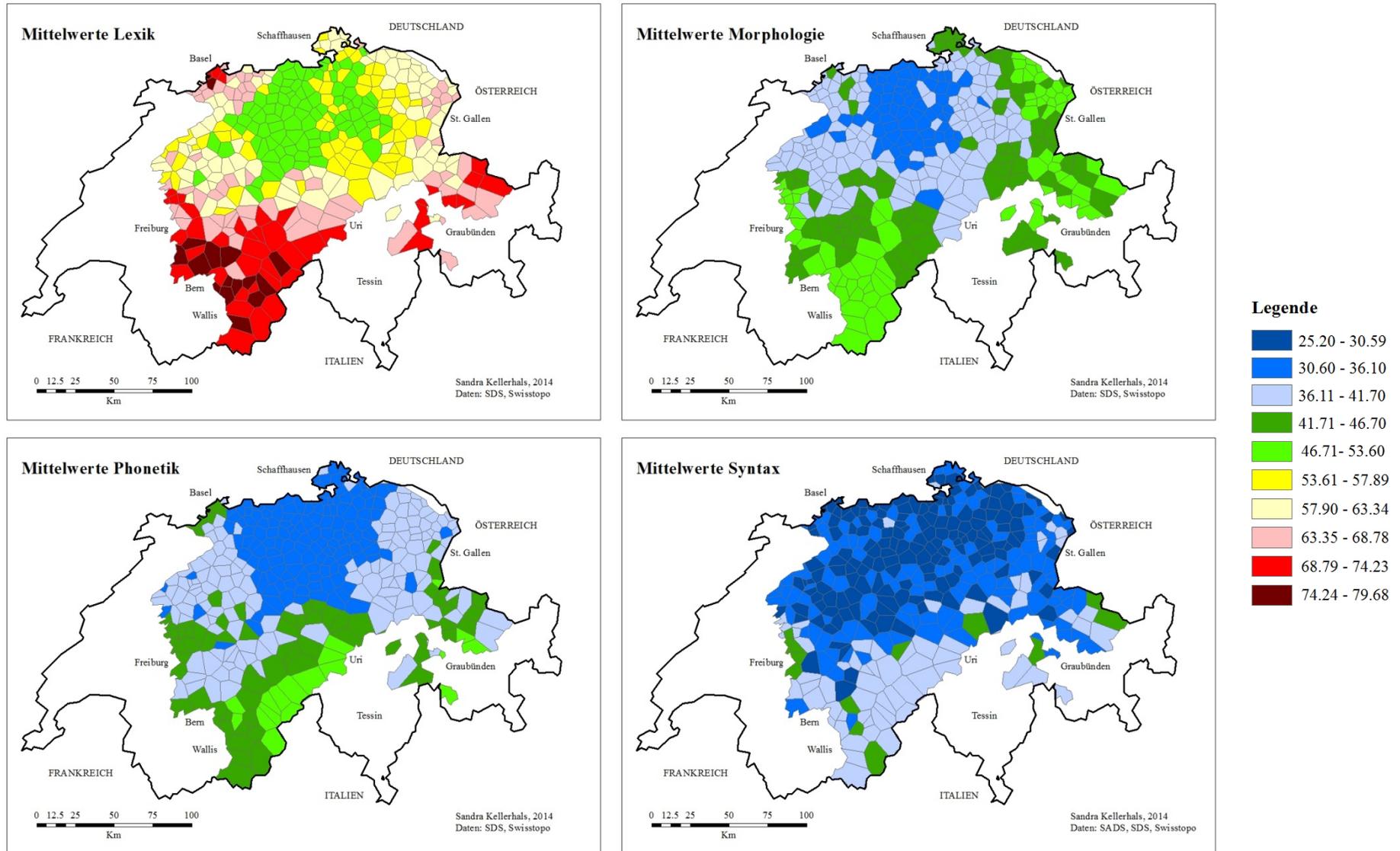


Abbildung 15 Mittelwerte der linguistischen Distanzen, Klassenbildung «gleiche Intervalle», Lexik, Morphologie, Phonetik und Syntax

Minima

Die Parameterkarten der Minima in Abbildung 16 zeigen für jede linguistische Ebene diejenigen Regionen, in welchen Orte besonders kleine Distanzwerte zu benachbarten Orten aufweisen. Die phonetische Ebene verfügt über die meisten solcher Dialektkerne und über besonders viele Orte, welche zu einem benachbarten Ort sogar eine linguistische Distanz von Null aufweisen (26, weisse Schraffur auf blauer Fläche). Im Vergleich dazu weist die Lexik sechs solche Orte auf, die Morphologie deren drei und die Syntax gar keine. Die Dialektkerne in Glarus, Bern, Luzern und Zürich, welche sich in der phonetischen Ebene erkennen lassen, sind zum Teil identisch mit den Dialektkernen in der lexikalischen Ebene. Die Minima in der morphologischen Ebene stimmen jedoch nicht mit denjenigen der anderen Ebenen überein. In der syntaktischen Ebene kommen wenige bis gar keine Dialektkerne vor. Die unterste Klasse enthält Distanzwerte von 6.54 bis 11.61. Ähnliche Werte entsprechen in der lexikalischen Ebene der zweituntersten und in der morphologischen und phonetischen Ebene bereits der drittuntersten Klasse. Die kleinsten Werte, welche in Orten im Süden Luzerns in der syntaktischen Ebene erkennbar sind, stimmen allerdings überein mit den Orten in der Phonetik und teilweise in der Lexik. Eine mögliche Erklärung für die fehlenden Dialektkerne auf der syntaktischen Ebene ist, dass aus dieser Ebene viele Variablen mit in die Analyse einfließen und andererseits auch relativ viele Varianten pro Variable vorkommen. Somit sind die Möglichkeiten für die Orte sich in irgendeiner Variable zu unterscheiden grösser als dies auf der phonetischen oder lexikalischen Eben der Fall ist, wo wenige Varianten pro Variable vorkommen (Phonetik) oder wenige Variablen in die Analyse miteinfließen (Lexik).

Zusammenfassend lässt sich in Bezug auf die syntaktische Ebene bisher sagen, dass die Orte sich linguistisch zwar insgesamt schwächer voneinander unterscheiden als in den anderen Ebenen (tiefe Mittelwerte der Distanzen), dass es jedoch keine Orte gibt, die sich gar nicht unterscheiden und relativ wenige, deren linguistische Distanzen nur leicht voneinander abweichen. Dies deutet darauf hin, dass sich in der syntaktischen Ebene wohl am schwersten klar trennbare Dialektregionen finden lassen.

Ein interessanter Aspekt soll hier noch in Bezug auf die morphologische Ebene hervorgehoben werden. Vergleicht man die Karte, die anhand der 378 Orte erstellt wurde mit derjenigen, die alle 565 SDS-Orte berücksichtigt, stellt man fest, dass sich der grösste Teil der zusätzlichen Orte in der untersten Klasse einfügt, gefolgt von der zweit tiefsten Klasse. So tauchen auf der Karte mit 565 Ortspunkten deutlich mehr Dialektkerne auf als auf derjenigen mit 378 Orten. Das heisst, dass überdurchschnittlich viele zusätzliche Orte in der 565er-Karte zu mindestens einem anderen Ort eine sehr kleine linguistische Distanz aufweisen. Abbildung 17 gibt eine Übersicht über die Unterschiede der beiden morphologischen Karten. In den anderen Ebenen sieht die Karte mit 378 praktisch identisch aus wie diejenige mit 565 Ortspunkten, da sich die meisten zusätzlichen Orte ungefähr gleichverteilt hauptsächlich in den beiden mittleren Klassen eingliedern. Als Beispiel ist dies anhand der lexikalischen Ebene in Abbildung 18 ersichtlich. Beide Beispiele deuten darauf hin, dass keine grösseren Informationsverluste auftreten, wenn nur 378 Orte betrachtet werden. Auf der morphologischen Ebene ist der grösste Teil der wegfallenden Orte zu mindestens einem verbleibenden Ort sprachlich sehr ähnlich. Die Beispiele der anderen linguistischen Ebenen zeigen, dass die 378 Orte als repräsentative

Auswahl der 565 Orte betrachtet werden können. Auch bei Vergleichen mit anderen Parameterkarten (z.B. Mittelwerten) zwischen Kartenbildern mit 378 und 565 Ortspunkten sieht das Bild jeweils praktisch identisch aus, mit dem einzigen Unterschied, dass dasjenige mit 565 Orten feiner aufgelöst ist.

Schiefen

Abbildung 19 zeigt für jeden Ort pro Ebene die Symmetrie bzw. Schiefe der Häufigkeitsverteilung der linguistischen Distanzen zu allen anderen Orten. Positive Schiefe-Werte sind in Rottönen dargestellt, negative in Blautönen und Werte um Null in Lilatönen. Orte mit positiven Schiefe-Werten weisen eine rechtschiefe Verteilung und daher relativ wenig hohe, dafür viele tiefe, Distanzwerte auf. Dies deutet auf einen hohen Sprachausgleich der betreffenden Orte mit den anderen Orten hin, da sie sich von eher wenigen Orten stark unterscheiden. Ähnlich wie bei den Mittelwerten fällt hier zuerst der starke Gegensatz zwischen der lexikalischen und der syntaktischen Ebene auf. Während in der lexikalischen Ebene viele zum Teil stark negative Schiefe-Werte vorkommen, überwiegen auf syntaktischer Ebene die positiven Werte. Die morphologische und phonetische Ebene befinden sich zwischen den beiden Extremen. Bei der Phonetik fällt aufgrund der häufigen negativen Schiefe-Werte auf, dass fast in der gesamten Schweiz ein geringer Sprachausgleich zwischen den Orten herrscht, abgesehen von einem Band, das sich vom Norden Aargaus bis in den Süden St. Gallens zieht. Dies bedeutet, dass in vielen Orten wenig tiefe Distanzwerte zu anderen Orten vorkommen. Bei den Parameterkarten der Minima hingegen wurde deutlich, dass diese wenigen tiefen Werte dann sehr tief sind. Die morphologische Ebene weist grundsätzlich ebenfalls einen eher geringen Sprachausgleich auf, abgesehen von den Regionen Luzern, Aargau, Zürich und Nordostschweiz, wo relativ viele Werte um Null herum vorkommen. In allen Karten grenzt sich das Wallis als eine Region mit besonders geringem Sprachausgleich ab. Ebenso lässt sich grundsätzlich erkennen, dass der Sprachausgleich gegen Süden und Westen geringer ist, als gegen Norden und Osten. Auf der syntaktischen Ebene zeigt der hohe Sprachausgleich vieler Orte, ausser den im Wallis gelegenen, auch hier wieder, dass es eher schwierig ist, die Orte nach ihren syntaktischen Eigenschaften von einander abzugrenzen.

6 Resultate

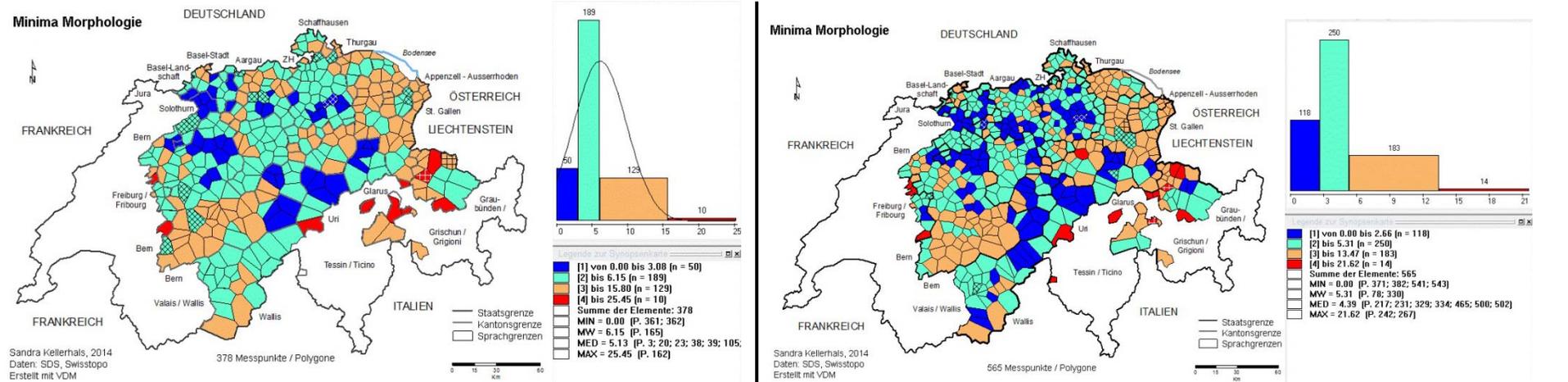


Abbildung 17 Vergleich der Minima der linguistischen Distanzen auf morphologischer Ebene, links mit 378, rechts mit 565 Ortspunkten

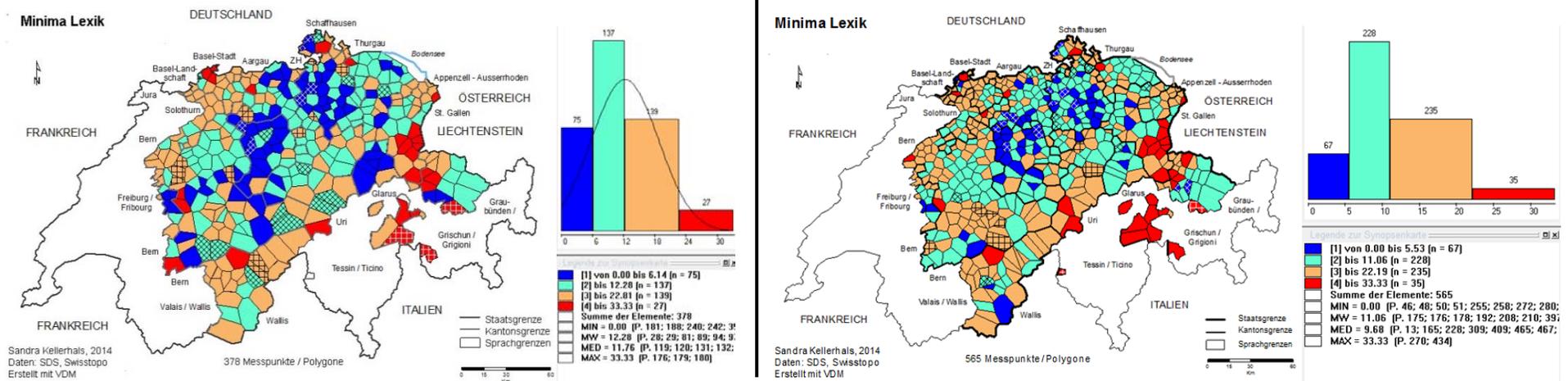


Abbildung 18 Vergleich der Minima der linguistischen Distanzen auf lexikalischer Ebene, links mit 378, rechts mit 565 Ortspunkte

6 Resultate

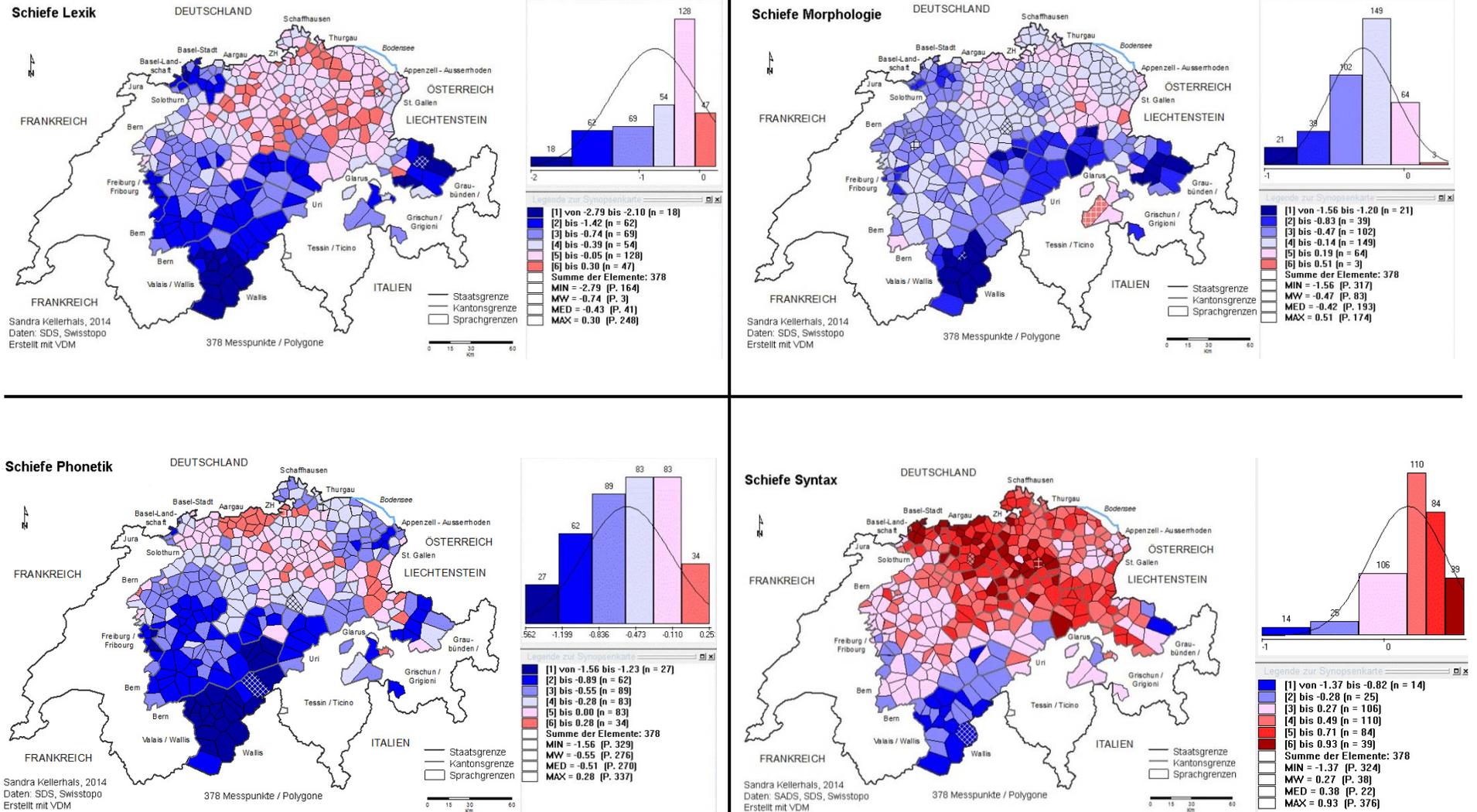


Abbildung 19 Schiefe der Verteilung der linguistischen Distanzen, Klassenbildung «MINMWMAX», Lexik, Morphologie, Phonetik und Syntax

6.1.2 Multidimensionale Skalierung

Die Multidimensionale Skalierung erfüllt hier zwei verschiedene Funktionen. Erstens werden die Resultate aus der MDS-Analyse mit allen Daten des SDS- und des SADS visuell mit denjenigen der einzelnen Ebenen verglichen. Zweitens dient die MDS-Analyse aller SDS- und SADS-Daten als Vorlage für die Clusteranalyse, welche im nächsten Abschnitt besprochen wird.

In oberen Teil von Abbildung 20 ist im Grossformat das Resultat der MDS-Analyse aller SDS- und SADS-Daten ersichtlich, während die Resultate der Analysen der einzelnen Ebenen im Kleinformat darunter zu finden sind. In allen Karten lassen sich, trotz der zum Teil kontinuierlichen Übergänge, grob Dialektregionen erkennen. Die erkannten Regionen werden im Folgenden kurz definiert:

Lexik (links oben) → 9 Dialektregionen

- | | |
|------------------------------|--|
| 1) Basel, Solothurn Nord | 6) Zürich, St. Gallen Nord, AR, Glarus |
| 2) Solothurn Süd, Bern Nord | 7) Ostschweiz (SH, TG, AI) |
| 3) Berner Oberland, Freiburg | 8) Zentralschweiz (OW, NW, UR, SZ, ZG Süd) |
| 4) Wallis | 9) St. Gallen Süd, Graubünden |
| 5) Aargau, Luzern, Zug Nord | |

Morphologie (rechts oben) → 8 Dialektregionen

- | | |
|--------------------------------|--|
| 1) Basel, Solothurn, Bern Nord | 5) Zürich |
| 2) Berner Oberland, Freiburg | 6) Ostschweiz (SH, TG, AR, AI, SG Nord) |
| 3) Wallis, Graubünden Süd | 7) Zentralschweiz (OW, NW, UR, SZ, GL, SG Süd) |
| 4) Aargau, Luzern, Zug Nord | 8) Graubünden Nord |

Phonetik (links unten) → 10 Dialektregionen

- | | |
|--------------------------------|---|
| 1) Basel | 6) Zürich, Schwyz Nord, St. Gallen Süd |
| 2) Solothurn, Bern Nord | 7) Nordostschweiz (SH, TG, SG Nord, AR, AI) |
| 3) Berner Oberland West, FR | 8) Zentral-/Ostschweiz (ZG Süd, SZ Süd, GL) |
| 4) Wallis, Berner Oberland Ost | 9) Zentralschweiz (OW, NW, UR) |
| 5) Aargau, Luzern | 10) Graubünden |

Syntax (rechts unten) → 6 Dialektregionen

- | | |
|--|--|
| 1) BS, AG, LU, ZG, SZ GL,
OW, NW, UR Nord | 4) Wallis, Uri Süd, Freiburg |
| 2) BL, SO, BE Nord | 5) Nordostschweiz (ZH, SH, TG, SG, AR, AI) |
| 3) Berner Oberland | 6) Graubünden |

Alle SDS- und SADS- Daten → 10 Dialektregionen

- | | |
|------------------------------|--|
| 7) Basel, Solothurn Nord | 12) Zürich |
| 8) Solothurn Süd, Bern Nord | 13) Nordostschweiz (SH, TG, SG Nord, AR, AI) |
| 9) Berner Oberland, Freiburg | 14) Zentralschweiz (OW, NW, UR, SZ, ZG Süd) |
| 10) Wallis | 15) Glarus und SG Süd |
| 11) Aargau, Luzern, Zug Nord | 16) Graubünden |

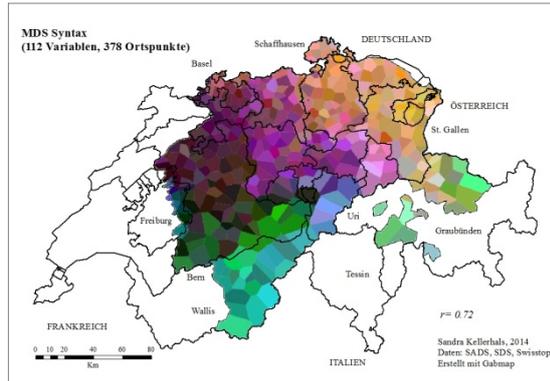
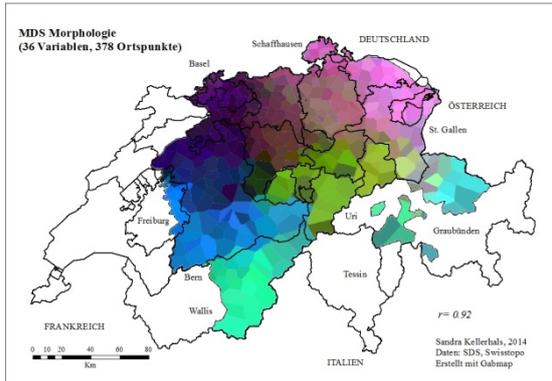
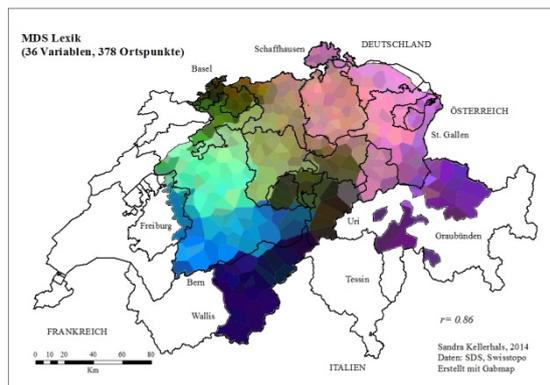
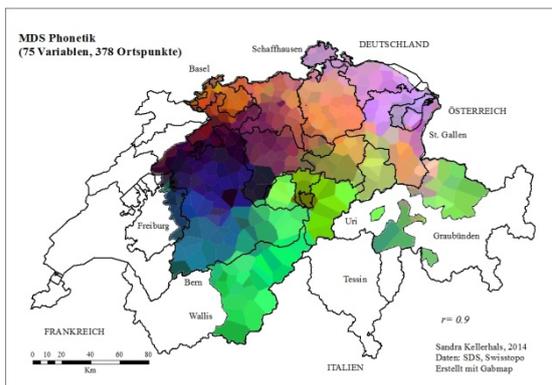
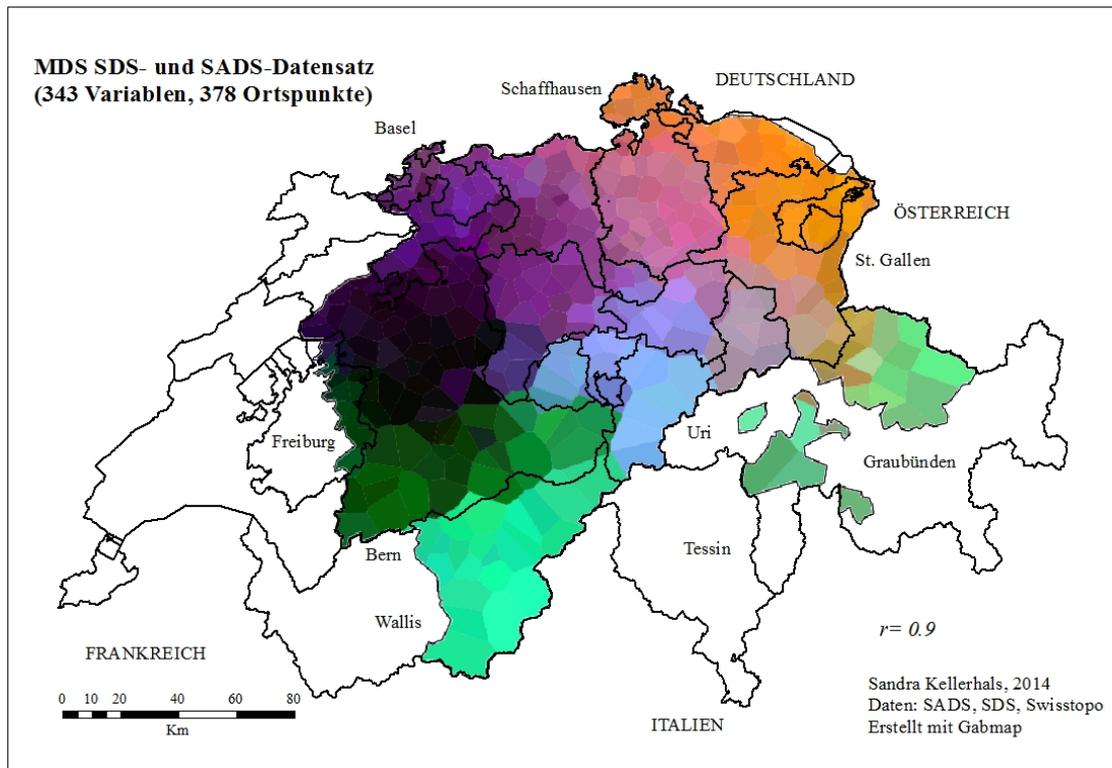


Abbildung 20 MDS aller SDS- und SADS-Daten (oben gross) und der einzelnen Ebenen (unten klein)

Grundsätzlich haben die MDS-Karten der lexikalischen, morphologischen und phonetischen Ebene eine ähnliche Struktur wie die MDS-Karte aus allen SDS- und SADS-Daten. Die syntaktische MDS-Karte weist hingegen weniger gut unterscheidbare Dialektregionen auf. Zudem sind die syntaktischen Dialektregionen viel stärker fragmentiert. An dieser Stelle sei auf die Diskussion in Kapitel 7 verwiesen, welche die gefundenen Dialektregionen mit traditionellen dialektologischen Grenzen vergleicht, die aus Hotzenköcherle et al. (1984) bekannt sind.

Wie im Abschnitt 4.4.2 bereits beschrieben, lässt sich eine MDS auch in zwei Dimensionen darstellen. Sich linguistisch näherstehende Orte werden darin auch in zwei Hauptdimensionen näher platziert. Eine solche Darstellung ist vor allem im Hinblick auf die Clusteranalyse interessant. Es lässt sich nämlich feststellen, ob sich bereits eine natürliche Gruppierung der Orte in zwei Dimensionen ergibt. So kann besser abgeschätzt werden, wie sinnvoll eine Einteilung in Cluster ist. Abbildung 21 zeigt jeweils die zweidimensionale MDS-Analyse der Daten jeder linguistischer Ebene.

Auf der syntaktischen Ebene scheint es besonders schwierig zu sein, Cluster zu bilden. Im unteren Teil der Darstellung lassen sich am ehesten noch einige Gemeinden im Freiburg (grüner Bereich), einige im Wallis (rot) und einige im Graubünden (blau) abgrenzen. In den anderen Ebenen finden sich zumindest teilweise deutlichere Abstände zwischen einigen Punktegruppen. Im nächsten Abschnitt wird dennoch versucht die Dialektlandschaft aller linguistischen Ebenen der Schweiz mittels einer hierarchischen Clusteranalyse abzubilden.

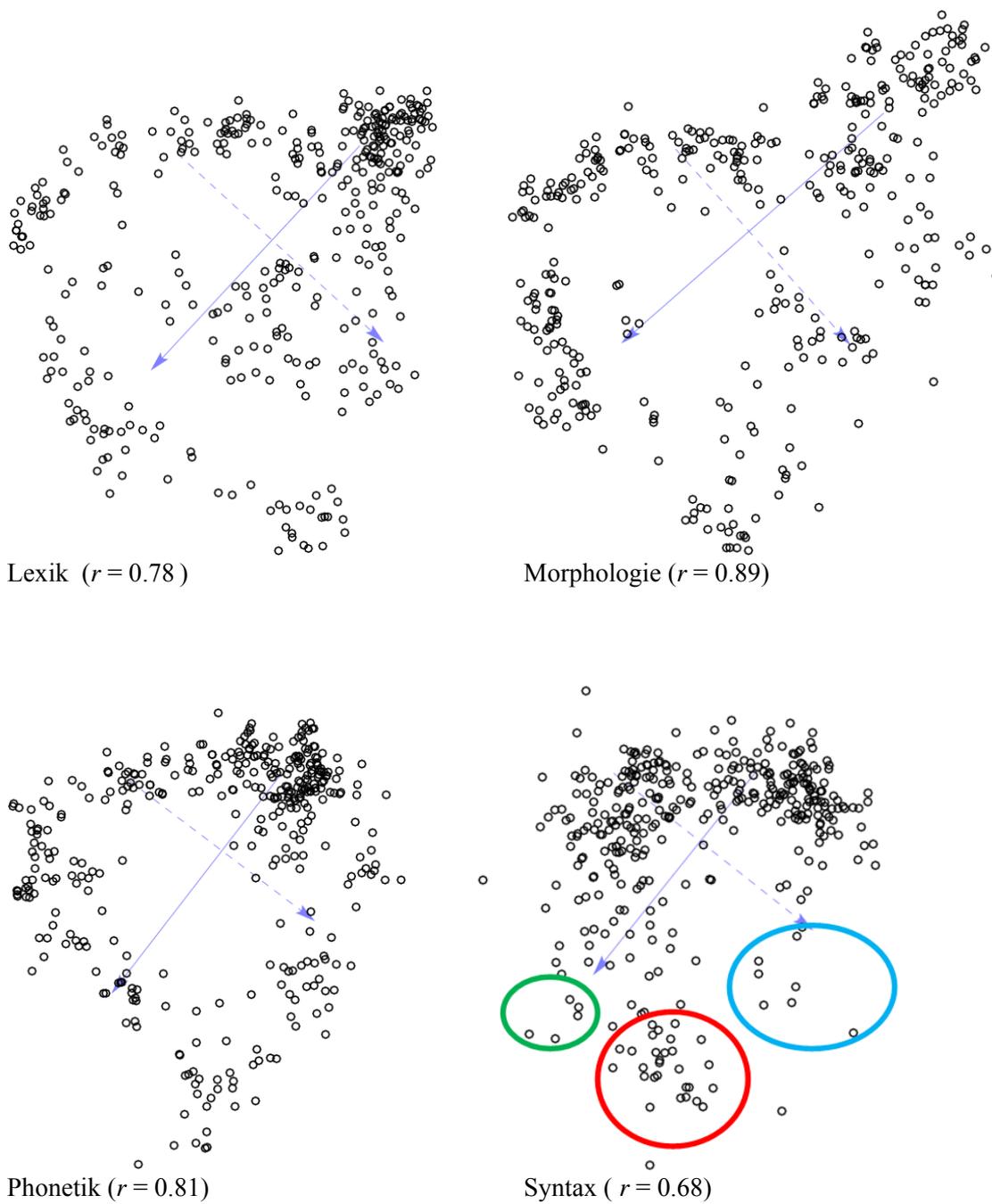


Abbildung 21 MDS-Graphiken (zweidimensionale Abbildungen der linguistischen Distanzen der 378 Orte) der vier linguistischen Ebenen

6.1.3 Hierarchische Clusteranalyse

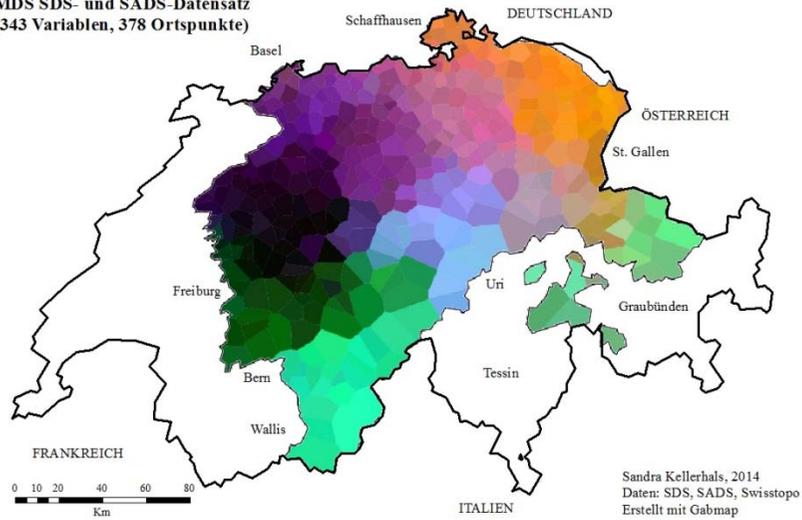
Wie bereits im Abschnitt 5.2 erwähnt, ist die Auswahl des Clusterbildungsalgorithmus sowie die Anzahl Cluster entscheidend für das Resultat der Clusteranalyse. Nach einem Vergleich der Resultate aus der MDS-Analyse mit den Resultaten verschiedener Clusteranalysen hat sich der Ward-Algorithmus als der stabilste und geeignetste Clusterbildungsalgorithmus herauskristallisiert.

Abbildung 22 zeigt nochmals das Resultat aus dieser MDS-Analyse sowie zum Vergleich dasjenige des «Fuzzy Clusterings» mit dem UPGMA-Algorithmus. Darunter sind jeweils die Resultate aus der Clusteranalyse, generiert mit dem Ward-Algorithmus und zehn Clustern, mit allen SDS- und SADS-Daten einmal mit Gabmap und einmal mit VDM abgebildet. Wie aus der Abbildung ersichtlich ist, bildet die Clusteranalyse die Resultate aus den anderen Analysen ziemlich genau ab. Auch stimmt die Clusterbildung, welche mit den VDM-Daten, mit nur einer Variante pro Variable und Ort zustande gekommen ist, weitgehend mit derjenigen aus Gabmap überein, wo mehrere Varianten pro Variable und Ort möglich sind. Bei stichprobenartigen Versuchen mit anderen Clusterbildungsalgorithmen als dem Ward-Algorithmus ist dies hingegen teilweise nicht der Fall gewesen.

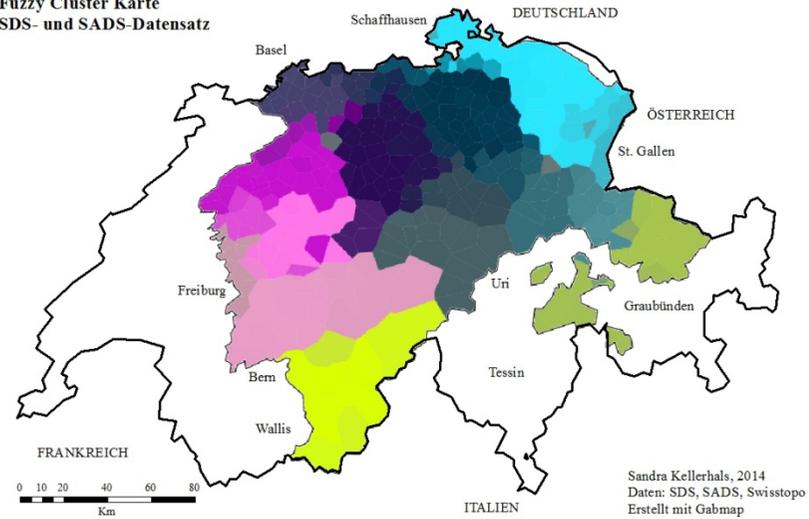
Zur weiteren Überprüfung der Cluster kann zum einen die zweidimensionale MDS-Analyse aller SDS- und SADS-Daten zu Hilfe genommen werden, welche entsprechend der gebildeten Cluster eingefärbt worden ist (Abbildung 23). Darauf ist ersichtlich, wie gut sich die Cluster in zwei Dimensionen wirklich unterscheiden. Wie in Abbildung 23 zu sehen ist, lassen sich vor allem die Orte im nördlichen Teil der Schweiz relativ schlecht in Cluster einteilen. Zum anderen kann im Dendrogramm eingesehen werden, wie weit sich die Schnittlinie bei zehn Clustern von Schnittpunkten entfernt befindet (Abbildung 24). Je weiter entfernt sich die Schnittlinie von den Knotenpunkten befindet, umso eher sind die Cluster gerechtfertigt. Wie aus Abbildung 24 ersichtlich ist, befindet sich die Schnittlinie nirgendwo unmittelbar bei einem Knotenpunkt.

6 Resultate

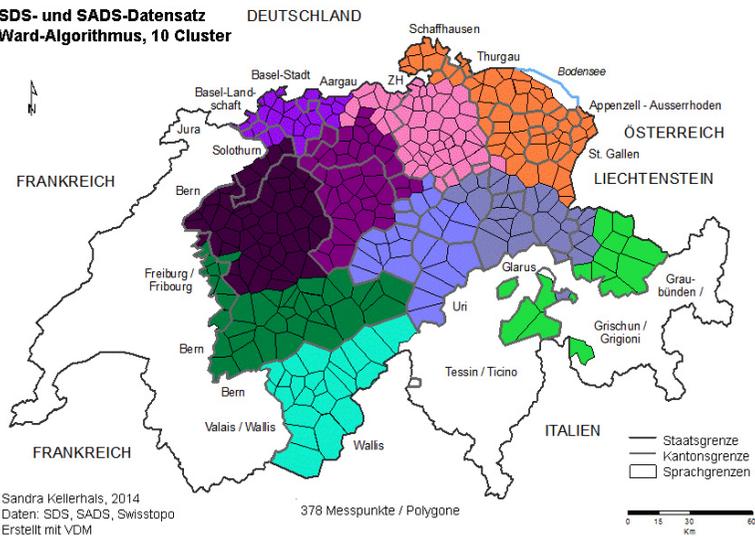
**MDS SDS- und SADS-Datensatz
(343 Variablen, 378 Ortspunkte)**



**Fuzzy Cluster Karte
SDS- und SADS-Datensatz**



**SDS- und SADS-Datensatz
Ward-Algorithmus, 10 Cluster**



**SDS- und SADS-Datensatz
Ward-Algorithmus, 10 Cluster**

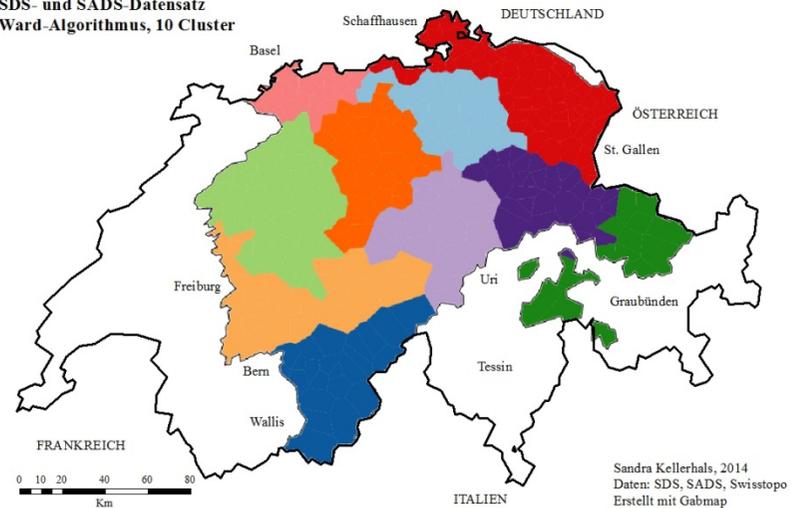


Abbildung 22 MDS-Karte, «Fuzzy Cluster»-Karte («UPGMA»-Algorithmus), Clusterkarte Ward-Algorithmus (10 Cluster) VDM und Gabmap

6 Resultate

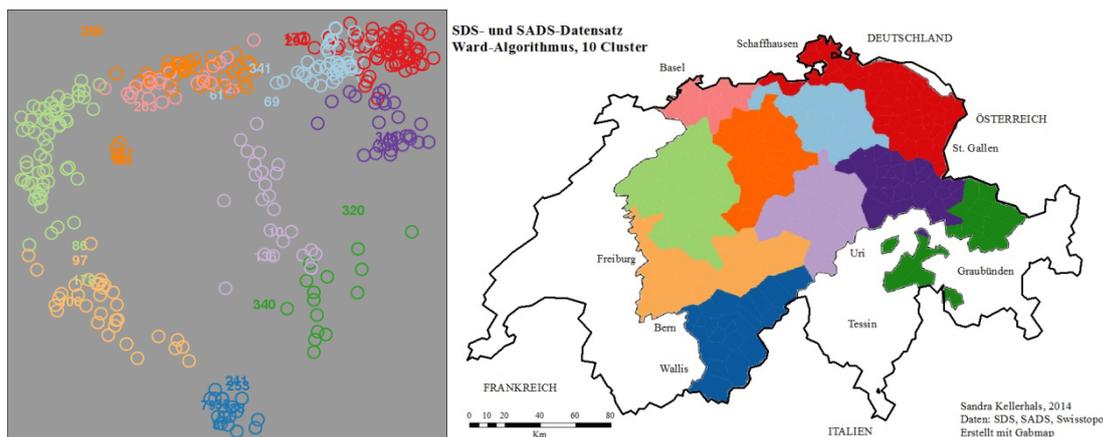


Abbildung 23 Clusterkarte aller SDS- und SADS-Daten, Ward-Algorithmus (10 Cluster) mit zweidimensionaler MDS-Graphik (Gabmap)

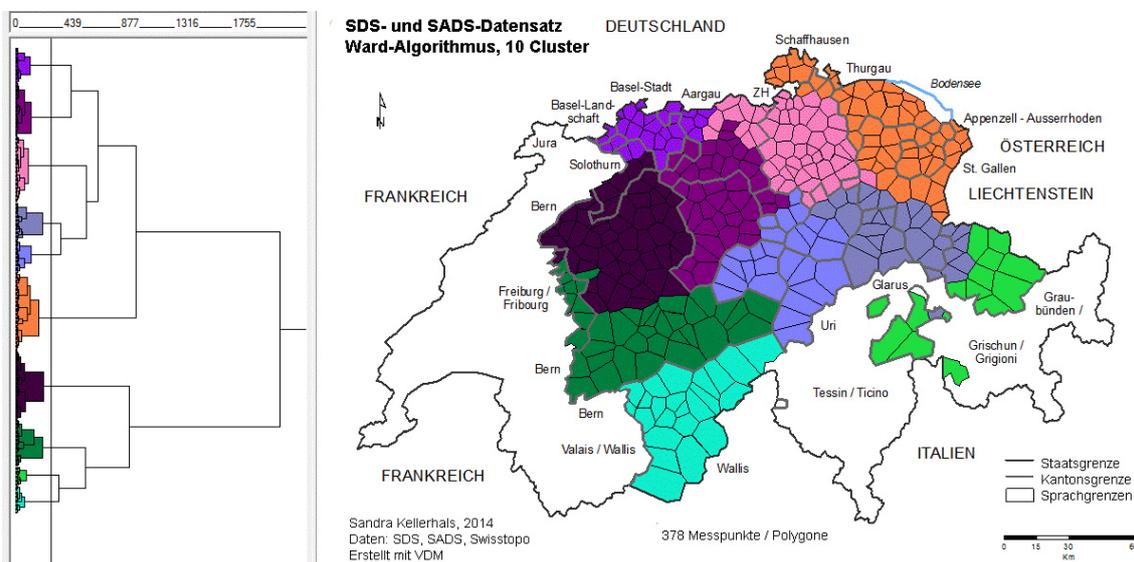


Abbildung 24 Clusterkarte aller SDS- und SADS-Daten, Ward-Algorithmus (10 Cluster) mit Dendrogramm (VDM)

Nachdem der Entscheid für den Ward-Algorithmus mit zehn Clustern für die Darstellung der linguistischen Distanzen aller SDS- und SADS-Daten gefallen ist, soll nun dieser Algorithmus ebenfalls mit zehn Clustern auf die Daten der einzelnen linguistischen Ebenen angewandt werden. Anschliessend kann eruiert werden, welcher prozentuale Anteil der Orte jeweils wieder in dasselbe (oder ein vergleichbares) Cluster eingeteilt wurde, wie bei der Analyse mit dem ganzen Datensatz. Dies gibt dann einen Hinweis auf die Frage, welche linguistische Ebene die Dialektlandschaft aller Daten am stärksten beeinflusst.

Abbildung 25 zeigt die Clusterkarten der vier verschiedenen linguistischen Ebenen, alle jeweils mit zehn Clustern und dem Ward-Algorithmus in VDM generiert. Über die Frage, welche der vier Karten der Karte aller SDS- und SADS-Daten in Abbildung 24 am ähnlichsten ist, gibt Tabelle 8 Auskunft. Sie zeigt für jede linguistische Ebene den Anteil der 378 Orte, der in der Clusteranalyse der entsprechenden linguistischen Ebene wieder demselben Cluster zugeteilt worden ist, wie bei der Analyse mit allen SDS- und SADS-Daten.

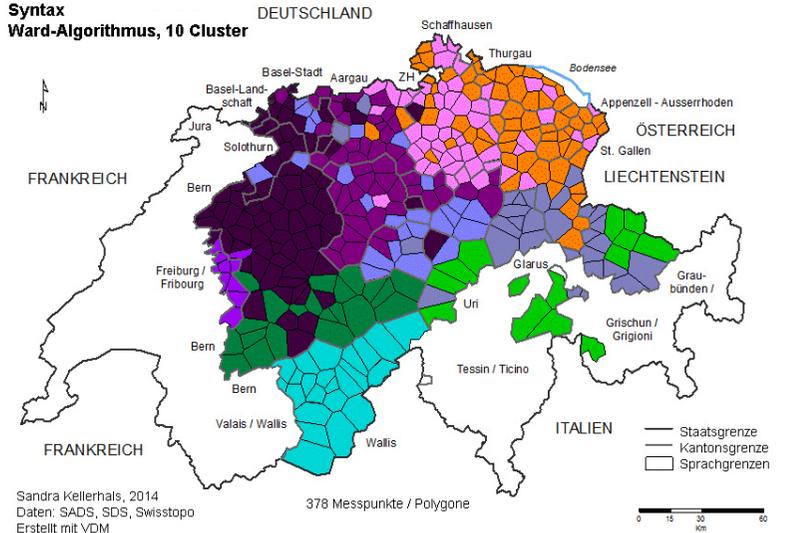
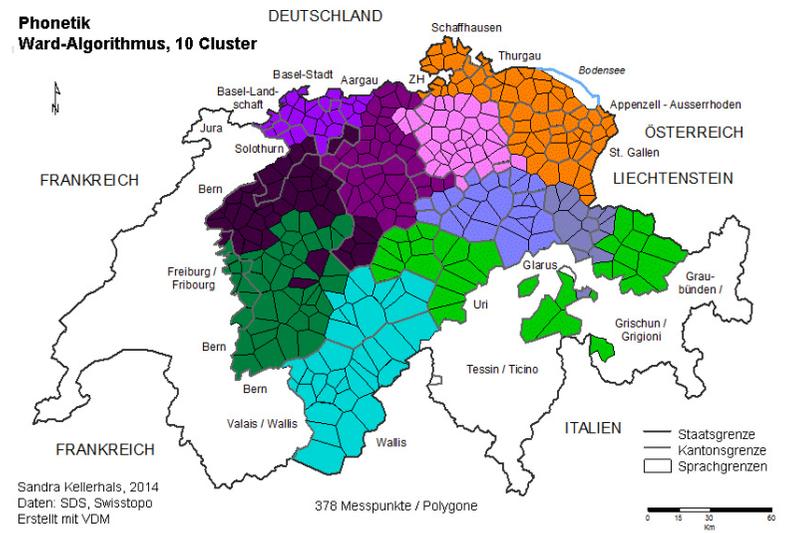
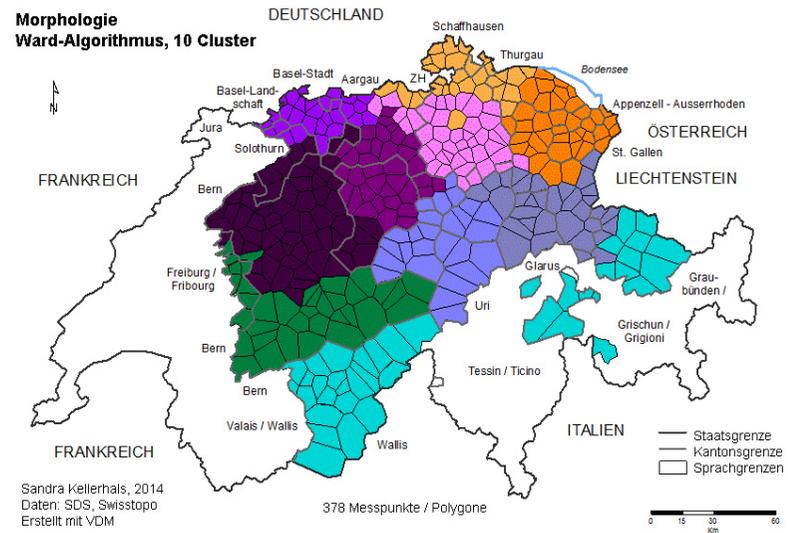
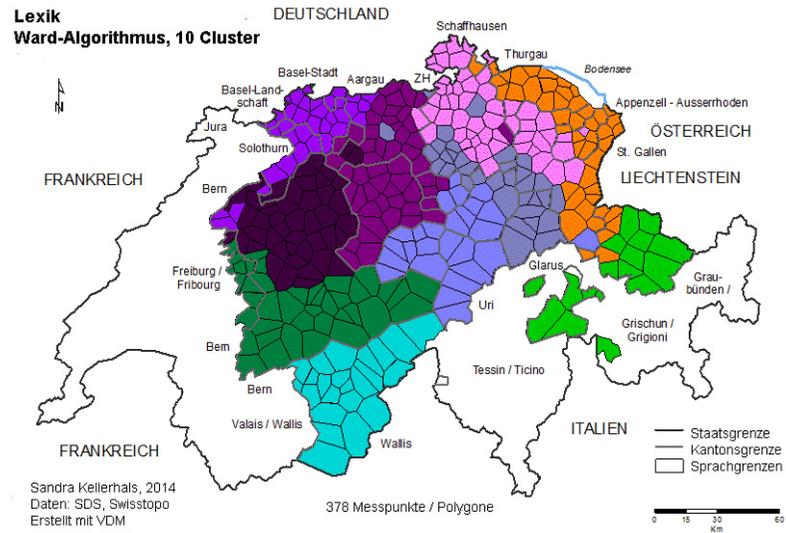


Abbildung 25 Clusteranalyse mit Ward-Algorithmus und zehn Clustern der einzelnen linguistischen Ebenen Lexik, Morphologie, Phonetik, Syntax

	Lexik	Morphologie	Phonetik	Syntax
<i>% Anteil Orte identisch klassiert wie in Clusteranalyse aller Daten</i>	78.3	83.9	76.2	66.9

Tabelle 8 Identisch klassierte Orte wie in Clusteranalyse aller Daten pro linguistische Ebene

Meist ist klar welches Cluster aus einer einzelnen linguistischen Ebene welchem Cluster aus der Karte aller Ebenen entspricht. Allerdings entsteht in der morphologischen und syntaktischen Ebene ein zusätzliches Cluster (Morphologie: hellorange Nordostschweiz, Syntax: hellviolett Freiburg) wodurch ein anderes verschwindet (Morphologie: kein hellgrünes Cluster in Graubünden, Syntax: kein hellviolett Cluster in Basel). In diesem Fall sind die Orte des neu entstandenen Clusters der einzelnen Ebene denjenigen im Cluster aus der Analyse aller Daten gegenübergestellt worden, das in der Analyse der einzelnen Ebenen verschwunden ist. Die Übereinstimmung für die in diesen Clustern enthaltenen Orte liegt dann bei 0 %. So kann diese Änderung in den Clustern erfasst werden.

Wie aus Tabelle 8 ersichtlich ist, stimmt die Einteilung der Orte in der morphologischen Ebene am ehesten mit derjenigen aller Daten überein. Die Lexik belegt den zweiten Rang, gefolgt von der Phonetik. Mit 66,9 % identisch klassierter Orte ist die Einteilung der Orte in der syntaktischen Ebene derjenigen aller Daten am wenigsten ähnlich. Das Kartenbild der Clusteranalyse auf der syntaktischen Ebene ist auch am stärksten fragmentiert. Die räumlich einheitlichsten Cluster sind im Wallis und in Freiburg zu finden (vgl. zwei dimensionale MDS-Analyse, Abbildung 21, rechts unten).

Ob die Zuteilung der Orte in die Cluster einer linguistischen Ebene ähnlich ist wie in diejenigen aller Daten, hängt von der Ähnlichkeit der zugrundeliegenden Distanzmatrizen ab. Die hohe Übereinstimmung der Cluster aus der morphologischen Ebene mit denjenigen aller Daten könnte daher kommen, dass aus der morphologischen Ebene auch die grösste Anzahl Variablen (120 Variablen) miteinflusst. Allerdings kann dieser Effekt nicht allein für das Resultat verantwortlich sein, da sonst die hohe Übereinstimmung der Lexik (36 Variablen) mit der Gesamtmatrix nicht erklärt werden könnte. Im nächsten Abschnitt werden die Korrelationskoeffizienten zwischen den Distanzmatrizen der linguistischen Ebenen und der Gesamtmatrix bestimmt, wobei unter anderem auch das Problem der unterschiedlichen Anzahl Variablen pro Ebene berücksichtigt wird, wie dies bereits im Abschnitt 4.5.1 beschrieben worden ist.

6.2 Numerischer Vergleich: Korrelationsanalyse

Vor der Berechnung der Korrelationen zwischen den linguistischen Distanzen der einzelnen Ebenen und den aggregierten Distanzen aller Ebenen sind die Daten der einzelnen Ebenen auf ihre räumliche Autokorrelation und ihre interne Konsistenz geprüft worden. Wie in Abschnitt 4.5.1 beschrieben eignen sich zur Bestimmung dieser Grössen die *Lokale Inkohärenz* (Nerbonne und Kleiweg 2007) und *Cronbachs Alpha* (Cronbach 1951). Beide Grössen konnten mittels Gabmap berechnet werden. Die Resultate sind in Tabelle 9 zusammengestellt.

Alle Ebenen weisen für Cronbachs Alpha Werte von über 0.7 auf. Die interne Konsistenz der Daten kann somit in jeder linguistischen Ebene als hinreichend angesehen werden (Nunnally

und Bernstein 1994). Am höchsten ist die interne Konsistenz der Daten auf der morphologischen Ebene, gefolgt von der Phonetik und der Lexik. Die syntaktischen Daten erreichen die geringste interne Konsistenz unter den linguistischen Ebenen und dies obwohl die Anzahl der Variablen, welche die Grösse auch massgeblich beeinflusst, in der lexikalischen Ebene weitaus und in der phonetischen Ebene etwas geringer ist. Auch bei der Lokalen Inkohärenz, welche die Verlässlichkeit der Daten anhand des lokalen Zusammenhangs zwischen der geographischen und der linguistischen Distanz misst, schneidet die Syntax am schlechtesten ab, da sie den höchsten Wert aufweist. Die Morphologie erreicht hier den tiefsten Wert, wodurch die Distanzmessungen auf dieser Ebene als die Verlässlichsten angesehen werden können. Die Phonetik und Lexik befinden sich auch hier in der Mitte. Schaut man sich die Werte der Lokalen Inkohärenz der drei Ebenen Lexik, Morphologie und Phonetik mit 565 Ortspunkten an, so sind diese dort sogar noch etwas höher. Nach diesem Mass nimmt die räumliche Autokorrelation und somit die Verlässlichkeit der Daten mit zunehmender Anzahl Ortspunkte also nicht zu. Die Werte der linguistischen Distanzen mit 565 Ortspunkten von Cronbachs Alpha sind jedoch jeweils um 0.01 höher als mit 378 Ortspunkten, was auf eine leicht höhere interne Konsistenz schliessen lässt. Besonders gut schneiden die linguistischen Distanzen aus den Daten aller Ebenen bei beiden Grössen ab, was auch mit der grossen Anzahl Variablen zusammenhängt.

Linguistische Ebene	Anzahl Ortspunkte	Anzahl Variablen	Lokale Inkohärenz	Cronbachs Alpha
<i>Lexik</i>	378	36	1.06	0.83
<i>Morphologie</i>	378	120	0.77	0.91
<i>Phonetik</i>	378	75	0.92	0.85
<i>Syntax</i>	378	112	3.41	0.77
<i>Alle Ebenen</i>	378	343	0.66	0.95
<i>Lexik</i>	565	36	1.29	0.84
<i>Morphologie</i>	565	120	0.80	0.92
<i>Phonetik</i>	565	75	1.04	0.86

Tabelle 9 Lokale Inkohärenz und Cronbachs Alpha der linguistischen Ebenen

Wie bereits in Abschnitt 4.5.1 erwähnt ist die vektorisierte Distanzmatrix, die aus allen Ebenen gebildet worden ist, vor der Korrelation mit den Distanzberechnungen der einzelnen Ebenen standardisiert worden. Dies macht einerseits Sinn, weil nicht von jeder linguistischen Ebene dieselbe Anzahl Variablen in die gesamte Distanzmatrix einfliesst und andererseits, weil sich die Varianzen bzw. die Standardabweichungen der Distanzen in den einzelnen Ebenen unterscheiden (s. Tabelle 10).

	Standardabweichung
<i>Lexik</i>	0.166
<i>Morphologie</i>	0.133
<i>Phonetik</i>	0.125
<i>Syntax</i>	0.083

Tabelle 10 Standardabweichung der Distanzen pro linguistische Ebene

Streuen die Distanzwerte stark um den Mittelwert, wie dies bei der Lexik der Fall ist, beeinflusst diese Variation diejenige der Gesamtmatrix stärker als im Falle einer geringen Streuung, wie bei der Syntax. Mit standardisierten Werten in der aggregierten Matrix aller

Ebenen sind sowohl dieser als auch der Einfluss der unterschiedlichen Anzahl Variablen pro Ebene ausgeglichen.

Tabelle 11 zeigt die *Pearson Produkt-Moment-Korrelationskoeffizienten* der linguistischen Ebenen untereinander und zu den standardisierten Distanzen aller Ebenen gemeinsam (Total). Das Bestimmtheitsmass (R^2) gibt darüber Auskunft, welcher Anteil der Varianz in der einen Ebene durch die Varianz in der anderen Ebene erklärt werden kann. Hohe Werte sind jeweils grün dargestellt (je höher umso dunkler), tiefere rot (je tiefer umso dunkler), und die Extremwerte sind fett hervorgehoben. Als erstes fällt auf, dass sich die Werte der Korrelation der Ebenen untereinander unterscheiden, je nachdem ob die Distanzmatrizen aus VDM oder Gabmap exportiert wurden. Da VDM immer nur eine Variante pro Ort und linguistische Variable akzeptiert, Gabmap aber mehrere, unterscheiden sich die Distanzmatrizen aus den beiden Systemen, was zu diesen Differenzen in den Korrelationskoeffizienten führt. Die Tendenzen sind jedoch in beiden Fällen dieselben. Die Morphologie korreliert am stärksten mit der Lexik und der Phonetik, wobei bei Gabmap und VDM jeweils die Extremwerte vertauscht sind. Die schwächsten Korrelationen betreffen in beiden Systemen die Syntax, welche in Gabmap am schwächsten mit der Lexik korreliert und in VDM am schwächsten mit der Phonetik. Die Reihenfolge der Stärken der Korrelation stimmt also in den von VDM und Gabmap berechneten Distanzen bis auf die Extremwerte überein. Um einen Hinweis darauf zu erhalten, welche Daten verlässlicher sind wurde die Lokale Inkohärenz auch noch für die linguistischen Distanzen aus VDM berechnet. Sie ergaben in allen vier Ebenen leicht höhere Werte als diejenigen von Gabmap: Für die Lexik 1.21, für die Morphologie 0.86, für die Phonetik 1.08 und für die Syntax 3.42. Deshalb können die linguistischen Distanzen aus Gabmap als etwas verlässlicher angesehen werden. Sie sind es ja auch, die alle vorkommenden Varianten in einem Ort berücksichtigen und nicht nur eine. Die Reihenfolge der Stärke der Korrelationen der einzelnen Ebenen mit den Distanzen aller Variablen stimmt in VDM und Gabmap überein. Die Morphologie korreliert demnach am stärksten mit der Gesamtmatrix, gefolgt von der Phonetik und der Lexik, die praktisch gleichauf sind. Die Syntax korreliert am schwächsten mit der Gesamtmatrix.

	Gabmap		VDM	
	Korrelation Pearson (r)	Bestimmtheitsmass (R^2)	Korrelation Pearson (r)	Bestimmtheitsmass (R^2)
<i>Lexik-Morphologie</i>	0.789	62 %	0.763	58 %
<i>Lexik-Phonetik</i>	0.767	59 %	0.751	56 %
<i>Lexik-Syntax</i>	0.516	27 %	0.585	34 %
<i>Morphologie-Phonetik</i>	0.784	61 %	0.772	60 %
<i>Morphologie-Syntax</i>	0.555	31 %	0.618	38 %
<i>Phonetik-Syntax</i>	0.530	28 %	0.571	33 %
<i>Lexik-Total</i>	0.891	79 %	0.890	79 %
<i>Morphologie-Total</i>	0.907	82 %	0.906	82 %
<i>Phonetik-Total</i>	0.894	80 %	0.889	79 %
<i>Syntax-Total</i>	0.754	57 %	0.797	64 %

Tabelle 11 Pearson Produkt-Moment-Korrelationskoeffizienten und Bestimmtheitsmasse der linguistischen Ebenen untereinander und mit allen Ebenen aggregiert (Gabmap und VDM). Grün: hohe Werte, rot: tiefe Werte, fett: Extremwerte.

Würden sich die Korrelationskoeffizienten verändern, wenn mehr Ortspunkte in die Analysen miteinbezogen würden? Um dafür einen Anhaltspunkt zu erhalten, wurden zusätzlich die Korrelationskoeffizienten der Distanzen der SDS-Ebenen Lexik, Morphologie und Phonetik mit 565 Ortspunkten berechnet. Wie Tabelle 12 zeigt, ist das Bestimmtheitsmass mit 565 Orten jeweils um ein bis zwei Prozentpunkte höher als mit 378 Orten. Das Bestimmtheitsmass zwischen Lexik und Morphologie und Lexik und Phonetik ist hier gleich gross und etwas höher als dasjenige zwischen Lexik und Phonetik. Dies stimmt jedoch weitgehend mit den Resultaten aus den Analysen mit 378 Orten überein.

	Korrelation Pearson (Gabmap, 565 Orte)	Bestimmtheitsmass (R^2) 565 Orte	Bestimmtheitsmass (R^2) 378 Orte
<i>Lexik-Morphologie</i>	0.795	63 %	62 %
<i>Lexik-Phonetik</i>	0.777	60 %	59 %
<i>Morphologie-Phonetik</i>	0.796	63 %	61 %

Tabelle 12 Vergleich der Bestimmtheitsmasse aus den Distanzen mit 565 Orten

Wie im Abschnitt 4.5.1 erwähnt sind alle Korrelationen auch mit dem *Spearman's Rangkorrelationskoeffizienten* berechnet worden. Die Korrelationskoeffizienten aus den beiden Berechnungen sind nahezu identisch. Die Werte für r unterscheiden sich nur geringfügig und die Reihenfolge der Stärke der Zusammenhänge ist identisch mit derjenigen aus Tabelle 11.

Aufgrund des hohen n ($\frac{378 \cdot 377}{2} = 71'253$) sind alle Korrelationen signifikant auf dem 0,1 % Signifikanzniveau. Die mit dem *Manteltest* ermittelten p-Werte gehen gegen 0. Auch sind die Unterschiede zwischen den einzelnen Korrelationskoeffizienten aus Tabelle 11 nach der *Fisher z-Transformation* signifikant auf dem 5 % Niveau. Die einzige Ausnahme bildet dabei der Unterschied zwischen Lexik-Total und Phonetik-Total der mit VDM generierten Daten. Diese beiden Korrelationskoeffizienten führen dann aber gerundet auch zu dem identischen Bestimmtheitsmass. Ebenfalls nicht signifikant auf dem 5 % Niveau ist der Unterschied zwischen den Korrelationskoeffizienten Lexik-Morphologie und Morphologie-Phonetik bei der Analyse mit 565 Ortspunkten.

6.3 Bestimmende Faktoren der Cluster

Für die Bestimmung der Variablen, welche die Dialektlandschaft des Schweizerdeutschen determinieren, wird wieder die Clustereinteilung mit dem Ward-Algorithmus und zehn Clustern aus der hierarchischen Clusteranalyse in Abschnitt 6.1.3 als Basis genommen. Diese Einteilung kommt dem Resultat aus der MDS-Analyse mit allen Daten des SDS und des SADS visuell am nächsten.

Im Anhang E findet sich die gesamte Tabelle mit zehn Variablen pro Cluster, welche jeweils zwischen den negativen z-Werten innerhalb und den positiven z-Werten zwischen den Clustern die grösste Differenz aufweisen (vgl. Abschnitt 4.5.2). Diejenigen Variablen also, welche dieses Cluster am stärksten determinieren. Da einige Variablen in mehreren Clustern unter den ersten zehn sind, besteht das Set aus 79 Variablen. Keine Variable ist in mehr als vier Clustern unter

den ersten zehn zu finden. Wie aus der ersten Spalte in Tabelle 13 ersichtlich ist, stammen die meisten der 79 Variablen aus der morphologischen Ebene. Dies ist auch der Fall, wenn diejenigen Variablen, welche mehrfach vorkommen, entsprechend auch mehrfach gezählt werden, so dass das Total 100 ergibt (s. zweite Spalte). Da jedoch die Morphologie auch über die meisten in die Analyse einflussenden Variablen verfügt, ist in der dritten Spalte der prozentuale Anteil der Variablen im 79er-Set an der gesamten Anzahl der Variablen in der entsprechenden Ebene berechnet. In der vierten Spalte werden die mehrfach vorkommenden Variablen entsprechend auch mehrfach gezählt. In beiden Fällen weist die Lexik den höchsten Anteil auf. Aus der lexikalischen Ebene fliessen insgesamt nur 36 Variablen in die Analysen ein und 14 (bzw. 23, wenn Mehrfache berücksichtigt werden) davon sind unter den ausgewählten Variablen zu finden. Den zweithöchsten Anteil verzeichnet die Phonetik, gefolgt von der Morphologie, während die Syntax mit nur zwei Variablen im Set nicht viel zur Determination der Cluster beiträgt.

Linguistische Ebene (Anzahl Variablen)	Absoluter Anteil an den Cluster bestimmenden Variablen (79)	Absoluter Anteil an den Cluster bestimmenden Variablen (100)	%-Anteil der Cluster best. Variablen (aus 79) an allen Variablen	%-Anteil der Cluster best. Variablen (aus 100, Mehrfache berücksichtigt) an allen Variablen
<i>Lexik (36)</i>	14 (17.7 %)	23	38.9	63.9
<i>Morphologie (120)</i>	35 (44.3 %)	40	29.2	33.3
<i>Phonetik (75)</i>	28 (35.4 %)	35	37.3	46.7
<i>Syntax (112)</i>	2 (2.5 %)	2	1.8	1.8

Tabelle 13 Cluster bestimmende Variablen und linguistische Ebenen

Tabelle 14 zeigt jeweils die eine Variable pro Cluster, welche die grösste Differenz zwischen dem negativen z-Wert innerhalb des Clusters und dem positiven z-Wert zwischen den Clustern aufweist. Es fällt auf, dass der negative Wert innerhalb des Clusters meist deutlicher negativ ist als dies beim positiven Wert in die umgekehrte Richtung der Fall ist. Die einzigen Ausnahmen bilden die Variablen «3154 Vög[el]chen: Subst. Diminutiv» und «3262 gehen lassen: Wortstellung» welche die Cluster im Wallis und in der Nordostschweiz determinieren. Bei diesen Variablen ist der Effekt, dass sich die Ausprägungen im Cluster überdurchschnittlich von denjenigen in den restlichen Clustern unterscheidet, deutlicher als die Einheitlichkeit der Ausprägung in den Orten innerhalb des Clusters. Sieht man sich die Verteilung der Varianten der Variable «3154 Vög[el]chen» an, fällt auf, dass die im Wallis vorherrschende Variante («-ulti»/ «-elti»/ «-iti»/ «ilti») wirklich in keiner anderen Region der Schweiz vorkommt. Jedoch ist die Variante auch innerhalb des Wallis perfekt homogen. Da die Homogenität innerhalb des Wallis jedoch unter allen Variablen häufiger ist, als das ausschliessliche Auftreten im Wallis, weicht der negative z-Wert weniger stark von Null ab als der Positive. Bei den anderen Variablen, wo der negative z-Wert innerhalb des Clusters stärker ausgeprägt ist, überwiegt der Effekt, dass die Varianten in den Orten des Clusters eine überdurchschnittlich einheitliche Ausprägung haben.

In dieser Auswahl aus zehn (bzw. neun verschiedenen) Variablen stammen vier (bzw. drei verschiedene) Variablen aus der Lexik, drei aus der Morphologie, zwei aus der Phonetik und eine aus der (Morpho-)Syntax.

Differenz	z innerhalb Cluster	z zwischen Clustern	Variable
-2.029546	-1.369854	0.659692	1134 t[ie]f: aobd. iu
-1.92811	-1.508274	0.419836	5112 Partikel: Grussformeln
-2.00428	-1.962975	0.041305	4176 nur
-2.652561	-0.53456	2.118001	3154 Vög[el]chen: Subst. Diminutiv
-2.049373	-1.733885	0.315488	6027 Zeitadverb: jeweils**
-1.408248	-1.287258	0.12099	3075 (an)fangen: 2./3.Pers. Sg. Ind. Präs.
-2.268966	-2.062026	0.20694	3047 haben: Pl. Ind. Präs.
-1.605122	-0.936313	0.668809	2138 ge[rn]: mhd. -rn
-2.409103	-2.303769	0.105334	6027 Zeitadverb: jeweil**s
-1.705452	-0.8012	0.904252	3262 gehen lassen: Wortstellung

Tabelle 14 Cluster bestimmende Variablen mit der grössten Differenz zwischen den z-Werten pro Cluster. (**: Variable 6027 weist in zwei verschiedenen Clustern die grösste Differenz zwischen den z-Werten auf)

Abbildung 26 vergleicht die Resultate aus den MDS-Analysen mit den zuvor ausgewählten 79 Variablen (zehn pro Cluster) und einem Set von 79 zufälligen Variablen mit der MDS-Analyse aller SDS- und SADS-Variablen. Auch die Resultate aus einem kleineren Set von 26 Variablen, das jeweils die drei Variablen pro Cluster enthält, welche die grösste Differenz zwischen den z-Werten aufweisen, werden gezeigt. Da auch hier einige Variablen mehrfach vorkommen, enthält das Set 26 Variablen und nicht 30. Auch diese Karte wird mit derjenigen eines Sets von 26 zufälligen Variablen verglichen. Die untersten beiden Karten vergleichen schliesslich MDS-Analysen, bei denen links nur noch für jedes Cluster die bezüglich der Differenz der z-Werte erstklassierte Variable in die Analyse miteinbezogen wird (neun Variablen) und rechts neun zufällige Variablen in die Analyse einfließen. Der Einfachheit halber werden fortan alle Sets von Variablen, welche mittels der Differenzen der z-Werte als Cluster-determinierend bestimmt worden sind, MIN-Sets genannt, diejenigen welche aus zufälligen Variablen bestehen, ZF-Sets. Die beiden Karten mit 79 Variablen sehen der Basiskarte mit allen Variablen (fortan ORG-Karte) beide relativ ähnlich. Einige Unterschiede sind jedoch dennoch erkennbar: In der ZF-Set-Karte ist die Dialektgrenze zwischen Aargau und Zürich im Gegensatz zur MIN-Set- und ORG-Karte nicht erkennbar. Auch die Grenze um die Region Basel fehlt auf dieser Karte und diejenige um die Zentralschweiz wirkt schwächer. Dafür ist die Grenze, welche das Berner Oberland und Freiburg vom Berner Unterland abgrenzt deutlicher zu sehen als auf der MIN-Set-Karte. Somit ist die ZF-Karte diesbezüglich näher an der ORG-Karte. Ungefähr die selben Unterschiede sind auch in den Karten mit 26 Variablen zu erkennen, wobei die Grenze zwischen dem Wallis und dem Berner Oberland in dieser ZF-Set-Karte zusätzlich schwächer ausgeprägt ist als in allen bisher beschriebenen Karten. Erst bei der Betrachtung der Karten mit nur neun Variablen sticht der Unterschied zwischen der MIN-Set- und der ZF-Set-Karte besonders ins Auge. Die ZF-Set-Karte ist stark fragmentiert. Das Wallis, Berner Oberland und Teile Graubündens heben sich jedoch, auch bei nur neun Variablen, von den anderen Gebieten ab. Auch eine schwache Grenze um die Zentralschweiz ist erkennbar und zusätzlich eine sich abhebende Region im Süden St.Gallens. Die MIN-Set-Karte zeigt sich deutlich näher an der ORG-Karte. Nur die Regionen Zürich und Bern sind stärker fragmentiert und die Grenze zwischen Zürich und der Ostschweiz fehlt. Dafür grenzt sich der Kanton Freiburg deutlich ab.

Zusammenfassend kann aus dem visuellen Vergleich der MDS-Analyse festgehalten werden, dass sich erst beim Set von neun Variablen deutliche Unterschiede zwischen den Karten des MIN- und des ZF-Sets ergeben, wobei das MIN-Set die ORG-Karte ziemlich genau abbildet.

Um jedoch auch einen quantitativen Anhaltspunkt über die Unterschiede zwischen den Karten der Sets bei einer sich verändernden Anzahl Variablen zu erhalten, ist in Abbildung 27 die hierarchische Clusteranalyse (Ward-Algorithmus mit zehn Clustern) auf die verschiedenen Sets angewandt worden. Neben dem visuellen Vergleich ist hier auch wieder ein numerischer möglich, indem für jede Variante der Anteil der Orte bestimmt wird, welche in der entsprechenden Karte wieder demselben Cluster zugeteilt worden ist wie in der ORG-Karte, bei welcher der Ward-Algorithmus mit zehn Clustern auf alle vorhandenen Variablen angewandt worden ist. Die Methode entspricht der in Abschnitt 6.1.3 beschriebenen. Der visuelle Vergleich zeigt ähnlich wie bei dem Vergleich der MDS-Analysen erst bei neun Variablen markante Unterschiede zwischen der MIN-Set und der ZF-Set-Karte. Auch der numerische Vergleich in Tabelle 15 bestätigt diese Erkenntnis. Der Anteil der identisch klassierten Orte wie in der ORG-Karte ist bei 79 Variablen in beiden Karten sehr hoch, leicht höher in der MIN-Set-Karte (93.1 %; ZF-Set-Karte 90.2 %). Nahezu identisch sind die Anteile der beiden Karten mit 26 Variablen (82.5 % bzw. 82.3 %). Erst bei dem Set von neun Variablen liegt der Anteil identisch klassierter Orte in der MIN-Set-Karte deutlich höher (77.8 %) als in der ZF-Set-Karte (47.9 %).

Anzahl Variablen im Set	% Anteil Orte identisch klassiert wie in Clusteranalyse aller Daten	
	Minimales Set	Zufallsset
79 (10 pro Cluster)	93.1	90.2
26 (3 pro Cluster)	82.5	82.3
9 (1 pro Cluster)	77.8	47.9

Tabelle 15 Vergleich der identisch klassierten Orte wie in Clusteranalyse aller Daten: Minimales Set – Zufallsset, mit 79, 26 und 9 Variablen

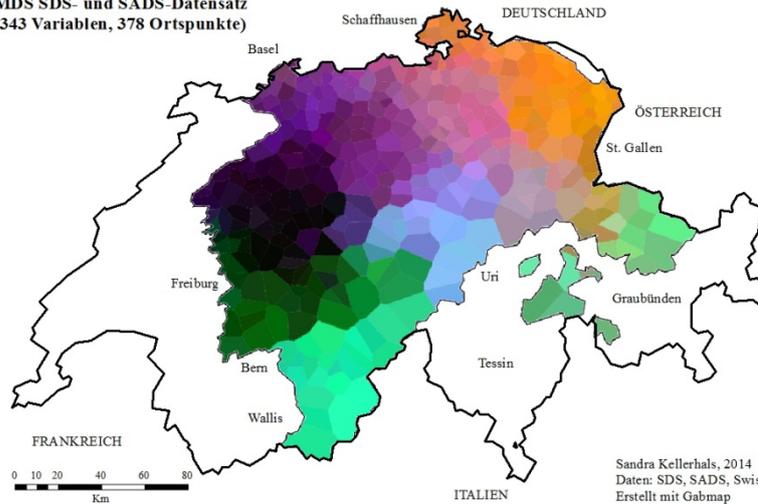
Mit der Analyse von bestimmenden Faktoren pro Cluster konnte ein minimales Set an Variablen gefunden werden, das mit nur einer Variable pro Cluster zu einer Übereinstimmung von fast 80 % mit der Clusteranalyse aller Variablen geführt hat. Je mehr clusterbestimmende Variablen ins Set einfließen, umso geringer wird der Unterschied zu einem Set mit zufälligen Variablen. Dies hängt auch damit zusammen, dass in das Set mit nur neun Variablen jeweils nur die Variablen mit der grössten Differenz zwischen dem negativen z-Wert innerhalb und dem positiven z-Wert zwischen den Clustern fließen. Bei jeder zusätzlichen Variable, die ins Set mit 26 bzw. 79 Variablen kommt, ist diese Differenz etwas kleiner.

Grundsätzlich ist der Effekt der Homogenität der Ausprägungen innerhalb eines Clusters in den gefundenen Variablen stärker als derjenige der Unterschiedlichkeit zu den Ausprägungen in den anderen Clustern. Perfekt trennende Variablen, bei denen diese beiden Effekte stark ausgeprägt sind, konnte nur eine, für das Cluster im Wallis, gefunden werden («3154 Vög[el]chen»). Auch bei Stichproben mit anderen zugrundeliegenden Clusteralgorithmen als dem Ward-Algorithmus wurden keine anderen perfekt trennenden Variablen gefunden.

Korreliert man die linguistischen Distanzen aus dem minimalen Set mit neun Variablen mit denjenigen aus allen Variablen, beträgt der Pearson-Produkt-Moment-Korrelationskoeffizient $r = 0.75$. Somit erklärt die Varianz dieses Sets 56 % der Varianz aller Variablen. Führt man

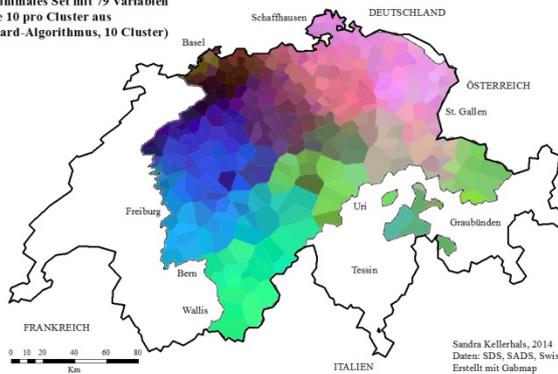
dieselbe Analyse mit dem Set aus neun zufällig ausgewählten Variablen durch, erhält man einen Korrelationskoeffizienten von $r = 0.65$ und daraus ein Bestimmtheitsmass von 42 %. Auch diese von der Clusteranalyse unabhängige Betrachtung bestätigt also, dass die Übereinstimmung mit den linguistischen Distanzen aller Daten mit dem MIN-Set höher ist als mit dem ZF-Set.

MDS SDS- und SADS-Datensatz
(343 Variablen, 378 Ortspunkte)



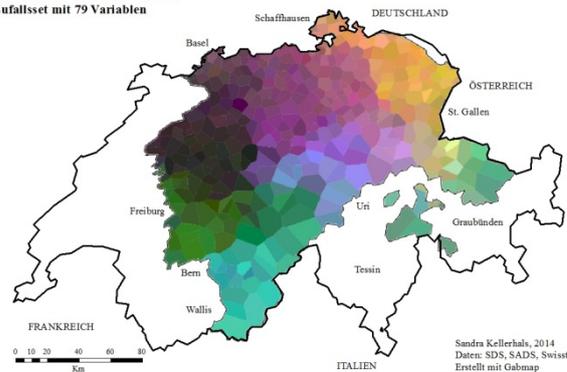
Sandra Kellerhals, 2014
Daten: SDS, SADS, Swisstopo
Erstellt mit Gabmap

MDS SDS- und SADS-Datensatz
Minimales Set mit 79 Variablen
(je 10 pro Cluster aus
Ward-Algorithmus, 10 Cluster)



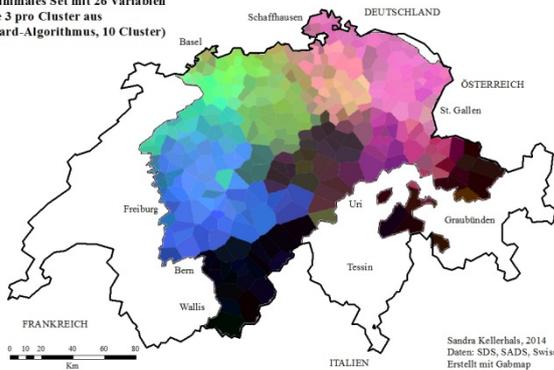
Sandra Kellerhals, 2014
Daten: SDS, SADS, Swisstopo
Erstellt mit Gabmap

MDS SDS- und SADS-Datensatz
Zufallsset mit 79 Variablen



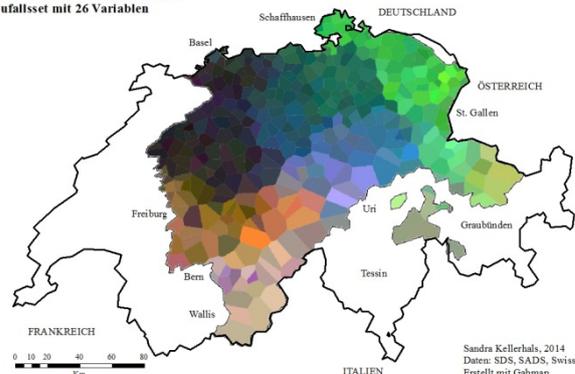
Sandra Kellerhals, 2014
Daten: SDS, SADS, Swisstopo
Erstellt mit Gabmap

MDS SDS- und SADS-Datensatz
Minimales Set mit 26 Variablen
(je 3 pro Cluster aus
Ward-Algorithmus, 10 Cluster)



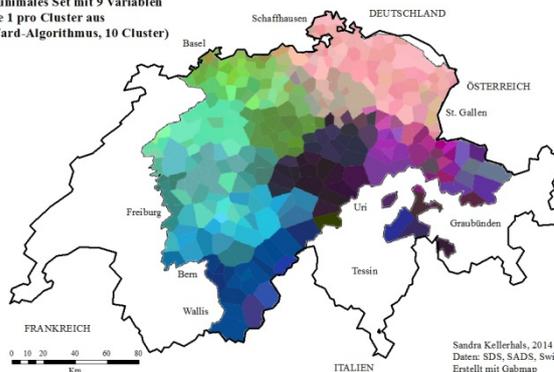
Sandra Kellerhals, 2014
Daten: SDS, SADS, Swisstopo
Erstellt mit Gabmap

MDS SDS- und SADS-Datensatz
Zufallsset mit 26 Variablen



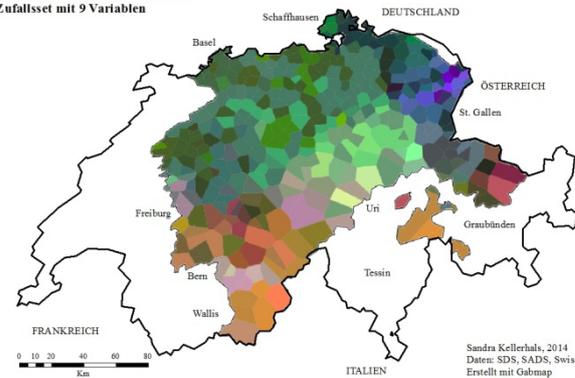
Sandra Kellerhals, 2014
Daten: SDS, SADS, Swisstopo
Erstellt mit Gabmap

MDS SDS- und SADS-Datensatz
Minimales Set mit 9 Variablen
(je 1 pro Cluster aus
Ward-Algorithmus, 10 Cluster)



Sandra Kellerhals, 2014
Daten: SDS, SADS, Swisstopo
Erstellt mit Gabmap

MDS SDS- und SADS-Datensatz
Zufallsset mit 9 Variablen



Sandra Kellerhals, 2014
Daten: SDS, SADS, Swisstopo
Erstellt mit Gabmap

Abbildung 26 Vergleich der MDS- Analysen mit Minimales Set – Zufallsset mit 79, 26 und 9 Variablen

6 Resultate

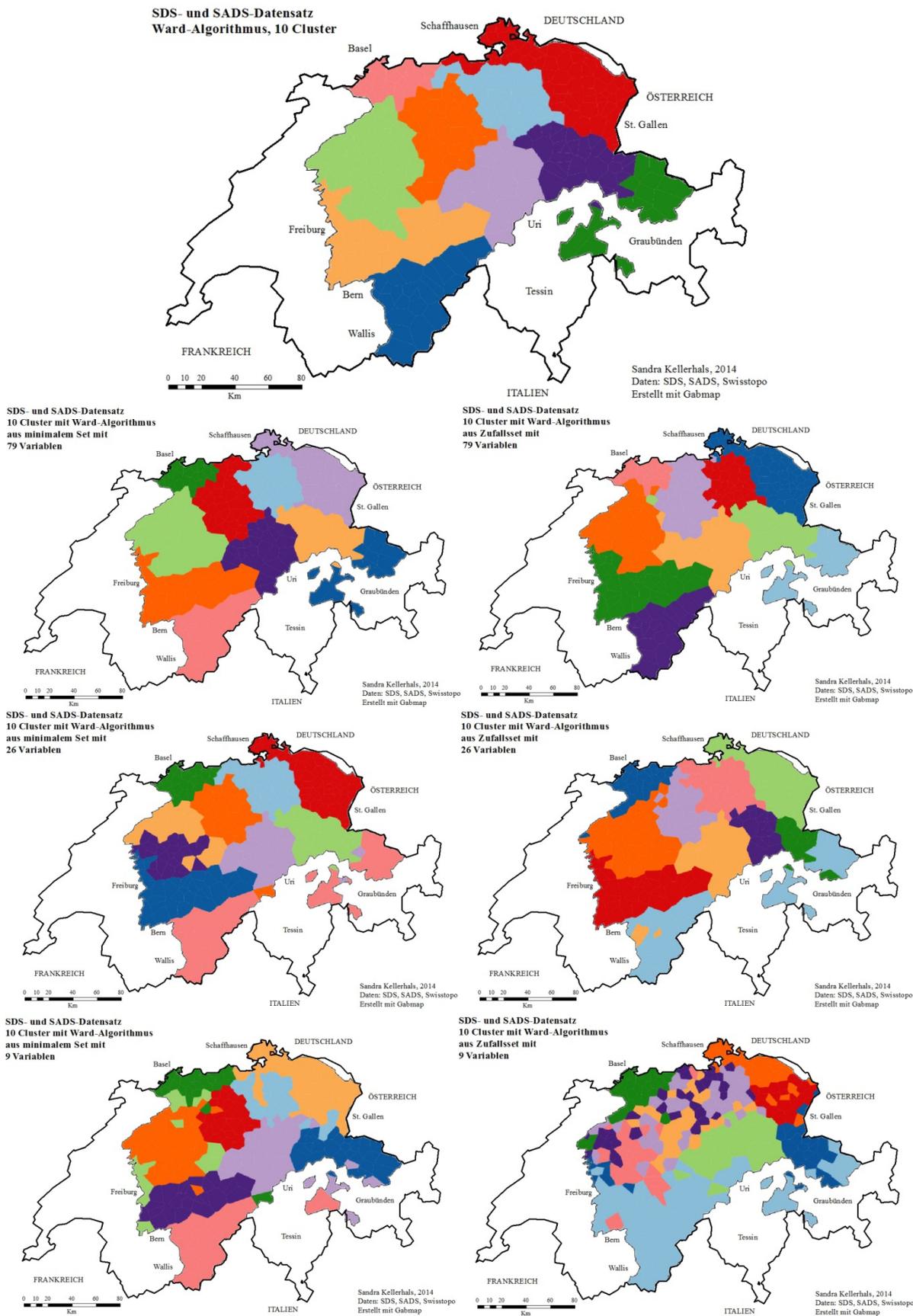


Abbildung 27 Vergleich der Clusteranalysen mit 79, 26 und 9 ausgewählten und zufälligen Variablen

7 Diskussion

In diesem Kapitel werden in einem ersten Teil die Methoden und Resultate dieser Arbeit reflektiert und wo möglich mit den Ergebnissen und Herangehensweisen anderer Studien verglichen. Als Vergleichsstudien in Bezug auf die visuellen Ergebnisse bzw. die gefundenen Dialektlandschaften dienen die einzigen bisher vorhandenen dialektometrischen Studien der Deutschschweiz: Goebel et al. (2013), Scherrer (2012) und Kelle (2001) (Abschnitt 7.1). Die gefundenen Resultate aus den numerischen Vergleichen können mit Spruit (2008) verglichen werden, der die Zusammenhänge zwischen den verschiedenen linguistischen Ebenen Lexik, Phonetik und Syntax anhand von niederländischen Dialekten untersucht hat (Abschnitt 7.2). Der letzte Teil des Kapitels diskutiert die Resultate inhaltlich und liefert Interpretationsmöglichkeiten (Abschnitt 7.3).

7.1 Visuelle Vergleiche: Parameterkarten, Multidimensionale Skalierung, hierarchische Clusteranalyse

Für die Darstellung der Dialektlandschaften, die auf aggregierten linguistischen Distanzen basieren, sind Kartierungsmethoden verwendet worden, die sich möglichst gut für Vergleiche zwischen den linguistischen Ebenen eignen. In Girnth (2010) ist eine allgemeine Zusammenstellung von Kartierungsmethoden für sprachliche Daten zu finden, die auch auf die Kartierung dialektometrischer Analysen eingeht. Darin sind neben den hier verwendeten Möglichkeiten auch solche erwähnt, die in dieser Arbeit nicht berücksichtigt worden sind. Es handelt sich dabei um Ähnlichkeits- und Isoglossen-Karten. In Ähnlichkeitskarten werden Dialektähnlichkeiten von einem bestimmten Ort aus dargestellt (Goebel 1982a, 2006b, 2010). Diese Art der Darstellung macht vor allem Sinn, wenn die Dialekte dynamisch untersucht werden können. In VDM ist es möglich, wiederholt auf jeweils einen anderen Ort zu klicken und dabei die sich teilweise kontinuierlich und teilweise abrupt verändernden Dialektlandschaften zu beobachten. Diese Methode ist in der vorliegenden Arbeit nicht verwendet worden, da die Karten statisch präsentiert werden und da die linguistische Stellung eines bestimmten Ortsdialekts und dessen Umfeld nicht im Vordergrund steht. Isoglossen-Karten hätten sich ebenfalls zur Darstellung von linguistischen Distanzen zwischen den Orten geeignet (Goebel 2010; Goebel et al. 2013). Je grösser die linguistische Distanz zwischen zwei Orten, umso dicker und markanter wird die Linie der Polygonseite gestaltet, welche die beiden Orte trennt. In dieser Arbeit wurde der Fokus jedoch mehr auf die flächenhaften Darstellungen von Dialektregionen gelegt und daher auf diese Darstellungsweise verzichtet.

Parameterkarten sind vor allem von den Salzburger Dialektologen verwendet worden um Charakteristiken von Verteilungen linguistischer Distanzen zu untersuchen. Es existieren zahlreiche Studien, wo diese Methodik die Eigenschaften romanischer (Goebel 1995, 1984, 2006a, 2006b), insbesondere französischer (Goebel 2010, 1982a, 2006b) und italienischer (Goebel 1982a), aber auch germanischer Dialekte (Goebel et al. 2013) aufzeigt. Meist wird in diesen Studien, wie auch in der hier vorliegenden, der Parameter der Schiefe (Goebel 2006a, 1982a, 2006b, 2010; Goebel et al. 2013) verwendet, der den Sprachausgleich in einer bestimmten Region misst. Auch Dialektkerne mittels Maxima der Ähnlichkeitsberechnungen zu suchen, ist eine beliebte Methode (Goebel 2006b, 2010; Goebel et al. 2013), welche in der vorliegenden Studie analog anhand von Minima der Distanzen angewandt worden ist.

In Goebel et al. (2013) sind die Schiefe-Werte der Häufigkeitsverteilung der Ähnlichkeiten mit allen SDS-Daten gemeinsam visualisiert worden, basierend auf den von Scherrer (2012) digitalisierten Karten. In Abbildung 28 und 29 sind die Schiefe-Werte der Ähnlichkeiten aller in der vorliegenden Arbeit verwendeten SDS-Daten dargestellt. Dies ermöglicht, einen Eindruck zu erhalten, welchen Einfluss die Anpassungen an den Daten für diese Arbeit im Vergleich zu denjenigen von Scherrer (2012) gehabt hat. Beide Karten sind jeweils mit demselben Klassenbildungsalgorithmus und derselben Anzahl Klassen dargestellt. Da sie auf Ähnlichkeiten und nicht auf Distanzen basieren, wie die in Abschnitt 6.1.1 präsentierten Schiefe-Werte, sind die Schiefe-Werte hier vorwiegend positiv und der Sprachausgleich der rot gefärbten Orte ist gering. Die Unterschiede der den beiden Karten (Abbildung 28 und Abbildung 29) zugrundeliegenden Daten sind im Abschnitt 4.1.1 beschrieben. Zusätzlich variiert die Vorgehensweise bei der Auswahl einer Variante pro Variable und Ort für VDM. In der Arbeit von Goebel et al. (2013) ist jeweils zufällig eine Variante ausgewählt worden, während in der vorliegenden Arbeit, wie im Abschnitt 4.2.3 beschrieben, die jeweils auf der einzelnen Karte links eingezeichnete Variante bevorzugt worden ist, da diese mit einer gewissen Berechtigung als die Wichtigere betrachtet werden kann (Trüb 2003).

Die Abbildungen zeigen, dass die Klassen der Schiefe-Werte sowohl was deren geographische Verteilung als auch was die Grenzwerte der Klassen betrifft, sehr ähnlich sind. Erst bei genauerem Hinschauen erkennt man, dass die Werte in vereinzelt Orten, vor allem im Kanton Luzern und an der Grenze Bern-Wallis nicht in dieselbe Klasse zu liegen kommen. Die Verfeinerung der Daten in der vorliegenden Arbeit hat also, gemessen an diesem Beispiel der Schiefen, erwartungsgemäss zu keinen grundlegenden Änderungen in der Dialektlandschaft, sondern nur zu kleineren Anpassungen geführt. Wie aber wirken sich die Verfeinerungen auf die Resultate der MDS- und hierarchischen Clusteranalysen aus? Die folgenden Abschnitte geben Aufschluss über diese Fragen.

In der Arbeit von Scherrer (2012) ist aus 196 SDS-Variablen (27 lexikalische, 110 morphologische und 59 phonetische Variablen) eine *MDS-Karte* generiert worden. Abbildung 30 zeigt oben diese Karte und unten diejenige mit den dieser Arbeit zugrundeliegenden 235 SDS-Variablen. In beiden Karten sind praktisch dieselben Dialektregionen erkennbar. Einzig der Süden St. Gallens bildet in Scherrers‘ Karte eine eigene Dialektregion, welche sich von derjenigen Glarus‘ unterscheidet, während diese beiden Regionen in der unteren Karte zu einer Dialektregion verschmelzen. Dies bestätigt, dass die MDS zu Recht als stabile Methode angesehen wird, deren Resultat sich bei einer leichten Anpassung der Daten nicht grundlegend ändert.



Figure 7.8.: Map resulting from multidimensional scaling applied to the SDS data set.

MDS SDS-Datensatz
(235 Variablen, 565 Orte)

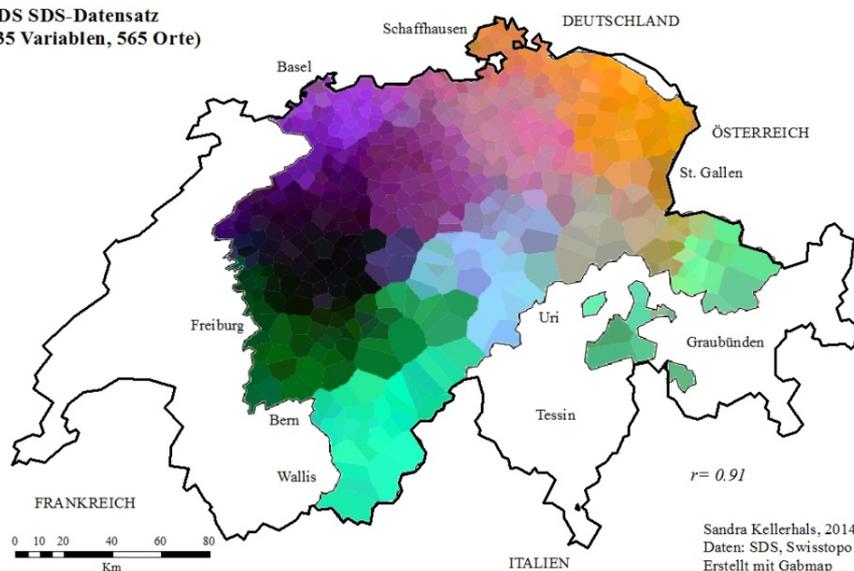


Abbildung 30 Vergleich der MDS-Analysen angewandt auf den SDS- Datensatz, oben mit 196 Variablen von Scherrer (2012, S. 263) unten mit den 235 Variablen aus dieser Arbeit

Aufgrund der MDS-Analysen mit allen Daten und den einzelnen linguistischen Ebenen sind im Abschnitt 6.1.2 Dialektregionen erkannt und definiert worden. Tabelle 16 zeigt nun inwiefern

diese gefundenen Regionen auch mit traditionellen dialektologischen Grenzen übereinstimmen, die aus Hotzenköcherle et al. (1984) bekannt sind (Scherrer 2012).

Traditionelle Grenzen nach Hotzenköcherle	MDS Lexik	MDS Morphologie	MDS Phonetik	MDS Syntax	MDS SDS- und SADS-Datensatz
<i>um Region Basel</i>	Ja, deutlich	Teilweise	Ja	Nein	Teilweise
<i>Zürich / Nordostschweiz</i>	Teilweise	Ja, schwach	Ja, deutlich	Nein	Ja
<i>Aargau / Zürich</i>	Ja, schwach	Teilweise	Teilweise	Ja, schwach	Teilweise
<i>Bern / Luzern</i>	Ja, deutlich	Ja, schwach	Ja, schwach	Ja, schwach	Ja
<i>Berner Oberland, Freiburg / Bern Nord</i>	Ja, deutlich	Ja, deutlich	Teilweise	Teilweise	Ja
<i>um Zentralschweiz</i>	Ja, deutlich	Ja, deutlich	Teilweise	Nein	Ja
<i>Wallis / Bern</i>	Ja, deutlich	Ja, deutlich	Teilweise	Ja	Ja

Tabelle 16 Übersicht über traditionelle dialektologische Dialektgrenzen aus Scherrer (2012) nach Hotzenköcherle et al. (1984) und die Übereinstimmungen mit den MDS-Analysen.

Wie aus Tabelle 16 ersichtlich ist, können die traditionellen dialektologischen Grenzen in allen Ebenen, mit Ausnahme der syntaktischen, zumindest teilweise oder schwach wieder erkannt werden. Auch die Übereinstimmung der MDS-Karte aller Daten mit den traditionellen Grenzen ist relativ hoch. Die stabilste Grenze stellt diejenige zwischen dem Wallis und Bern dar. Tendenziell sind die Grenzen in der Süd- und Westschweiz deutlicher wieder erkannt worden als diejenigen in der Nord- und Ostschweiz. Dass die traditionelle Dialektologie Grenzen definiert hat, die nicht mit denjenigen der syntaktischen Ebene übereinstimmen, erstaunt nicht. Den in traditionellen Arbeiten beschriebenen Dialektgrenzen liegen keine syntaktischen Daten zugrunde. Zudem haben auch unsere weiteren Analysen gezeigt, dass sich die Dialektlandschaften syntaktischer Phänomene von denjenigen der anderen Ebenen am stärksten unterscheiden.

Wie bereits im Abschnitt 5.2 erwähnt, gilt die *hierarchische Clusteranalyse* als sensibler auf Veränderungen in den Daten, als die Multidimensionale Skalierung, wobei der Ward-Algorithmus als eher stabil hervorgetreten ist. Ein Vergleich mit den Ergebnissen aus Goebel et al. (2013) hat dies bestätigt. Eine aus allen SDS-Daten mit dem Ward-Algorithmus und zehn Clustern generierte Karte aus Goebel et al. (2013) unterscheidet sich nur geringfügig von einer vergleichbaren Karte, die aus allen SDS-Daten der vorliegenden Arbeit generiert worden ist. Lediglich ein Cluster in der Region Luzern hat sich, zugunsten des westlichen und östlichen Clusters, gegenüber demjenigen aus unserer Studie etwas verkleinert.

Wie aber unterscheiden sich die gefundenen Cluster aus verschiedenen Studien, denen jeweils etwas andere Daten zugrunde liegen bei Verwendung eines anderen Algorithmus? Kelle (2001) verwendet in seiner Studie mit 170 Variablen aus der phonetischen und morphologischen Ebene den «Complete-Linkage»-Algorithmus und bildet damit bis zu sechs Cluster. Dabei findet er neben dem Wallis als Dialektregion eine viergeteilte Deutschschweiz, in der sich eine Nordost- und Südost- sowie eine Nordwest- und Südwest-Gruppe abzeichnen. Das sechste Cluster bilden die abgetrennten südlichen Orte, welche im Ortsnetz der vorliegenden Arbeit nicht vertreten

sind. Auch die als Kultur- und Dialektgrenze geltende «Brünig-Napf-Reuss-Linie», welche sich von der Aaremündung im Nordosten des Kantons Aargau bis zum Furka-Pass zieht (Kelle 2001) ist zu erkennen (s. Abbildung 31).

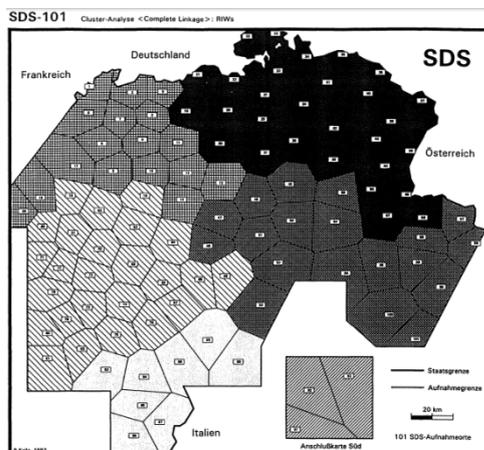


Abbildung 31 Clusterkarte mit SDS-Daten (170 Variablen, 101 Ortspunkte) aus Kelle (2001, S. 28) «Complete Linkage»-Algorithmus (6 Cluster)

Auch Scherrer (2012) hat zu Vergleichszwecken mit Kelle (2001), seine 196 Variablen aus dem SDS mit dem «Complete Linkage»-Algorithmus in sechs Cluster eingeteilt (Abbildung 32).

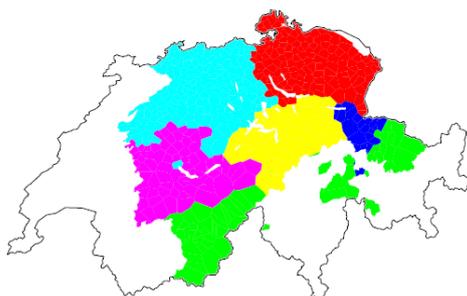


Figure 7.6.: Six clusters obtained with the Complete Linkage algorithm on the SDS data set.

Abbildung 32 Clusterkarte mit SDS-Daten (196 Variablen, 565 Ortspunkte) aus Scherrer (2012, S. 260), «Complete Linkage»-Algorithmus (6 Cluster)

In Scherrers' Karte beinhaltet ein einziges Cluster die Walliser und Bündner Dialekte. Dafür ist im Südosten ein zusätzliches Cluster entstanden. Die ansonsten viergeteilte Deutschschweiz im Norden der Alpen ist auch hier zu erkennen, jedoch sind die Cluster etwas anders angeordnet als bei Kelle (2001). Mit dem verfeinerten SDS-Datensatz (235 Variablen) aus der vorliegenden Arbeit entsteht nochmals ein etwas anderes Bild. Die durch VDM generierte Karte mit jeweils einer Variante pro Ort (Abbildung 33, links) sieht derjenigen von Scherrer sehr ähnlich. Das Cluster in der Nordostschweiz ist bei Scherrer etwas weniger gegen Westen ausgedehnt als dies in der linken Karte der Abbildung 33 und in Kelles Karte der Fall ist. Generiert man die Karte allerdings mit Gabmap, wo mehrere Varianten pro Ort enthalten sein können, verändert sich das Bild markant (Abbildung 33, rechts).

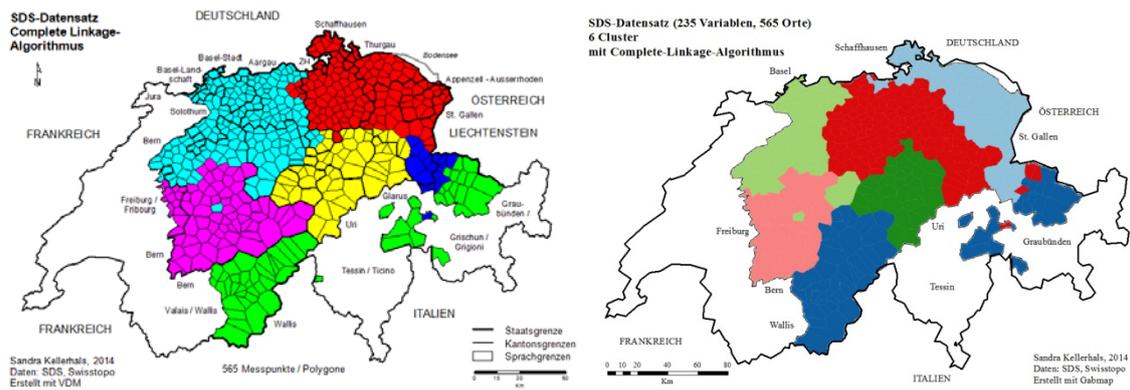


Abbildung 33 Clusterkarte mit SDS-Daten (235 Variablen, 565 Ortspunkte), «Complete Linkage»-Algorithmus (6 Cluster), links mit VDM (1 Variante pro Ort), rechts mit Gabmap (mehrere Varianten pro Ort)

Der «Complete Linkage»-Algorithmus vergleicht alle Orts-Paare in zwei Clustern und wählt dasjenige mit der grössten Distanz zueinander aus. Dieser Vorgang wird mit allen Clusterpaaren durchgeführt und am Ende die beiden Cluster zusammen geführt, deren ausgewähltes Orts-Paar die geringste «grösste Distanz» hat. Durch die unterschiedlichen Distanzmatrizen, welche aus VDM und Gabmap durch die Tatsache resultieren, dass in der einen Anwendung Mehrfachantworten berücksichtigt werden können und in der anderen nicht, können diese Orts-Paare mit der grössten Distanz variieren, wodurch andere Cluster gebildet werden. Bei diesem Algorithmus können also geringfügige Änderungen im Datensatz zu grossen Änderungen in der Clusterbildung führen. Wie Abbildung 34 zeigt, scheint vor allem die Variation in der Phonetik für das veränderte Bild verantwortlich zu sein (rechts). Sieht man sich die Clustereinteilung der Morphologie an (links), ähnelt diese dem Bild aus Kelle (2001) eher, obwohl die Schweiz nördlich der Alpen fünfgeteilt ist und sich die Walliser und Bündner Dialekte wieder in einem Cluster befinden. Die Dialektunterschiede entlang der erwähnten Brünig-Napf-Reuss Linie (rot in Abbildung 34) treten jedoch in beiden Ebenen zutage.

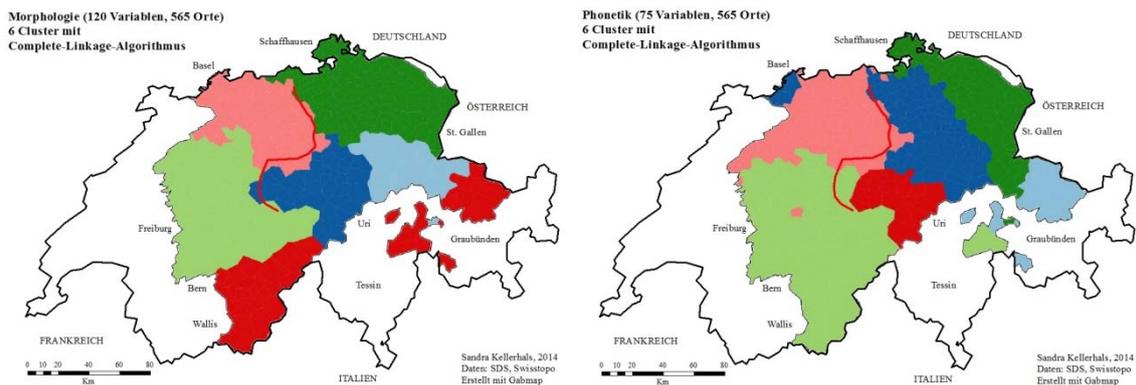


Abbildung 34 Clusterkarten Morphologie (links) und Phonetik (rechts), «Complete Linkage»-Algorithmus (6 Cluster) aus Gabmap mit Brünig-Napf-Reuss-Linie (rot)

Scherrer (2012) stellt fest, dass der «Complete Linkage»-Algorithmus gewöhnlich nicht mehr für dialektometrische Studien verwendet wird. Er wendet daher den «WPGMA»-Algorithmus auf sein SDS-Datenset an und bildet zehn Cluster (Abbildung 35, links). Wendet man denselben Algorithmus mit derselben Clusteranzahl auf den verfeinerten SDS-Datensatz der vorliegenden Arbeit an, sieht die Dialektlandschaft recht unterschiedlich aus. Nur die Cluster in den

Randregionen der Nordwest, Süd und (Nord)ostschweiz sind in den beiden Karten vergleichbar. Auch dieser Algorithmus reagiert also mit grossen Änderungen in der Clusterbildung bei eher geringfügigen Änderungen im Datensatz.

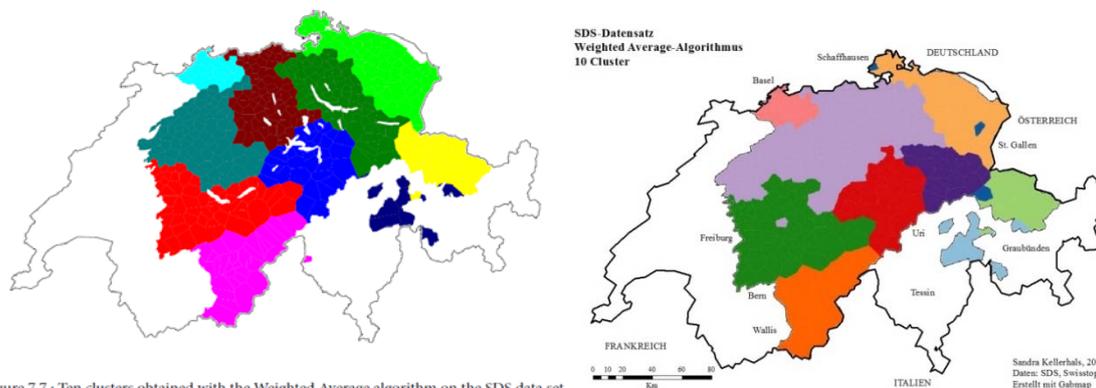


Figure 7.7.: Ten clusters obtained with the Weighted Average algorithm on the SDS data set.

Abbildung 35 Clusterkarte mit SDS-Daten links aus Scherrer (2012, S. 261) mit 196 Variablen, rechts mit verfeinertem Datensatz (235 Variablen), «WPGMA»-Algorithmus (10 Cluster)

Es konnte gezeigt werden, dass die hierarchische Clusteranalyse eine eher instabile Methode ist, die je nach Anpassungen im Datensatz und Algorithmus andere Dialektregionen zum Vorschein bringt. Weiter hat sich gezeigt, dass der Ward-Algorithmus am wenigsten stark auf Änderungen im Datensatz reagiert hat. Diese Erkenntnisse machen deutlich, wie wichtig es ist, die Resultate der hierarchischen Clusteranalyse nicht isoliert, sondern zusammen mit den Ergebnissen aus der MDS-Analyse zu betrachten und mithilfe eines «Fuzzy-Clusterings» zu validieren.

Ein solches «Fuzzy Clustering» hat Scherrer (2012) mit den Daten des SADS durchgeführt. Er hat dabei jedoch nicht wie in der vorliegenden Studie jeweils nur die Antworten berücksichtigt, welche die Gewährspersonen als natürlichste genannt haben, sondern alle akzeptierten Antworten und auch alle 118 Fragen in die Analyse miteinbezogen.

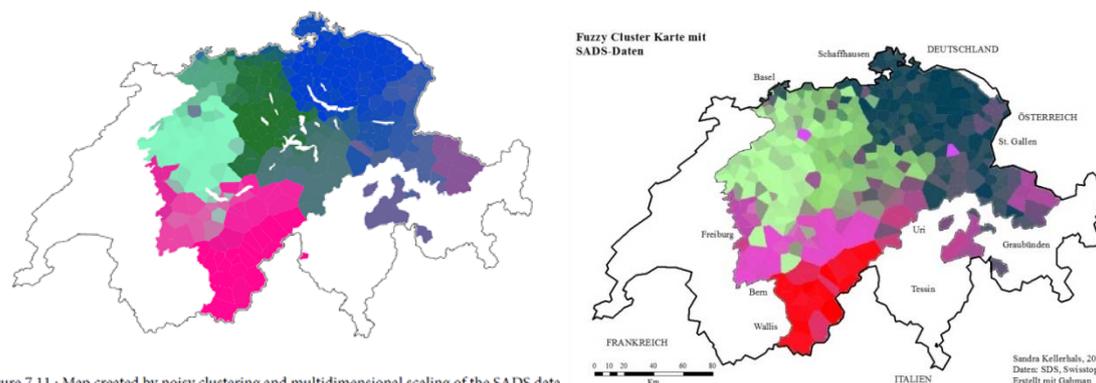


Figure 7.11.: Map created by noisy clustering and multidimensional scaling of the SADS data.

Abbildung 36 «Fuzzy Cluster»-Karten mit «UPGMA» und «WPGMA»-Algorithmus, rechts aus Scherrer (2012, S. 267) mit 118 Variablen und allen akzeptierten Antworten, links mit 108 Variablen und allen natürlichen Antworten.

Während in der Karte von Scherrer (2012) die Ost-West-Grenzen zwischen Bern, Luzern, Aargau und Zürich deutlich hervortreten, ist in der rechten Karte in Abbildung 36 lediglich die Grenze zwischen Aargau und Zürich erkennbar. Dafür ist in der rechten Karte die Dialektgrenze

zwischen Bern und Wallis deutlicher sichtbar. Auch erscheinen Freiburg und das Berner Oberland als eine abgetrennte Dialektregion. In der Ostschweiz sind die beiden Karten fast identisch, wobei die rechte Karte insgesamt stärker fragmentiert erscheint. Dieser Vergleich zeigt wieder den Einfluss der Datenauswahl auf das Ergebnis. In Anlehnung an die Erhebungsmethode des SDS und um einen möglichst guten Vergleich mit diesen Daten zu ermöglichen, sind in dieser Studie die natürlichsten Antworten des SADS-Datensatzes verwendet worden. Eine Idee für eine zukünftige Studie wäre der Einbezug aller akzeptierten Varianten des SADS in den Vergleich mit dem SDS. Für einen Vergleich zwischen dem SADS-Datensatz, in dem alle akzeptierten und demjenigen, indem nur die natürlichen Antworten berücksichtigt werden, sei auf Scherrer (2012) verwiesen. Er zeigt, dass Karten, die auf den natürlichen Antworten basieren grundsätzlich fragmentierter sind als solche, die auf allen akzeptierten Varianten beruhen. Als Grund dafür vermutet er, dass es für die Gewährspersonen schwierig ist eine Variante auszuwählen und diese Befragungstechnik daher eher zu Ausreißern neigt.

7.2 Numerischer Vergleich: Korrelationsanalyse

Spruit (2008) hat für die niederländischen Dialekte ebenfalls die Pearson-Produkt-Moment-Korrelationen zwischen den linguistischen Ebenen untersucht und die Syntax miteinbezogen. Im Gegensatz zur vorliegenden Arbeit beinhaltet bei Spruit (2008) die «Phonetik» sowohl die phonetisch/ phonologische als auch die morphologische Ebene. Aus der Lexik und der Phonetik sind 125 Karten in die Analyse miteinbezogen worden und aus der Syntax 145. Das untersuchte Gebiet verfügt über 70 Ortspunkte. Ein weiterer Unterschied zur vorliegenden Studie besteht darin, dass die linguistischen Distanzen auf phonetischer Ebene mittels Levenshtein-Distanz berechnet worden sind, während die Berechnung auf lexikalischer und syntaktischer Ebene mittels Hamming-Distanz und dem Gewichteten Identitätswert erfolgt ist. In der vorliegenden Studie sind alle Distanzen mittels Hamming-Distanz bzw. Relativem Identitätswert ermittelt worden. Tabelle 17 zeigt den direkten Vergleich zwischen den Bestimmtheitsmassen aus Spruit (2008) und der vorliegenden Studie.

	R² vorliegende Studie (Gabmap)	R² (Spruit 2008)
<i>Phonetik-Lexik</i>	59 %	38 %
<i>Lexik-Syntax</i>	27 %	25 %
<i>Syntax-Phonetik</i>	28 %	42 %

Tabelle 17 Vergleich mit den Bestimmtheitsmassen aus Spruit (2008)

In beiden Studien sind sich die Lexik und die Syntax am wenigsten ähnlich. Während in Spruit (2008) die Phonetik sowohl mit der Lexik als auch mit der Syntax relativ stark korreliert, ist die Syntax in der vorliegenden Studie mit den beiden anderen Ebenen relativ schwach verbunden. Im Vergleich zu der Studie von Spruit (2008) ist die Korrelation zwischen der Phonetik und der Lexik in der vorliegenden Studie relativ hoch. Es wäre interessant herauszufinden, wie diese unterschiedlichen Zusammenhänge zwischen den Ebenen zustande kommen. Denkbar wäre,

dass die unterschiedlichen Methoden (Levenshtein und GIW vs. RIW) oder die unterschiedlichen Definitionen der «phonetischen» Ebene eine Rolle spielen. Auch könnten die verschiedenen Werte der internen Konsistenz der Daten einen Einfluss auf die Resultate haben. Spruit (2008) erwähnt, dass seine Resultate aus den Daten der lexikalischen Ebene aufgrund des niedrigen Cronbachs Alpha-Wertes (0.75) weniger verlässlicher sein könnten. In unserer Studie sind die Cronbachs-Alpha-Werte auf der syntaktischen Ebene am niedrigsten (0.77). Es könnte aber auch an der unterschiedlichen Sprache (Niederländisch vs. Schweizerdeutsch), der unterschiedlichen Anzahl einflussender Variablen oder allfälligen unterschiedlichen Vorgehensweisen bei der Klassierung und Auswahl der Anzahl berücksichtigter Varianten pro Variable liegen. Auch wäre ein weiterer Schritt, den Einfluss der Geographie auf die linguistischen Distanzen mit zu berücksichtigen, wie dies Spruit (2008) tut. Andere Studien, welche die Syntax in die dialektometrische Analyse miteinbeziehen, z.B. Uiboed et al. (2013), vergleichen diese nicht direkt mit den anderen linguistischen Ebenen und lassen sich daher nicht direkt mit dieser Studie vergleichen.

7.3 Bestimmende Faktoren der Cluster

Die in dieser Studie verwendete Methode zur Bestimmung der determinierenden Faktoren der Cluster aus Cöltekin et al. (2012) ist aufgrund ihrer Einfachheit gewählt worden und weil sie bereits in Gabmap implementiert ist. Andere Methoden zum Detektieren von bestimmenden Faktoren sind beispielsweise in Prokic (2007) beschrieben. Dort wird am Beispiel von bulgarischen Dialekten untersucht, welche Unterschiede in der Aussprache in welchem Mass für Dialektgrenzen verantwortlich sind. Wieling und Nerbonne (2011) untersuchen die Wichtigkeit von phonetischen Phänomenen anhand deren Repräsentativität («Representativeness») und Besonderheit («Distinctiveness»). Nerbonne (2006) führt mit den Daten des «Linguistic Atlas of the Middle and South Atlantic States» eine Faktorenanalyse durch, um herauszufinden welche phonetischen Merkmale häufig zusammen auftreten. Nerbonne (2009) hat sich anhand von deutschen Dialekten angeschaut, welche linguistischen Distanzmatrizen von einzelnen Worten am besten mit den einzelnen Dimensionen der MDS-Analyse korrelieren.

In der vorliegenden Studie ist mit den gefundenen determinierenden Faktoren jeweils ein Set an minimalen Variablen gebildet worden, das anhand der MDS- und hierarchischen Clusteranalyse mit einem entsprechenden Set zufälliger Variablen verglichen worden ist. Um sicher zu gehen, dass die Resultate nicht zufällig zustande gekommen sind, sollten sie mit weiteren zufällig generierten Sets verglichen werden. Auch wäre es interessant zu sehen, wie sich die Resultate verändern, wenn die Analyse mit einem anderen Clusterbildungsalgorithmus, mit einer anderen Anzahl Cluster oder angewandt auf die einzelnen sprachlichen Ebenen durchgeführt würde. Weiter könnten auch diejenigen Variablen gesucht werden, welche über alle Cluster die höchste durchschnittliche Differenz in den z-Werten aufweisen.

7.4 Interpretation der Resultate

Alle erhaltenen Resultate deuten darauf hin, dass die aggregierten Phänomene der Syntax nicht im selben Masse raumbildend sind wie diejenigen der anderen Ebenen. Die untersuchten Phänomene der Lexik beispielsweise scheinen räumlich eher homogen zu sein. Wenn also ein Wort in einem Ort verwendet wird, kommt es eher auch in der direkten Umgebung vor, meist jedoch nicht in irgendeiner weiter entfernt liegenden Region des untersuchten Sprachgebietes. Auch auf der phonetischen Ebene fällt auf, dass die Phänomene oft räumlich homogen auftreten. Bei der Syntax ist dies weniger der Fall. Sieht man sich die Karten der einzelnen Phänomene der syntaktischen Ebene an, kommt es zwar teilweise vor, dass die Phänomene raumbildend sind. Die einzelnen Gebiete, die sich dann aufgrund der geographischen Verteilung der syntaktischen Varianten ergeben sind dann im Allgemeinen grossflächig und teilen beispielsweise die Schweiz in ein östliches und westliches Gebiet auf (z.B. bei den Phänomenen «Finalanschluss» und «Verdoppelung lassen») oder auch in ein nördliches und südliches (z.B. beim Passiv). Oft ist es jedoch auch so, dass Varianten in den unterschiedlichsten Regionen der Schweiz verwendet werden und sich ein stark fragmentiertes Raumbild ergibt. Dies könnte verschiedene Gründe haben. Die syntaktischen Phänomene lassen sich allgemein nicht leicht isolieren und strukturieren (Löffler 2003, S. 109). Somit gibt es auch unterschiedliche Herangehensweisen, was die Auswahl der syntaktischen Phänomene und Varianten betrifft, die in eine Studie einfließen sollen. In der vorliegenden Arbeit sind beispielsweise nicht nur strukturell ähnliche Varianten als gültig akzeptiert worden, sondern auch funktional ähnliche. Dies lässt sich anhand des Beispiels der Übersetzungsfrage III.1 «das Eis beginnt zu schmelzen» erläutern. Darin geht es eigentlich um die Verwendung des Verbes «anfangen». Nun werden jedoch nicht nur die Varianten mitberücksichtigt, welche auch «anfangen» beinhalten, sondern auch «schmilzt s Iis», in welcher das Verb «anfangen» ganz weggelassen wird. In anderen Ansätzen werden strikt nur bestimmte, vorgegebene Varianten akzeptiert. Ausserdem liessen sich in einigen Fällen auch die Antworten auf entsprechende SADS-Fragen auch mehrfach klassifizieren, was zu einer grösseren Anzahl an «Arbeitskarten» (Goebel 1982a, S. 779) bei einer gleichzeitigen Reduktion der jeweils möglichen Antworten führen würde. Als Beispiel dafür kann die Frage III.28 betrachtet werden, die das Phänomen des Komparativs erforscht. Erstens könnte hier untersucht werden, ob der Komparativ ein- oder zweiteilig ist und zweitens wie die entsprechenden Bestandteile realisiert sind. Verschiedene Phänomene hätten also in mehr Karten mit weniger Varianten aufgeteilt werden können. Auf diese Weise und mit der ausschliesslichen Berücksichtigung von strukturell ähnlichen Varianten, hätte die Anzahl Varianten pro Variable verringert werden können. Eventuell hätte dies zu anderen Resultaten geführt und vielleicht lassen sich die höheren Korrelationen der Syntax mit der Phonetik aus der Studie von Spruit (2008) auch mit solchen unterschiedlichen Herangehensweisen erklären. Allerdings kann darüber nur spekuliert werden. Weniger Varianten pro Variable heisst ja auch nicht zwingend, dass diese dann raumbildend sind.

Dass die Syntax nicht im selben Masse raumbildend ist wie die Lexik, die sogar durchschnittlich noch mehr Varianten pro Variable aufweist, zeigt sich auch an den wenigen syntaktischen Variablen, die im Gegensatz zu lexikalischen als Cluster determinierend aufgetreten sind. Nur das Cluster in der Ostschweiz wird am stärksten durch eine

7 Diskussion

Wortstellungskarte determiniert. Interessant wäre eine Untersuchung, ob solche Varianten, die in einer bestimmten Region homogen auftreten und in anderen Regionen nicht, auch in der Wahrnehmung der Leute als typisch gelten.

8 Fazit

Diese Arbeit hat die Dialekte der Deutschschweiz zum ersten Mal anhand einer dialektometrischen Studie untersucht, die alle vier linguistischen Ebenen Lexik, Morphologie, Phonetik und Syntax miteinbezieht und direkt miteinander vergleicht.

Dieses Kapitel fasst das Erreichte kurz zusammen. Der erste Teil greift die Forschungsfragen nochmals auf und liefert prägnante Antworten. Anschliessend werden die Grenzen der Arbeit aufgezeigt. Den Abschluss bilden Ideen für zukünftige Fragestellungen in Bezug auf die dialektometrische Analyse von schweizerdeutschen Dialekten und die Erforschung von sprachlichen Phänomenen im Raum.

8.1 Beantwortung der Forschungsfragen

Die erste Forschungsfrage lautet:

- 1) *Welche Unterschiede gibt es in der räumlichen Verteilung von aggregierten Distanzen morphologischer, phonetischer, lexikalischer und syntaktischer Phänomene in der Deutschschweiz?*

Diese Forschungsfrage ist einerseits mit visuellen Vergleichen anhand von, für jede linguistische Ebene, generierten Parameterkarten (Mittelwerte, Minima und Schiefen), MDS-Karten und hierarchischen Clusteranalysen beantwortet worden.

Die Analysen der Parameterkarten haben gezeigt, dass sich die lexikalische und die syntaktische Ebene in Bezug auf die Verteilung der Mittelwerte und der Schiefen besonders stark unterscheiden. Bei den Minima finden sich die stärksten Gegensätze in der phonetischen und syntaktischen Ebene. Allgemein haben die Analysen angedeutet, dass sich auf der syntaktischen Ebene am schwersten klar trennbare Dialektregionen finden lassen. Auch aus den MDS-Karten lassen sich auf syntaktischer Ebene deutlich weniger Dialektregionen bzw. Grenzen ausmachen als in den andern drei Ebenen, was auch eine Einteilung in Cluster auf dieser Ebene besonders schwierig macht.

In einer anderen Herangehensweise sind die Verteilungen der linguistischen Distanzen in den vier sprachlichen Ebenen anhand von Korrelationskoeffizienten zwischen den Ebenen verglichen worden. Während die Varianz in der lexikalischen Ebene jeweils zu rund 60 % von der morphologischen bzw. phonetischen Ebene erklärt werden kann, sind es bei der syntaktischen Ebene jeweils nur um die 30 % der Varianz, welche sich durch eine der anderen Ebenen erklären lassen.

Diese erste Forschungsfrage ist in den folgenden untergeordneten Fragen konkretisiert worden, welche hier ebenfalls noch kurz beantwortet werden sollen:

Ähneln sich die Dialektregionen- oder Landschaften, welche sich aus den einzelnen linguistischen Ebenen ergeben oder fällt eine bestimmte Ebene von den anderen ab?

Die Multidimensionale Skalierung und die hierarchische Clusteranalyse haben gezeigt, dass sich die Dialektregionen der lexikalischen, morphologischen und phonetischen Ebene eher ähneln, während die syntaktische Ebene von den anderen abfällt.

Welche linguistische Ebene ist hauptsächlich für die Dialektlandschaft der Deutschschweiz verantwortlich?

Die Korrelationsanalyse der linguistischen Distanzen der einzelnen Ebenen mit den standardisierten aggregierten Distanzen aller Ebenen hat gezeigt, dass die Morphologie am stärksten korreliert mit den totalen Distanzen und daher hauptsächlich für die Dialektlandschaft der Deutschschweiz verantwortlich zu sein scheint. Auch die hierarchische Clusteranalyse hat gezeigt, dass in der morphologischen Ebene der grösste Anteil der Orte identisch klassiert wird wie in der Clusteranalyse aller Daten.

Korrelieren Karten bzw. Distanzmatrizen, in denen phonetische, lexikalische, morphologische und syntaktische Phänomene aggregiert wurden am ehesten mit Karten, in denen a) nur phonetische, b) nur morphologische, c) nur lexikalische oder d) nur syntaktische Phänomene berücksichtigt wurden?

Die Reihenfolge der Bestimmtheitsmasse mit der standardisierten totalen Gesamtmatrix ist:

b) Morphologie 82 % vor a) Phonetik 80 % vor c) Lexik 79 % vor d) Syntax 57 %

Die zweite Forschungsfrage lautet:

2) *Welche linguistischen Phänomene trennen speziell gut?*

Nach der Methode von Cöltekin et al. (2012) und dem Clustern der Dialekte nach dem Ward-Algorithmus mit zehn Clustern sind die folgenden Variablen gefunden worden, welche diese Cluster am besten determinieren, also besonders gut trennen:

Lexik	Morphologie	Phonetik	Syntax
5112 Partikel: Grussformeln	3154 Vög[el]chen: Subst. Diminutiv	1134 t[ie]f: aobd. iu	3262 gehen lassen: Wortstellung
4176 nur	3075 (an)fangen: 2./3.Pers.Sg.Ind.Präs.	2138 ge[rn]: mhd. rn	
6027 Zeitadverb: jeweils (2x)	3047 haben: Pl.Ind.Präs.		

Tabelle 18 Die Cluster bestimmenden Variablen nach linguistischer Ebene

Allerdings konnten, mit einer Ausnahme, keine «perfekt trennenden» Variablen gefunden werden, deren Ausprägungen innerhalb eines Clusters stark homogen sind und die sich gleichzeitig zu den Ausprägungen in den anderen Clustern stark unterscheiden. Meist ist die Homogenität stärker ausgeprägt als die Unterschiedlichkeit. Berechnet man für jedes Cluster die zehn Variablen, welche dieses am besten determinieren, findet sich keine Variable, die es für mehr als vier Cluster in die jeweilige Auswahl schafft.

Lässt sich ein minimales Set an Variablen finden, das besonders stark mit den aggregierten Phänomenen aller Ebenen korreliert?

Die linguistischen Distanzen aus dem Set der in Tabelle 18 aufgeführten neun Variablen erklärt die Varianz der Distanzen aller vorhandenen Variablen zu 56 % während ein zufälliges Set aus neun Variablen diese nur zu 42 % erklärt.

Falls sich ein solches finden lässt, stammen diese Variablen vorwiegend aus einer bestimmten linguistischen Ebene?

Nimmt man aus der Clusterbildung nach Ward mit zehn Clustern jeweils die ersten zehn Variablen, welche diese Cluster am besten determinieren, erhält man ein Set mit 79 Variablen. Absolut gesehen stammen daraus die meisten Variablen aus der morphologischen Ebene (35), gefolgt von der Phonetischen (28) und der Lexikalischen (14). Nur zwei Variablen sind der syntaktischen Ebene zuzuschreiben. Der relative Anteil im Set enthaltener Variablen an den vorhandenen Variablen einer bestimmten Ebene ist in der Lexik am höchsten (38.9 %), gefolgt von der Phonetik (37.7 %) und der Morphologie (29.2 %). Auch hier bildet die Syntax mit 1.8 % das Schlusslicht.

8.2 Grenzen

Die Studie hat gezeigt, dass sich Dialektlandschaften der Deutschschweiz, welche sich aus den neueren Daten des Syntaktischen Atlas der deutschen Schweiz ergeben, vergleichen lassen mit den vor rund 60 Jahren erhobenen Daten des Sprachatlas der Deutschen Schweiz, der die linguistischen Ebenen der Lexik, Morphologie und Phonetik abdeckt. Trotzdem sind diesem Vergleich auch Grenzen gesetzt, die hier zusammengefasst werden.

Ortsnetz

Beim SDS konnten nicht alle Ortspunkte berücksichtigt werden, da das Ortspunktenetz vom SADS weniger dicht ist. Die durchgeführten Vergleiche zwischen den Datensätzen des SDS mit allen 565 Ortspunkten und dem reduzierten Ortsnetz haben zwar jeweils keinen markanten Unterschied bedeutet, trotzdem ist das Ausmass des Informationsverlustes nicht genau abschätzbar. Auch sind 38 Ortspunkte aus dem SADS, welche keine exakte Entsprechung im SDS haben, einem ausgewählten SDS-Punkt zugeordnet worden, um möglichst viele Ortspunkte in die Analysen miteinfließen lassen zu können. Diese Zuteilung ist mit den Linguistinnen und Linguisten des SADS-Teams abgesprochen worden. Trotzdem ist es denkbar, dass sich die Dialekte in den zugeteilten Orten eventuell in einigen Variablen unterschieden hätten. Spruit (2008) hat in seiner Studie aus diesem Grund nur die Ortspunkte in den Vergleich zwischen zwei Datensätzen miteinbezogen, welche exakt übereinstimmten, wodurch jedoch sein Ortspunktenetz stark ausgedünnt wurde.

Datenauswahl

Ebenfalls ist nicht klar, wie sich die Resultate der Analysen verändert hätten, wenn von den Antworten des SADS nicht nur jeweils die natürlichste Antwort einer ausgewählten Gewährsperson, sondern die akzeptierten Varianten aller im Ort befragten Personen in die Analysen miteinbezogen worden wären. Gemäss Scherrer (2012) führt der Einbezug aller akzeptierten Varianten pro Ort zu einer Dialektlandschaft, in der die geographische Distanz stärker mit der linguistischen Distanz korreliert, als dies bei der Auswahl einer natürlichsten Variante der Fall ist. Eventuell wäre die so generierte syntaktische Dialektlandschaft den anderen Ebenen ähnlicher gewesen, als dies die Resultate in der vorliegenden Studie gezeigt haben.

Informantenauswahl

Die Arbeit gibt auch keinen Aufschluss darüber, in welchem Ausmass sich die Resultate verändert hätten, wenn statt, in Anlehnung an die Methode des SDS, jeweils nur eine Gewährsperson pro Ort zu berücksichtigen, jeweils die dominante natürlichste Antwort pro Ort ausgewählt worden wäre. Es ist nicht klar in wieviel Prozent der Fälle die ausgewählte Gewährsperson nicht die dominante Variante pro Ort als natürlichste genannt hat.

Erhebungszeitraum

Auch die Einflüsse der unterschiedlichen Erhebungsmethoden und des unterschiedlichen Erhebungszeitraums zwischen den beiden Datensätzen konnten mangels korrespondierenden Variablen nur an einem einzigen Beispiel untersucht werden (s. Abschnitt 4.1.5). Es ist also nicht genau quantifizierbar, welche gefundenen Unterschiede zwischen der syntaktischen Ebene und den anderen linguistischen Ebenen auf den Zeitunterschied zurückzuführen sind.

Geographie

Bei den Analysen der Zusammenhänge zwischen den linguistischen Distanzen der einzelnen Ebenen ist der Faktor Geographie nicht miteinbezogen worden. Es ist also nicht untersucht worden, zu welchem Anteil die Zusammenhänge zwischen den linguistischen Distanzen rein auf die geographischen Distanzen zwischen den Orten zurückzuführen sind, wie dies Spruit (2008) im Anschluss an seine Korrelationsanalysen noch getan hat.

Determinierende Variablen

In der vorliegenden Studie ist nur anhand einer Methode, angewandt auf die Variablen aller Ebenen gemeinsam, ein minimales Set an besonders gut trennenden Variablen bestimmt worden. Dieses ist dann jeweils auch nur mit einem zufällig gebildeten Set verglichen worden. Ob das gefundene minimale Set auch im Vergleich zu weiteren zufällig generierten Sets besser abgeschnitten hätte, wurde nicht eruiert.

Interpretation

Zu guter Letzt ist die Auswahl und Interpretation der Resultate durch eine Geographin ohne grossen linguistischen Hintergrund erfolgt. Es wäre interessant zu wissen, wie Linguisten und Linguistinnen die Fülle an generierten Resultaten interpretieren würden und auf welche Erkenntnisse sie die Schwerpunkte gesetzt hätten.

8.3 Ausblick

Mit den hier aufbereiteten Daten liessen sich noch eine Reihe anderer Analysen durchführen. Zum Beispiel könnten die Korrelationsanalysen mittels VDM visualisiert und so festgestellt werden, in welchen Regionen die linguistischen Distanzen eines Ortes zu allen anderen Orten zwischen zwei sprachlichen Ebenen besonders gut korrelieren. Weiter wäre auch der Einbezug geographischer Distanzen in die Analysen interessant. Im Rahmen des SNF-Projektes SynMod, das ebenfalls eine Zusammenarbeit zwischen dem Geographischen Institut und dem Deutschen Seminar der Universität Zürich darstellt, wird dieser Frage bereits nachgegangen.

Eine Überprüfung der Methoden und Ergebnisse durch Linguistik-Experten könnte Aufschluss darüber geben, ob die vorgenommenen Datenreduktionen sinnvoll waren. Beispielsweise könnte anhand des vergleichbaren Phänomens zwischen dem SDS und dem SADS-Datensatz überprüft werden, welche Gewährspersonen aus dem SADS am ehesten dieselben Antworten gegeben haben wie diejenigen im SDS. So könnte festgestellt werden, ob die Auswahl einer Gewährsperson aufgrund ihres Alters sinnvoll war oder nicht. Auch könnten die Antworten der ausgewählten Gewährspersonen mit den jeweils dominanten Antworten pro Ort verglichen und deren Übereinstimmung überprüft werden. Weichen die Antworten stark voneinander ab, wäre zu überlegen, die Studie nochmals mit den dominanten Antworten aus dem SADS durchzuführen.

Durch die Digitalisierung zusätzlicher Karten aus dem SDS könnte die Anzahl der in zukünftige Studien einflussenden Variablen ausgeglichen werden. Es wäre interessant zu sehen, wie dies vor allem die visuellen Ergebnisse beeinflussen würde.

Für die Bestimmung derjenigen Variablen, welche besonders gut trennen, könnten andere Methoden ausprobiert und die Ergebnisse verglichen werden.

Schliesslich wäre es interessant, die durch die Dialektometrie gefundenen Dialektgrenzen der Deutschschweiz mit kleinräumigen natürlichen Grenzen (z.B. Topographie) in Bezug zu setzen und beispielsweise in einem Dreidimensionalen Modell zu visualisieren.

Die Themenvielfalt deutet darauf hin, dass einer weiteren Zusammenarbeit zwischen dem Geographischen Institut und dem Deutschen Seminar zur Erforschung sprachlicher Variation im Raum nichts im Wege steht.

Literaturverzeichnis

- Bart, Gabriela; Glaser, Elvira (2011): Vergleich zwischen Punktsymbol- und Flächenkarten im Rahmen des SADS-Projekts. Esri Campusday. Universität Zürich. Zürich, 16.09.2011.
- Bortz, Jürgen (1993): Statistik. Für Sozialwissenschaftler. 4., vollst. überarb. Aufl. Berlin [u.a.]: Springer (Springer-Lehrbuch).
- Bucheli, Claudia (2005): Passiv im Schweizerdeutschen (Linguistik online, 24, 3/05). Online verfügbar unter http://www.linguistik-online.de/24_05/bucheli.html, zuletzt aktualisiert am 28.11.2005, zuletzt geprüft am 17.03.2014.
- Bucheli, Claudia (2008): Neue Technik, alte Probleme: auf dem Weg zum Syntaktischen Atlas der Deutschen Schweiz (SADS). In: Stephan Elspass und Werner König (Hg.): Sprachgeographie digital. Die neue Generation der Sprachatlanten. 1. Aufl. Hildesheim, New York: G. Olms (Germanistische Linguistik, 190-191, 2008).
- Bucheli, Claudia; Glaser, Elvira; Seiler, Guido (2012): Is a syntactic dialectology possible? Contributions from Swiss German. In: Andrea Ender, Adrian Leemann und Bernhard Wälchli (Hg.): Methods in contemporary linguistics. Berlin [etc.]: De Gruyter Mouton (Trends in linguistics. studies and monographs, 247), S. 93–120.
- Bucheli, Claudia; Glaser, Elvira (2002): The Syntactic Atlas of Swiss German Dialects: empirical and methodological problems. In: Sjeff Barbiers, Leonie Cornips und Susanne van der Kleij (Hg.): Syntactic Microvariation. Elektronische Publikation. Online verfügbar unter <http://www.meertens.knaw.nl/books/synmic/>.
- Burrough, P. A.; McDonnell, Rachael (1998): Principles of geographical information systems. Oxford, New York: Oxford University Press (Spatial information systems).
- Bussmann, Hadumod; Gerstner-Link, Claudia (2002): Lexikon der Sprachwissenschaft. 3., aktualisierte und erw. Aufl. Stuttgart: Kröner.
- Castellazzi, B. (2010): Guide to export Thiessen maps from ArcGIS 9.3 to VDM. Universität Salzburg.
- Christen, Helen; Friedli, Matthias; Glaser, Elvira (2011): Der Kleine Sprachatlas der deutschen Schweiz: Von der Idee zum Volksbuch. In: Verein für das Schweizerdeutsche Wörterbuch. Schweizerdeutsches Wörterbuch: Bericht über das Jahr 2010, S. 25–49.
- Cöltekin, Cagri; Nerbonne, John; Prokic, Jelena (2012): Detecting shibboleths. In: *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*, S. 72–80.
- Cronbach, Lee J. (1951): Coefficient alpha and the internal structure of tests. In: *Psychometrika* 16 (3), S. 297–334. DOI: 10.1007/BF02310555.
- Embleton, Sheila (1993): Multidimensional Scaling as a Dialectometrical Technique: Outline of a Research Project. In: Reinhard Köhler und Burghard B. Rieger (Hg.): Contributions to Quantitative Linguistics. Dordrecht: Springer Netherlands, S. 267–276.
- Fleischer, Wolfgang; Helbig, Gerhard; Lerchner, Gotthard (2001): Kleine Enzyklopädie, deutsche Sprache. Frankfurt am Main: P. Lang.
- Girnth, Heiko (2010): Mapping language data. In: Alfred Lameli, Roland Kehrein und Stefan Rabanus (Hg.): Language and space. Language mapping: an international handbook of linguistic variation. Berlin, New York: De Gruyter Mouton (Handbooks of linguistics and communication science, 30.2), S. 98–145.

- Goebel, Hans (1982a): Ansätze zu einer computativen Dialektometrie. In: Werner Besch, Ulrich Knoop, Wolfgang Putschke und Herbert Ernst Wiegand (Hg.): *Dialektologie. Handbuch zur dt. und allg. Dialektforschung*. Berlin: de Gruyter (Handbooks of linguistics and communication science, 1), S. 778–792.
- Goebel, Hans (1982b): *Dialektometrie. Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie*. Wien: Verlag der österreichischen Akademie der Wissenschaften.
- Goebel, Hans (1984): *Dialektometrische Studien: Anhand italo-romanischer, rätomanischer und galloromanischer Sprachmaterialien aus AIS und ALF*. 3 Vol. Tübingen: Max Niemeyer.
- Goebel, Hans (1995): *Dialektometrische Beschreibung der Romania*. In: Günter Holtus, Michael Metzeltin und Christian Schmitt (Hg.): *Lexikon der Romanistischen Linguistik (LRL). Les différentes langues romanes et leurs régions d'implantation du Moyen Âge à la Renaissance*. Tübingen: M. Niemeyer, S. 977–1003.
- Goebel, Hans (2004): *VDM - Visual Dialectometry. Vorstellung eines dialektometrischen Software-Pakets auf CD-ROM*. In: Wolfgang Dahmen, Günter Holtus, Johannes Kramer, Michael Metzeltin, Schweickard Wolfgang und Winkelmann Otto (Hg.): *Romanistik und neue Medien. Romanistisches Kolloquium XVI*. Tübingen: Gunter Narr Verlag, S. 209–241.
- Goebel, Hans (2006a): (Meta)Sprachliche Kon- und Divergenzen im Bereich der Sprachlandschaft Ladinien. Ein dialektometrischer Vergleich anhand subjektiver und objektiver Dialektdaten. In: *Ladinia* 2006 (30), S. 223–283.
- Goebel, Hans (2006b): *Recent Advances in Salzburg Dialectometry*. In: *Literary and Linguistic Computing* 21 (4), S. 411–435. DOI: 10.1093/lc/fql042.
- Goebel, Hans (2010): *Dialectometry and quantitative mapping*. In: Alfred Lameli, Roland Kehrein und Stefan Rabanus (Hg.): *Language and space. Language mapping: an international handbook of linguistic variation*. Berlin, New York: De Gruyter Mouton (Handbooks of linguistics and communication science, 30.2), S. 433–464.
- Goebel, Hans; Scherrer, Yves; Smečka, Pavel (2013): *Kurzbericht über die Dialektometrisierung des Gesamtnetzes des „Sprachatlasses der deutschen Schweiz“ (SDS)*. In: Karina Schneider-Wiejowski, Birte Kellermeier-Rehbein und Jakob Haselhuber (Hg.): *Vielfalt, Variation und Stellung der deutschen Sprache: de Gruyter*, S. 153–176.
- Grieve, Jack; Speelman, Dirk; Geeraerts, Dirk (2011): *A statistical method for the identification and aggregation of regional linguistic variation*. In: *Lang Var Change* 23 (02), S. 193–221. DOI: 10.1017/S095439451100007X.
- Heeringa, Wilbert (2004): *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Dissertation. University of Groningen, Groningen. Faculty of Arts, the Humanities Computing department.
- Hogg, Robert V.; McKean, Joseph W.; Craig, Allen T. (2013): *Introduction to mathematical statistics*. 7th ed. Boston: Pearson.
- Hotzenköcherle, Rudolf; Baumgartner, Heinrich (1962-1997, 2003): *Sprachatlas der deutschen Schweiz*. Bern: Francke Verlag.
- Hotzenköcherle, Rudolf; Bigler, Niklaus; Schläpfer, Robert; Börlin, Rolf (1984): *Die Sprachlandschaften der deutschen Schweiz*. Aarau: Sauerländer (Reihe Sprachlandschaft, Bd. 1).
- Hotzenköcherle, Rudolf; Trüb, Rudolf (Hg.) (1962): *Sprachatlas der deutschen Schweiz. Band I, Lautgeographie. Vokalqualität*. Bern: Francke.
- Kelle, Bernhard (2001): *Zur Typologie der Dialekte in der deutschsprachigen Schweiz: Ein dialektometrischer Versuch*. In: *Dialectologia et Geolinguistica* 2001 (9), S. 9–34.

- Lanwermeijer, Manuela; Womelsdorf, Tobias (2009): Digitaler Wenker-Atlas (DiWA). Kartierungsverfahren. Unter Mitarbeit von Alfred Lameli, Tanja Giessler, Roland Kehrein, Alexandra Lenz, Karl-Heinz Müller, Jost Nickel et al. Hg. v. Jürgen Erich Schmidt und Joachim Herrgen. Forschungszentrum Deutscher Sprachatlas. Marburg. Online verfügbar unter <http://www.diwa.info/Geschichte/Kartierungsverfahren.aspx>, zuletzt geprüft am 18.02.2014.
- Leinonen, Therese (2013): Gabmap user documentation. Manual fuzzy clustering. Unter Mitarbeit von Peter Kleiweg, Charlotte Gooskens, Rinke Colen und John Nerbonne. Information Science. Groningen. Online verfügbar unter <http://www.gabmap.nl/~app/doc/manual/clustering-fuzzy.html>, zuletzt aktualisiert am 27.02.2013, zuletzt geprüft am 11.03.2014.
- Löffler, Heinrich (2003): Dialektologie. Eine Einführung. Tübingen: Narr (Narr Studienbücher).
- Manly, B.F. J. (2007): Randomization, bootstrap and Monte Carlo methods in biology. 3rd ed. Boca Raton, FL: Chapman & Hall/ CRC (Texts in statistical science).
- Mantel, N. (1967): The detection of disease clustering and a generalized regression approach. In: *Cancer Res.* 27 (2), S. 209–220.
- Nerbonne, John (2006): Identifying Linguistic Structure in Aggregate Comparison. In: *Literary and Linguistic Computing* 21 (4), S. 463–475. DOI: 10.1093/lc/fql041.
- Nerbonne, John (2009): Data-Driven Dialectology. In: *Language and Linguistics Compass* 3 (1), S. 175–198.
- Nerbonne, John (2010): Mapping Aggregate Variation. In: Alfred Lameli, Roland Kehrein und Stefan Rabanus (Hg.): *Language and space. Language mapping: an international handbook of linguistic variation*. Berlin, New York: De Gruyter Mouton (Handbooks of linguistics and communication science, 30.2), S. 476–501.
- Nerbonne, John; Colen, Rinke; Gooskens, Charlotte; Kleiweg, Peter; Leinonen, Therese (2011): Gabmap - a Web Application for Dialectology. In: *Dialectologia et Geolinguistica* 2011 (Special issue II), S. 65–89. Online verfügbar unter <http://www.raco.cat/index.php/Dialectologia/article/view/245345/328591>.
- Nerbonne, John; Heeringa, Wilbert; van der Hout, Erik; van der Kooij, Peter; Otten, Simone; van de Vis, Willem (1996): Phonetic Distance between Dutch Dialects. In: Gert Durieux und Walter Daelemans (Hg.): *CLIN VI. Papers from the sixth CLIN Meeting*. Antwerp: University of Antwerp, UIA, Center for Dutch Language and Speech, S. 185–202.
- Nerbonne, John; Kleiweg, Peter (2007): Toward a Dialectological Yardstick. In: *Journal of Quantitative Linguistics* 2007 (14(2)), S. 148–167. Online verfügbar unter <http://www.let.rug.nl/nerbonne/papers/yardstick.pdf>, zuletzt geprüft am 10.02.2014.
- Nerbonne, John; Kretzschmar, William (2003): Introducing Computational Methods in Dialectometry. In: *Computers and the Humanities* (37), S. 245–255.
- Nunnally, Jum C.; Bernstein, Ira H. (1994): *Psychometric theory*. 3rd ed. New York: McGraw-Hill (McGraw-Hill series in psychology).
- Pfister, Beat; Kaufmann, Tobias (2008): *Sprachverarbeitung: Grundlagen und Methoden der Sprachsynthese und Spracherkennung*. Berlin, Heidelberg: Springer.
- Prokic, Jelena (2007): Identifying linguistic structure in a quantitative analysis of dialect pronunciation. In: *Proceedings of the ACL (Association for Computational Linguistics), Student Research Workshop*, S. 61–66.
- Rumpf, Jonas; Pickl, Simon; Elspass, Stephan; König, Werner; Schmidt, Volker (2009): Structural analysis of dialect maps using methods from spatial statistics. In: *Zeitschrift für Dialektologie und Linguistik*, S. 280–308. DOI: 10.2307/25735021.

- Scherrer, Yves (2011): Neue dialektometrische Versuche mit schweizerdeutschem Dialektmaterial. Geolinguistik-Workshop. Freiburg im Breisgau, 12/2011.
- Scherrer, Yves (2012): Generating Swiss German Sentences from Standard German – A Multi-dialectal Approach. Dissertation. Universität Genf, Genf.
- Schrambke, Renate (2010): Traditional dialect geography. In: Peter Auer und Jürgen Erich Schmidt (Hg.): Language and Space. Volume 1 Theories and Methods (Handbooks of linguistics and communication science, 30.1), S. 107–125.
- Séguy, Jean (1971): La relation entre la distance spatiale et la distance lexicale. In: *Revue de Linguistique Romane* (35), S. 335–357.
- Sibler, Pius (2011): Visualisierung und geostatistische Analyse mit Daten des Syntaktischen Atlas der Deutschen Schweiz (SADS). Masterarbeit. Universität Zürich, Zürich. Geographisches Institut.
- Spruit, Marco René (2008): Quantitative perspectives on syntactic variation in Dutch dialects. Utrecht: LOT.
- Staffeldt, Sven (2010): Einführung in die Phonetik, Phonologie und Graphematik des Deutschen. Ein Leitfaden für den akademischen Unterricht. 1. Aufl. Tübingen: Stauffenburg-Verl (Stauffenburg-Einführungen, 21).
- Tavakol, Mohsen; Dennick, Reg (2011): Making sense of Cronbach's alpha. In: *Int. J. Medical Education* 2, S. 53–55. DOI: 10.5116/ijme.4dfb.8dfd.
- Torgerson, Warren S. (1952): Multidimensional scaling: I. Theory and method. In: *Psychometrika* 17 (4), S. 401–419. DOI: 10.1007/BF02288916.
- Trüb, Rudolf (Hg.) (2003): Sprachatlas der deutschen Schweiz Abschlussband. Werkgeschichte, Publikationsmethode, Gesamtregister. Unter Mitarbeit von Lily Trüb.
- Uiboaed, K.; Hasselblatt, C.; Lindstrom, L.; Muischnek, K.; Nerbonne, J. (2013): Variation of verbal constructions in Estonian dialects. In: *Literary and Linguistic Computing* 28 (1), S. 42–62. DOI: 10.1093/lc/fqs053.
- Wagner, Silke (2003): Optimierung des Demographic Clustering Algorithmus. Diplomarbeit. Universität Konstanz, Konstanz.
- Ward, Joe (1963): Hierarchical Grouping to optimize an objective function. In: *Journal of American Statistical Association* 58 (301), S. 236–244.
- Wenker, Georg (1881): Sprachatlas von Nord- und Mitteldeutschland. Auf Grund von systematisch mit Hilfe der Volksschullehrer gesammeltem Material aus circa 30.000 Orten. Straßburg.
- Wieling, Martijn; Nerbonne, John (2011): Bipartite spectral graph partitioning for clustering dialect varieties and detecting their linguistic features. In: *Computer Speech & Language* 25 (3), S. 700–715. DOI: 10.1016/j.csl.2010.05.004.
- Wilks, Daniel S. (2006): Statistical methods in the atmospheric sciences. 2nd ed. Amsterdam, Boston: Academic Press (International geophysics series, v. 91).
- Zastrow, Thomas (2011): Neue Analyse- und Visualisierungsmethoden in der Dialektometrie. Dissertation. Eberhard Karls Universität, Tübingen. Philosophische Fakultät.

Anhang

A. Liste der miteinbezogenen SDS-Karten¹³

Karten Nr.	Karten Titel	Karten Nr.	Karten Titel
Phonetik			
1015	B[e]tt: mhd. e (Primärumlaut)	2003	Schn[a]bel: mhd. a offene Silbe
1021	Sp[e]ck: mhd. ë (germ. e)	2015	M[a]gen: mhd. ä
1035	[e]ng: mhd. e vor Nas + Kon.	2022	f[ah]ren: mhd. a offene Silbe
1041	G[o]tte: mhd. o	2027	B[e]sen: Dehnung von alter Kürze in offener Silbe, mhd. ë
1045	D[o]nnstig: mhd. o	2044	s[a]gen: Dehnung von alter Kürze in offener Silbe, mhd. a
1048	Schl[i]tten: mhd. i	2069	Gesch[irr]: Dehnung vor rr
1050	Ch[u]chi: mhd. u	2071	schr[ei]ben: Kürze in offener Silbe vor ben
1052	F[ü]chse: mhd. ü	2073	m[a]len: Kürze von alter Länge in offener Silbe vor len
1054	Br[ü]cke: Uml./ nicht Uml. mhd. u vor ck, gg	2074	m[ah]len: mhd. a/ a
1055	R[ü]cken: Uml./ nicht Uml. von mhd. u vor gg	2077	r[ei]ten: Kürze von alter Länge vor t, k
1056	dr[ü]cken: Uml./ nicht Uml. mhd. u vor ck, gg	2084	Pfl[aum]e(n): mhd. û vor m(m)
1060	d[u]rch: Umlt/ nicht Uml. mhd. u	2087	s[ä]gen: Infinitiv
1060	[ü]ber: Umlt/ nicht Uml. mhd. u	2088	tr[a]gen: Infinitiv
1061	[A]bend: mhd. â	2093	Gr[a]s, sp[a]ren, Umlaut
1067	S[a]men: mhd. â	2094	[K]ind: germ. k im Anlt
1095	Schn[ee]: mhd. ê (di/ mono)	2095	-li[ch]: urd. kk hint.Vok i. Inlt u. Auslt
1095	Schn[ee]: mhd. ê	2095	drü[ck]en: urd. kk hint.Vok i. Inlt u. Auslt
1100	L[oh]n: mhd. ô vor n	2097	tr[ink]en: germ. -nk(-)
1105	[Ei]s: mhd. î vor Kons.	2113	wa[chs]t: mhd. -hs-
1106	M[au]s: mhd. û vor Kons.	2118	la[ng]: mhd. -nc (< -NG)
1107	M[äu]se: mhd. Iu = Umlt vor û	2119	fi[nd]en: mhd. -nd-
1109	G[ei]ss: mhd. ei	2120	Hu[nd]: mhd. -nt (< -ND)
1112	S[eil]: mhd. ei vor l	2138	ge[rn]: mhd. -rn
1121	[Au]gen: mhd. ou	2143	Da[rm]: mhd. -rm
1121	[Au]gen: Monophthong	2144	Läu[s]e: mhd. s > sch
1124	B[au]m: mhd. ou vor m	2147	fo[l]ge: mhd. l vor Kon (Qual.)
1129	räuchern: mhd. öu	2161	m[ähe]n: Übergangslt bei V. pura
1134	t[ie]f: aobd. iu	2162	br[ühe]n: Übergangslt bei V. pura

¹³ Im Vergleich zu Scherrer (2012) in diese Studie neu einflussende (**fett**) oder geänderte Variablen (*kursiv*)

1138	Kn[ie]: aobd. iu	2164	[D]eckel: mhd. d (germ. D) im Anlt
1148	schn[e]jen: mhd. î in Hiatt	2165	[T]ag: mhd. t (germ. d) im Anlt
1152	b[au]jen: mhd. û in Hiatt	2170	Gar[b]e: mhd. b,d,g hinter r
1152	b[au]jen: mhd. û in Hiatt (Ent)	2171	ho[b]len: mhd. b, d, g, s vor l (m, n)
1152	b[au]jen: mhd. û in Hiatt (w)	2172	Ra[d]: ahd.d im abs. Auslt. ety. Kurzvok
1156	n[eu]-e,-er,-es: ü in Hiatt	2186	Ta[nn]e: mhd. nn zw. Vokal
1156	n[eu]-e,-er,-es: ü in Hiatt (Ent)	2192	Sa[m]jen: mhd. m zw. Vokal
1160	Epf[e]: e > ö (Rund)	2198	Te[ll]er: mhd. ll vor -er
1161	Schw[e]ster: e > ö (Rund)		
1164	Br[i]lle: i > ü (Rund)		
1165	Milch: i > ü (Rund)		
Morphologie			
3001	Infinitiv	3096	sehen: 1./ 2./ 3. Sg. Präf.
3001	Infinitiv: Gerundium	3097	sehen: 1./ 2./ 3. Pl.
3001	End. Infinit.: schwa eä	3100	kommen: 1. Pers. Sg.
3001	End. Infinit.: schwa eäa	3100	kommen: 2./ 3. Sg.
3001	End. Infinit.: schwa eäau	3102	kommen: 1.-3. Pers. Pl.
3003	Präfix des stark. Verbs: Part. Perfkt	3105	können (Inversion): 2.Pl.
3003	Präfix des schwa. Verbs: Part. Perfkt	3106	mögen: 1. Pers. Sg.
3007	gebracht: Part. Perfekt (f.)	3107	mögen (Inversion): 1./ 3.Pers.Pl.
3008	gelaufen: Part. Perfekt (f.)	3107	mögen (Inversion): 2. Pers. Pl.
3009	tragen: Infinitiv	3108	dürfen: 3. Pers. Sg.
3009	tragen: Part. Perfekt	3108	dürfen: 3. Pers. Pl.
3016	verl[ie]ren: SS Vok (2.Kl. stark. V.)	3112	wollen: 2. Pers. Sg.
3016	sch[ie]ssen: SS Vok (2.Kl. stark. V.)	3113	wollen: 3. Pers. Sg.
3018	fl[ie]gen, l[ü]gen: SS Vok	3114	wollen: 1. Pers. Pl.
3019	ziehen: SS Vok (Sonderfall)	3131	das: best. Art. Nom. Akk. Sg. Neutr
3022	End. d. Normalv.: 1. Pers. Sg. Ind. Präs.	3132	die: best. Art. Nom. Akk. Sg. f
3028	kostet: Endung Normalv.: 3. Pers. Sg. Ind. Präs.	3134	der: best. Art. Nom. Sg. m
3031	End. d. Normalv.: 1. Pers. Pl. Ind. Präs.	3142	ein (I): unbest. Art. Nom. Akk. Sg. Neutr
3032	End. d. Normalv.: 2.Pers. Pl. Ind. Präs.	3144	eine: unbest. Art. Nom. Akk. Sg. f
3033	End. d. Normalv.: 3. Pers. Pl. Ind. Präs.	3145	ein: unbest. Art. Nom. Akk. Sg. m
3040	End. d. Kurzverb.: 1. Pers. Sg. Ind. Präs.	3152	Trög[lein]: Subst. Diminutiv
3044	End. d. Kurzverb.: 1. Pers. Pl. Ind. Präs.	3154	Vög[el]chen: Subst. Diminutiv
3044	End. d. Kurzverb.: 2. Pers. Pl. Ind. Präs.	3159	Bern[erin]: Subst. f. Derivation
3044	End. d. Kurzverb.: 3. Pers. Pl. Ind. Präs.	3162	Wirt[in]: Subst. f. Derivation
3046	haben: Infinitiv	3163	Zeit[ung]:Subst. f. Derivation
3046	haben: Part. Perfekt	3164	Stiel-Stiel[e]:Subst. m. Pl. -e/ 0 (I)
3046	haben: 1. Pers. Sg. Ind. Präs.	3187	Subst. f. Pl. -i
3047	haben: Pl. Ind. Präs.	3195	Personalpr.: ich (voll)
3048	haben: 2./3. Pers. Sg. Ind. Präs.	3195	Personalpr.: mich (voll)

3050	sein: Infinitiv/ Part. Perfekt	3201	Personalpr.: sie (f. Sg.)
3050	<i>sein: Infinitiv/ Part. Perfekt (Pl)</i>	3201	Personalpr.: sie (Pl.)
3050	sein: 1. 2. 3. Pers. Pl. Ind. Präs	3203	Personalpr.: wir (i/ ier)
3051	sein: 1. 3. Pers. Sg. Konj. Präs.	3203	Personalpr.: wir
3054	tun: Part. Perfekt	3204	Personalpr.: ihr (2. Pers. Pl.)
3055	tun: Gerundium	3205	ihm: Personalpronomen
3056	gehen: 1. Pers. Sg. Ind. Präs.	3207	Personalpr.: uns
3056	kommen: Inf./ Part. Perfekt	3209	Personalpr.: euch
3057	<i>gehen (Vok): 2./ 3. Pers. Sg. Ind. Präs.</i>	3211	Possessivpr.: mein/ dein/ sein (m.)
3058	gehen (Vokal): Pl. Ind. Präs.	3212	Possessivpr.: mein/ dein/ sein (f.)
3058	gehen (Vokal): Pl. Ind. Präs.	3213	Possessivpr.: mein/ dein/ sein (n.)
3059	stehen: Pl. Ind. Präs.	3214	Possessivpr.: mein/ dein/ sein (pl.)
3064	<i>lassen: 1. Pers. Sg. Ind. Präs.</i>	3215	Possessivpr.: ihr (m.)
3066	lassen: Pl. Ind. Präs.	3216	Possessivpr.: ihr (f.)
3070	schlagen: 1. Pers. Sg. Ind. Präs.	3216	Possessivpr.: ihr (pl.)
3070	<i>schlagen: Inf.</i>	3217	Possessivpr.: unser
3071	<i>schlagen: 2./ 3. Pers. Sg. Ind. Präs.</i>	3217	Possessivpr.: unser (Dat.)
3072	schlagen: Pl. Ind. Präs.	3218	Possessivpr.: unser/ euer (f.)
3074	<i>(an)fangen: 1. Pers. Sg. Ind. Präs.</i>	3218	Possessivpr.: unser/ euer
3075	<i>(an)fangen: 2./ 3. Pers. Sg. Ind. Präs.</i>	3218	Possessivpr.: unser/ euer (pl.)
3076	(an)fangen: Pl. Ind. Präs.	3219	Possessivpr.: euer (n.)
3078	(an)fangen: Infinitiv	3220	Possessivpr.: mein (f., Dat.)
3080	liegen: 1.-3. Pl.	3223	Interrogativpr.: was
3081	liegen: 2./ 3. Pers. Sg.	3252	Adj.: gut[es] Nom. Akk. Sg. Neutr.
3086	müssen: 1./ 3. Pers. Sg.	3253	Adj.: alt[e] Nom. Akk. Pl. m./ f.
3087	<i>müssen: Ind. Pl.</i>	3254	Adj.: ganz[e] Nom. Akk. Sg. f.
3090	geben: 1.-3. Pl.	3259	du es/ es du: Wortstellung
3091	<i>nehmen: 1.-3. Pl.</i>	3260	ihr es/ es ihr: Wortstellung
3092	nehmen: 1. Pers. Sg.	3261	Syntax: [gewesen bin] f.
3092	<i>nehmen: 2./ 3. Pers. Sg.</i>	3262	gehen lassen: Wortstellung
3094	<i>geben/ nehmen: Vokalqual.</i>	3263	lass ihn gehen: erstarrte Infinitiv-Partikeln
3095	<i>nehmen: Part. Perfekt</i>	3265	es kommt regnen: erstarrte Infinitiv-Partikeln
3096	<i>sehen: 1./ 2./ 3. Sg.</i>	3266	wirst/ kommst: Umschr. inchoativen Funkt.
Lexik			
4149	alles	6020	Zeitadverb: gestern
4150	als: Konjunktion	6021	Zeitadverb: heute
4151	(ich will) auch: bet. Satzauslt.	6022	Zeitadverb: oft
4155	ein wenig	6026	Zeitadverb: immer
4159	gar: Adverb	6027	Zeitadverb: jeweils
4167	-nicht-: im Satzinlaut	6029	Zeitadverb: früher (Lautung)
4168	-nicht mehr: im Satzschluss	6030	Zeitadverb: bald

4171	nichts	6098	Ortsadv.: hier (von/ an dies. Ort)
4172	noch	6100	Ortsadverb: heim (Lautung)
4176	nur	6101	Ortsadverb: dort
4180	und	6105	Richtungsadv.: heraus
4181	sonst	6106	Richtungsadv.: hinaus
4183	Fragewort: wie?	6107	Richtungsadv.: herein
4184	wieder	6108	Richtungsadv.: hinein
5112	Partikel: Grussformeln	6109	Richtungsadv.: herauf
5115	Partikel: Dankformel	6110	Richtungsadv.: hinauf
5118	Partikel: schau	6111	Richtungsadv.: herab
5119	Partikel: ja	6112	Richtungsadv.: hinab

B. Liste mit den verwendeten SADS-Fragen

1. Fragebogen	Einbezogene, klassierte Varianten	Phänomen
I.1 Entschuldigung, ich habe zu wenig Kleingeld, um ein Billett zu lösen. (Übersetzung)	<ul style="list-style-type: none"> - für ... (zu) lösen - zum ... (zu)/ (zum) lösen - um ... (zu)/ (zum) lösen - für zum ... (zu) lösen - finaler Nebensatz mit dass bzw. damit - einfache Präpositionalphrase mit für 	Finalanschluss
I.2 Wem will er denn die schönen Blumen bringen? (Übersetzung)	<ul style="list-style-type: none"> - wem - xy wem 	Relativsatz
I.3 Ich habe den Fritz kommen hören. (Übersetzung)	<ul style="list-style-type: none"> - gehört kommen / kommen gehört - hören kommen / kommen hören 	Ersatzinfinitiv und Position der Verbalteile
I.4 Vervollständigen Sie den Antwortsatz; er soll Auskunft darüber geben, wo die Nachbarin ist („einkaufen“). (Ergänzung)	<ul style="list-style-type: none"> - go poschte - go poschte gange, gange go poschte - goge poschte - poschte - am poschte - bim poschte 	Go-Konstruktionen
I.5 Dä Chorb isch... (Ergänzung)	<ul style="list-style-type: none"> - umkippt - umkipptä 	Resultativ
I.6 Vervollständigen Sie den Antwortsatz; er soll Auskunft darüber geben, wozu die Politikerin Tabletten einnimmt („einschlafen“). (Ergänzung)	<ul style="list-style-type: none"> - zum ein(zu)schlafen - für ein(zu)schlafen - um ein(zu)schlafen - finiter Nebensatz mit dass bzw. damit - Nominalisierung für das (Ein-) Schlafen - für <z> einschlafen - für <z> schlafen - für <z> entschlafen 	Finalanschluss
I.7 Ist das dein neues Auto? (Ankreuzen)	<ul style="list-style-type: none"> - meiner Schwester - xy meiner Schwester - Nominalphrase mit Possessivpronomen - Nominalphrase mit Possessivpronomen + ‚Auto‘ 	Präpositionale Dativmarkierung (PDM)
I.8 Max soll Marianne beim Abwaschen helfen Max aber will sich drücken und antwortet: Ich habe gestern schon... (Ankreuzen):	<ul style="list-style-type: none"> - geholfen abwaschen / abwaschen geholfen - helfen abwaschen / abwaschen helfen 	Ersatzinfinitiv und Position der Verbalteile
I.9 Ich weiss nicht ober er mal heiraten wird. (Ankreuzen)	<ul style="list-style-type: none"> - will heiraten - heiraten will 	Position der Verbalteile im Nebensatz
I.10 Bruno könnte sich Martina an der Seite seines Sohnes vorstellen. Er sagt: (Ankreuzen)	<ul style="list-style-type: none"> - e ganz e - ganz e - e ganz 	Artikelverdoppelung
I.11 Ihre Mutter ruft an und will, dass Sie vorbeikommen. Sie haben sich gerade hingesetzt um zu lesen und antworten ihr ausweichend: (Ankreuzen)	<ul style="list-style-type: none"> - für ... (zu) lesen - zum ... (zu) lesen - um ... zu lesen - für zum ... (zu) lesen - zu-Infinitivanschluss ohne explizite Nebensatzeinleitung - finiter Nebensatz mit dass bzw. damit 	Finalanschluss
I.12 Sie geben Ihrer Tochter den Tipp, die Fischstäbchen gefroren anzubraten. (Ankreuzen)	<ul style="list-style-type: none"> - (ase) gefroren - (ase) gefroren-e - (ase) gefroren-er 	Koprädikatives Partizip/Adjektiv, prädikative Adjektiv-Verb-Fügung

I.13 <i>Es wird gearbeitet (Ankreuzen)</i>	- wird es gearbeitet - wird gearbeitet	Expletives <i>es</i>
I.14 <i>Frage weggelassen</i>		
I.15 <i>Habe ich dir das nicht bereits erzählt? (Ankreuzen)</i>	- han-i-der daas - ha-der daas - han-der dass - han ICH dir	Auslassung Pronomen du/ ich
I.16 <i>Davon habe ich auch schon gehört (Ankreuzen)</i>	- dadevoo han ich - davoo han ich - vo däm han ich - da han ich au scho devoo	Pronominalad- verbien (Doppe- lung, Stranding, Stellung)
I.17 <i>Familie Bertingers Kind ist krank auf die Schulreise mitgegangen. Maria empört sich: (Ankreuzen)</i>	- chrank (unflektiert) - chrankä, chranknä (erstarrt) - chranks (Kongruenz Neutrum) - chranki (Kongruenz Fem.)	Koprädikatives Partizip/Adjektiv, prädikative Adjektiv-Verb- Fügung
I.18 <i>Frau Terchert sagt zu ihrem Mann: „Wir haben keine Milch mehr.“ Der Mann sagt: (Ankreuzen)</i>	- go chaufe - re go chaufe	Partitiv
I.19 <i>Bruno und Peter unterhalten sich über das sündhaft teure Auto, das sich Brunos Schwester angeschafft hat. Bruno sagt zu Peter: (Ankreuzen)</i>	- bezahlt hat - hat bezahlt	Position der Verbalteile im Nebensatz
I.20 <i>Ihre Kollegin hat das Buch, das Sie ihr zum Geburtstag geschenkt hatten, ungelesen an ihre Mutter weitergegeben. Sie ärgern sich: (Ankreuzen)</i>	- dir geschenkt - xy dir geschenkt	Präpositionale Dativmarkierung (PDM)

2. Fragebogen	Einbezogene, klassierte Varianten	Phänomen
II.1 <i>Hast du die Uhr flicken lassen? (Übersetzung)</i>	- lassen flicken - flicken lassen - la flicken lassen / la lassen flicken	<i>lassen</i> -Verdoppe- lung und Stellung der Verbalteile
II.2a <i>Das ist doch die Frau, der ich schon lange das Buch bringen sollte (Übersetzung)</i>	- woni (mit/ ohne Resumptivpronomen) - (a/ i) dere woni - (a/ i) dere i/ ich/ ig	Relativsatz
II.2b <i>Das ist doch die Frau, der ich schon lange das Buch bringen sollte (Übersetzung)</i>	- bringen sollte - sollte bringen	Position der Verbalteile im Nebensatz
II.3 <i>Er lässt den Schreiner kommen. (Übersetzung)</i>	- lässt kommen - lässt (la) ... (la) kommen	<i>lassen</i> - Verdoppelung und Stellung der Verbalteile
II.4 <i>Du hast sicher viel zu erzählen (Übersetzung)</i>	- z verzelle - zum verzelle - z verzellid	Gerundium
II.5 <i>Ihr dürft alles liegen lassen. (Übersetzung)</i>	Modalverb - ligge la - la ligge - la liegen la Imperativ - lönd ... la ligge - lönd ... ligge	<i>lassen</i> - Verdoppelung und Stellung der Verbalteile
II.6 <i>Bruno Z. erklärt, warum er zu spät in die Gemeinderatssitzung gekommen ist: S Telefon hät grad glüütet, ... (Ankreuzen)</i>	- han welle gaa - han gaa welle - gaa han welle	Position der Verbalteile im Nebensatz

II.7 Erna wohnt zwar sehr abgelegen, hatte aber dennoch lange Zeit keinen Führerschein. Sie erklärt: (Ankreuzen)	- gelernt fahren / fahren gelernt - lernen fahren / fahren lernen	Ersatzinfinitiv und Position der Verbalteile
II.8 Sie treffen beim Einkaufen Ihren Nachbarn und schwatzen mit ihm über den Wetterbericht. Sie meint: (Ankreuzen)	- no öppis Schnee - no e chli Schnee - no chli Schnee - ø Schnee	Partitiv
II.9 Hanna meint, die Villa stehe immer noch zum Verkauf. Fredy aber informiert sie: (Ankreuzen)	- verkauft worden - verkauft kommen - verkauft-e kommen - verkauft-e werden - Zustandspassiv	Passiv
II.10 Ihr Nachbar lobt, wie schön gepflegt Ihr Garten doch sei. Bescheiden sagen Sie: (Ankreuzen)	- de vil de - vil de - de vil	Artikelverdoppelung
II.11 Ihr Bruder hat sich bei der Gartenarbeit die Hand verletzt. Als Frau Terchert Sie fragt, wie es ihm gehe, sagen Sie: (Ankreuzen)	- eingebunden - eingebunden-e - eingebunden-er	Resultativ
II.12 Frau Terchert befürchtet, dass ihre Nachbarin die Katze heimlich füttert. Als sie die Nachbarin darauf anspricht, weist diese den Vorwurf entschieden zurück: (Ankreuzen)	- euere - xy euere	Präpositionale Dativmarkierung (PDM)
II.13 Marianne rät ihrer Nachbarin bei Schlafproblemen, vor dem Schlafengehen Milch mit Honig zu trinken. Ausserdem betont sie: (Ankreuzen)	- (ase) heiss - (ase) heiss-e - (ase) heiss-er	Koprädikatives Partizip/Adjektiv, prädikative Adjektiv-Verb-Fügung
II.14 Frau Terchert erzählt über Hans, der sie gestern beim Essen gestört hat: (Ankreuzen)	- ässe han welle - ässe welle han - welle ässe han - han welle ässe	Position der Verbalteile im Nebensatz
II.15 Thomas will dem Lehrer zusammen mit Jürg einen Streich spielen. Jürg aber hat Gewissensbisse. Er sagt: (Ankreuzen)	- da wott i nüt mit z tue ha - da wott i nüt z tue ha mit - dademit wott i nüt z tue ha - da wott i (nüt) demit (nüt) z tue ha - da wott i nüt z tue ha demit - mit däm wott i nüt z tue ha - da(r)mit wott i nüt z tue ha	Gerundium
II.16 Sie sind gerade dabei ein Bild aufzuhängen. Eine Kollegin hört das Hämmern, klopft an und fragt verdutzt, was hier vorgeht. Sie erklären: (Ankreuzen)	- grad am e(s) Bild uf(z)hänke - grad e(s) Bild am ufhänke - grad draa e(s) Bild ufZhänke - grad ame(ne) Bild uf(z)hänke	Progressivkonstruktionen
II.17 Nach einem Hochzeitsfest von Freunden erzählt Hanna ihrer Nachbarin ausführlich, wie ungepflegt die Kinder ihrer Cousine in die Kirche gekommen sind. Hanna empört sich: (Ankreuzen)	- (ase) ungekämmt - (ase) ungekämmt-e - (ase) ungekämmt-er - de ungsräältä Wäg	Koprädikatives Partizip/Adjektiv, prädikative Adjektiv-Verb-Fügung
II.18 Ein Mann spricht Sie in der Stadt an und bedankt sich bei Ihnen für etwas. Ihre Freundin fragt Sie, woher er Sie kennt. Sie sagen: (Ankreuzen)	- woni (mit/ ohne Resumptivpronomen) - (a/ i) dem woni - (a/ i) dem i/ ich	Relativsatz
II.19 Sie waren im Wald Würste braten. Sie sind wirklich erschrocken, als sich plötzlich ein Fuchs aus dem Unterholz heranschlich. Am nächsten Tag erzählen Sie davon: (Ankreuzen)	- (z) schliiche cho/ cho (z) schliiche - (z) schleich-end(e) - z schleich-id gekommen - z schleich-ete gekommen - gschliche cho	Infinitivanschluss

<i>fII.20 Sie sind beim Autohändler. Er will Ihnen einen Porsche aufschwätzen. Sie winken ab: (Ankreuzen)</i>	- woni - das - das woni	Relativsatz
<i>fII.21 Martin fragt Susi etwas, aber wegen der lauten Musik versteht Susi kein Wort. Lisa, die verstanden hat, erklärt Susi: (Ankreuzen)</i>	- ob ø - ob d - ob de - ob du	Ausfall der 2. Sg.
<i>fII.22 Frau Terchert fragt Kevin ob das denn sein neues Dreirad sei. Kevin antwortet: (Ankreuzen)</i>	- s Peters - em Peter - em Peter siis - (das) vom Peter - em Peters - ghört em Peter	Possessivkonstruktionen
<i>fII.23 Dann fällt Kevin ein, dass das Dreirad Sandra gehört. Er sagt: (Ankreuzen)</i>	- s Sandras - de Sandra - de Sandra ires - vo de Sandra - em Sandra - de Sandra sis - ghört de Sandra - ghört em Sandra - daas vo de Sandra	Possessivkonstruktionen
<i>fII.24 Annas Mann kann überraschend nicht zum Fondue kommen. Anna sagt enttäuscht: (Ankreuzen)</i>	- grad am s Brot (z) schniide - grad s Brot am schniide / am Brot schniide / am Brot Z schniide	Progressivkonstruktionen
<i>fII.25 Gut habt ihr eine so grosse Wohnung: (Ankreuzen)</i>	- händ ir - dass ir - as ir	Anschlusslose (invertierte) Ergänzungssätze
<i>fII.26 Vreni hat an einer Party einen netten jungen Mann kennen gelernt. Sie fragt sich, ob er sich wohl bei ihr melden wird. Ihre Freundin Marianne meint dazu: (Ankreuzen)</i>	- wo ø - wo d - wo dass - wo dass d - wo du - wo de - wo dass du - wo dass de	Doppelte/erweiterte Konjunktionen / Auslassung Pronomen du/ ich
<i>fII.27 Erstaunt fragen Sie, warum ihr Kollege denn seinen Schuh ausziehe. Er antwortet: (Ankreuzen)</i>	- wil er nass isch - as er nass isch - dass er nass isch	Koprädikatives Partizip/Adjektiv, prädikative Adjektiv-Verb-Fügung
<i>fII.28 Sie erklären Ihrem Kollegen, wer der Mann ist, der ihnen zuwinkt: (Ankreuzen)</i>	- woni (mit/ ohne Resumptivpronomen) - mit dem woni/ich - mit dem i/ich/ig	Relativsatz
<i>fII.29 Peter O. behauptet, es brauche gar keine neue Turnhalle. Bruno Z. erhebt sich und entgegnet aufgebracht: (Ankreuzen)</i>	- eifach go z säge - eifach go säge - eifach z säge - eifach säge - eifach cho z'säge	Go-Konstruktionen
<i>fII.30 Wie heisst es in Ihrem Dialekt für „der Hund des Lehrers“? (Ankreuzen)</i>	- s Leerers Hund - em Leerer sin Hund - de Hund vom Leerer - em Leerers Hund	Possessivkonstruktionen
<i>fII.31 Wie heisst es in Ihrem Dialekt für „Ich habe Anna gesehen“? (Ankreuzen)</i>	- die Anna - die Anna + Flexionssuffix - Anna - das Anni - Anni + Flexionssuffix	Artikel bei Eigennamen

<i>fII.32 Wie heisst es in Ihrem Dialekt für „Ich habe Fritz gesehen“? (Ankreuzen)</i>	<ul style="list-style-type: none"> - den Fritz - den Fritz + Flexionssuffix - Fritz - Fritz + Flexionssuffix 	Artikel bei Eigennamen
--	--	------------------------

3. Fragebogen	Einbezogene, klassierte Varianten	Phänomen
<i>fIII.1 Seit Wochen ist der See gefroren und man kann Eislaufen gehen. Allerdings war es in den letzten Tagen nicht mehr besonders kalt. Sie sagen: (Übersetzung)</i>	<ul style="list-style-type: none"> - fängt s Iis aa (z) schmelze - fängt s IIs afa (z) schmelze - fängt slis ane schmelze - tuet s Iis afa (z) schmelze - wird s Iis (afa) schmelze - tuet s Iis schmelze - schmilzt s Iis 	<i>anfangen</i>
<i>fIII.2 Ihre Freundin dreht sich mehrfach suchend um: (Übersetzung)</i>	<ul style="list-style-type: none"> - wer - wem - wen - we - wels - genusunspezifische Formen 	Interrogativpronomen
<i>fIII.3 Marianne kommt mit einem Blumenstrauss nach Hause. Ihre Mutter fragt erstaunt: (Übersetzung)</i>	<ul style="list-style-type: none"> - für wer - für wem - für wen - für we - für wels - für welum - genusunspezifische Formen 	Interrogativpronomen
<i>fIII.4 Doch Marianne winkt ab: (Übersetzung)</i>	<ul style="list-style-type: none"> - für dich - für dir 	Personalpronomen/ Indefinitpronomen
<i>fIII.5 Max erwartet Besuch. Als es an der Tür klingelt, erklärt er, warum er die Schürze umgebunden hat: (Übersetzung)</i>	<ul style="list-style-type: none"> - agfange (z/ zum) choche - afa choche - agfange afa choche - afange choche - agfange ane choche - agfange mit choche / mit choche agfange/ mit choche ... agfange 	<i>anfangen</i>
<i>fIII.6 Frage weggelassen</i>		
<i>fIII.7 Sie informieren Ihre Nachbarin darüber, dass Susi ein Kind erwartet. Die Nachbarin weiss bereits Bescheid. Sie sagt: (Übersetzung)</i>	<ul style="list-style-type: none"> - mir es - es mir - s mirs - mir das - das mir 	Stellung der Pronomen
<i>fIII.8 Thomas hat in der Rekrutenschule angefangen zu rauchen. Seine Mutter ist darüber alles andere als begeistert. Er erzählt davon einem Kollegen: (Übersetzung)</i>	<ul style="list-style-type: none"> - han agfange (z/ zum) rauche - han afa rauche - han agfange afa rauche - han agfange ane rauche - agfange (mit) rauchen - agfange (z/ mit) rauche 	<i>anfangen</i>
<i>fIII.9 Ihr Nachbar bittet Sie, ihm einen Regenschirm auszuleihen. Sie mahnen ihn: (Übersetzung)</i>	<ul style="list-style-type: none"> - muesch mer en - muesch en mer - muesch mer de(r) - muesch de(r) mer 	Stellung der Pronomen
<i>fIII.10 Frage weggelassen</i>		
<i>fIII.11 Mich erwischt keiner: (Übersetzung)</i>	<ul style="list-style-type: none"> - mich - mir 	Personalpronomen/ Indefinitpronomen

<i>fIII.12 Hans will eine Fertigsuppe zubereiten. Dabei hält er sich genau an die Anleitung auf der Packung. Doch die Suppe läuft über. Marianne gibt ihm folgenden Tipp: (Übersetzung)</i>	<ul style="list-style-type: none"> - afangt (z/ zum) choche - faat afa choche - afangt afa choche - faat a choche - afangt ane choche - tun - mit choche afangt 	<i>anfangen</i>
<i>fIII.13 Er gibt sich keine Mühe. (Ankreuzen)</i>	<ul style="list-style-type: none"> - em - sich 	Reflexiv-konstruktion
<i>fIII.14 Fritz wundert sich, dass seine Frau nochmals in den Laden gehen will. Sie antwortet ihm: (Ankreuzen)</i>	<ul style="list-style-type: none"> - käme Brot - kä Brot me 	Verschiedenes
<i>fIII.15 Ein Freund bricht zu einer zweitägigen Bergtour auf. Sie raten ihm, Handschuhe mitzunehmen, denn: (Ankreuzen)</i>	<ul style="list-style-type: none"> - cha s/es - cha - isch s/es - isch 	Expletives <i>es</i>
<i>fIII.16 Sie erklären, warum man bei ihnen schlecht parkieren kann: (Ankreuzen)</i>	<ul style="list-style-type: none"> - aufgerissen - aufgerissen-e 	Resultativ
<i>fIII.17 Sie können sich immer noch nicht zum Kauf einer neuen Polstergruppe entscheiden. Sie sagen: (Ankreuzen)</i>	<ul style="list-style-type: none"> - uns - sich 	Reflexiv-konstruktion
<i>fIII.18 Eine Mutter empört sich über ihren Sohn, der in der Freizeit nur Videospiele macht: (Ankreuzen)</i>	<ul style="list-style-type: none"> - käs Buech nid - käs Buech - nid es Buech 	Doppelte Negation
<i>fIII.19 Die Mutter hängt im Garten zwei Badetücher zum Trocknen auf. Aber das Kind spritzt dort mit der Wasserpistole herum. Die Mutter ruft dem Kind zu: (Ankreuzen)</i>	<ul style="list-style-type: none"> - nass - nassi - netze 	Koprädikatives Partizip/Adjektiv, prädikative Adjektiv-Verb-Fügung
<i>fIII.20 Markus will in Amerika eine sehr gute Stellung als Top-Manager annehmen. Seine Frau will wegen der Kinder aber nicht wegziehen. Gegenüber ihrer besten Freundin behauptet sie: (Ankreuzen)</i>	<ul style="list-style-type: none"> - ihn selbst - sich selbst - ihm selbst - er selbst - e selbst - ihn - sich - ihm - schim sälber 	Reflexiv-konstruktion
<i>fIII.21 Die Grossmutter erklärt ihrem Enkel, warum sie erst mit fünfzig Jahren zum ersten Mal in die Ferien geflogen ist: (Ankreuzen)</i>	<ul style="list-style-type: none"> - niemer kä Gält - niemer Gält - me kä Gält 	Doppelte Negation
<i>fIII.22 Anneli sagt über ihre Schwester: (Ankreuzen)</i>	<ul style="list-style-type: none"> - als - weder - wie - wan 	Komparativ
<i>fIII.23 Sie haben gesehen, wie Hans auf dem glatten Eis auf dem Heimweg gestürzt ist. Aber er will sich nicht helfen lassen. Empört erzählen Sie Ihrer Familie: (Ankreuzen)</i>	<ul style="list-style-type: none"> - hinkend - hinkend + Vokal - g(h)inket - g(h)inkete - (g)hinkig - hinkege/ (g)hinklige - hinkends/ ghinkäds/ ds hinkete/ ds hinkedse 	<i>hinkend</i> als Konverb
<i>fIII.24 Susi erklärt, warum ihr kranker Grossvater keine Hauspflege beanspruchen will: (Ankreuzen)</i>	<ul style="list-style-type: none"> - öpper fremder - öpper fremds - öpper fremd 	Flexion des Adjektivs nach Indefinitpronomen

<i>fIII.25 Maria und Hanna wollen nicht mit auf den Sonntagsspaziergang. Der Vater sagt zur Mutter: (Ankreuzen)</i>	<ul style="list-style-type: none"> - als - wie - weder - (an)statt - wan 	Komparativ
<i>fIII.26 Als Sie beim Nachbarn eintreten, hören Sie aus dem Badezimmer, wie der kleine Kevin schreit. Der Nachbar erklärt: (Ankreuzen)</i>	<ul style="list-style-type: none"> - chunnt, bechunnt, kriegt, überchunnt, bechunnt über - werden gewaschen - kommen gewaschen 	Passiv
<i>fIII.27 Sie erzählen Ihrem Freund, was Sie auf dem Bild sehen: (Ankreuzen)</i>	<ul style="list-style-type: none"> - hinter e Puess (z) schriibe - dehinter e Puess (z) schriibe - am e Puess (z) schriibe - e Puess am schriibe - anere Puess (z) schriibe - am ø Puess schriibe 	Progressivkonstruktionen
<i>fIII.28 Verena fragt ihre Mutter, wie alt denn der Gemeinderat Bruno O. sei. Die Mutter antwortet, er sei schon über 50. Darauf sagt Verena erstaunt: (Ankreuzen)</i>	<ul style="list-style-type: none"> - weder - als/as - wie - weder as - weder als - weder dass - als dass - wan - wan dass - dass 	Komparativ

4. Fragebogen	Einbezogene, klassierte Varianten	Phänomen
<i>IV.1 Martina wäre eine ganz gute Gemeindepräsidentin. (Übersetzung)</i>	<ul style="list-style-type: none"> - e ganz e - e ganz - ganz e 	Artikelverdoppelung
<i>IV.2 Die können einen doch nicht so behandeln. (Übersetzung)</i>	<ul style="list-style-type: none"> - einem - ein(en) - eines 	Personalpronomen/ Indefinitpronomen
<i>IV.3 Ich habe es ihm schon geschickt. (Übersetzen)</i>	<ul style="list-style-type: none"> - ihm es - es ihm - ihm das - das dem - ihms ihm - s ihms 	Stellung der Pronomen
<i>IV.4 Wer ist das gewesen? (Übersetzen)</i>	<ul style="list-style-type: none"> - wer - we - wels 	Interrogativpronomen
<i>IV.5 Wir gehen Fussball spielen. (Übersetzung)</i>	<ul style="list-style-type: none"> - go tshutte - go z tshutte 	Go-Konstruktionen
<i>IV.6 Frage weggelassen</i>		
<i>IV.7 Jetzt kannst du anfangen. (Übersetzung)</i>	<ul style="list-style-type: none"> - afange - a(a)faa, a(a)fa, afan, afau 	<i>anfangen</i>
<i>IV.8 Ich habe es ihr gestern gegeben. (Übersetzung)</i>	<ul style="list-style-type: none"> - ihr-es - es-ihr 	Stellung der Pronomen
<i>IV.9 Marianne sitzt im Zug. Der Kontrolleur will das Billet sehen. Sie erschrickt und sagt: (Ankreuzen)</i>	<ul style="list-style-type: none"> - z löse - ø löse - zum ... löse 	Finalanschluss
<i>IV.10 Kurz vor einer Kreuzung fragt M. seine Beifahrerin, ob er abbiegen müsse. Sie antwortet: (Ankreuzen)</i>	<ul style="list-style-type: none"> - ø luege - go luege 	Go-Konstruktionen
<i>IV.11 Kevins Mutter will nicht glauben, dass dieser Annas Briefkasten gesprengt hat. Anna lässt sich nicht abbringen: (Ankreuzen)</i>	<ul style="list-style-type: none"> - er - ihn - ihm 	Personalpronomen/ Indefinitpronomen

IV.12 Marianne sagt zu ihrer Tocher, sie solle spielen gehen, bis ihr Vater nach Hause komme: (Ankreuzen)	- bis dass de - bis de - bis wa de	Doppelte/ erweiterte Konjunktionen
IV.13 Ihre Grossmutter bietet ihnen Kuchen an. Sie lehnen dankend ab: (Ankreuzen)	- welle wott i - welle tuen i - welle tät i - welle würi - welle möcht i	Verb- verdoppelung bei Betonung
IV.14 Der Grossvater kommt in die Stube, wo die Grossmutter bei Dämmerlicht liest. Er sagt zu ihr: (Ankreuzen)	- zum (zu) lesen - für (zu) lesen - um zu lesen - für zum zu lesen - für das Lesen	Finalanschluss
IV.15 Bruno ist mit Anita am Telefon. Diese will auch noch mit seiner Frau sprechen. Bruno ruft sie: (Ankreuzen)	- ob ø - ob d - ob du - ob de	Auslassung Pronomen du/ ich
IV.16 Frage weggelassen		
IV.17 Ihre Freundin glaubt nicht, dass vorhin Alt-Bundesrat Ogi an Ihnen vorbei gelaufen ist. Sie finden aber: (Ankreuzen)	- er - en - e	Personal- pronomen/ Inde- finitpronomen
IV.18 Sandra gesteht der Klavierlehrerin: (Ankreuzen)	- z üebe - ø üebe - zum üebe - ø üebe vergässe - z üebe vergässe	Finalanschluss
IV.19 Frau Terchert fragt Sie, ob Sie schon ein Geschenk für Alex haben. Sie antworten: (Ankreuzen)	- öppis ganz öppis - öppis ganz - ganz öppis	Artikel- verdoppelung
IV.20 Auf Ihrem Tisch stapeln sich die Unterlagen. Sie seufzen: (Ankreuzen)	- es weiss ein nid, wo... - me weis nid, wo...	anfangen
IV.21 Sie erzählen Ihrem Nachbarn, dass Hans erst mit 50 Jahren den Fahrausweis gemacht hat. Der Nachbar sagt erstaunt: (Ankreuzen)	- dass ... fahren gelernt hat - dass ... hat lernen fahren	Ersatzinfinitiv und Position der Verbalteile
IV.22 Lukas sitzt seit zwei Stunden an seinen Hausaufgaben. Der Grossvater lobt ihn: (Ankreuzen)	- bisch scho... - sii bisch... - du bisch (de)... - bisch du... - si tuesch...	Verb- verdoppelung bei Betonung
IV.23 Frage weggelassen		
IV.24 Sie haben dem Grossvater erzählt, dass jemand Kevin beim Basteln geholfen hat. Der Grossvater erinnert sich nicht mehr und fragt nach: (Ankreuzen)	- wer ... dass em - wer ... dass er em - wer ... er hät em - wer ... ø hät em - wer ... wer em - wer ... wer dass em - wer ... wer hät em - wer ... wo em - was ... wer em - was ... wer hät em - wer ... der em	Verschränkte Wh-Sätze
IV.25 Peter behauptet, dass seine Tochter schon mit 4 Jahren lesen konnte. Sie zweifeln daran und sagen: (Ankreuzen)	- dass ... lesen gelernt hat - dass ... hat lernen lesen	Ersatzinfinitiv und Position der Verbalteile

IV.26 Der Pfarrer hat mit einer Kollekte eine bedürftige Familie in der Gemeinde unterstützt. Sie erkundigen sich nochmals bei Erika, die Ihnen das erzählt hat: (Ankreuzen)	<ul style="list-style-type: none"> - wem ... dass de - wem ... dass im de - wem ... hät de - wem ... de ... hät - was ... wem dass de - was ... wem de - was ... wem hät - wem ... wo de - wer ... dass de - wer ... wäm de 	Verschränkte Wh-Sätze
IV.27 Ich habe es Anna gegeben. (Ankreuzen)	<ul style="list-style-type: none"> - Anna - der Anna - der Anna + Flexionssuffix - der Anni - (d)em Anna/ Anni - Annin 	Artikel bei Eigennamen
IV.28 Ich habe es Fritz gegeben. (Ankreuzen)	<ul style="list-style-type: none"> - dem Fritz - dem Fritz + Flexionssuffix - Fritz - Fritz + Flexionssuffix 	Artikel bei Eigennamen
IV.29 Basil will Jasmin ihr neues Dreirad wegnehmen. Sie wehrt sich: (Ankreuzen)	<ul style="list-style-type: none"> - das isch miir - das isch miis - das isch mii - das isch mine - das isch min 	Personalpronomen/ Indefinitpronomen
IV.30 Frage weggelassen		
IV.31 Erna und Verena treffen sich im Bus und schwatzen über die neue Nachbarin, die sich die Haare grün gefärbt hat. Erna empört sich, aber Verena findet: (Ankreuzen)	<ul style="list-style-type: none"> - gfallt tät mir au - gfiel mer au - gfallti mir au - täti mir au gfallt - würd mir au gfallt 	Präpositionale Dativmarkierung (PDM)
IV.32 Sie machen mit Ihrem Freund eine Bergtour. Ihr Freund ist nicht besonders gut trainiert und seufzt schon nach der ersten Steigung: (Ankreuzen)	<ul style="list-style-type: none"> - macht mi z schwitze - macht mir z schwitze - kommen + z/ zum/ ins schwitze - bringen + z/ zum/ ins - git mir z schwitze - tuet mi z schwitze - tuet mir tue z schwitze 	Personalpronomen/ Indefinitpronomen
IV.33 Hans fragt, wem das Velo vor dem Haus der Nachbarin gehöre. Seine Frau erklärt ihm: (Ankreuzen)	<ul style="list-style-type: none"> - dr Lehrerin (fIV33.1) - dr Lehrerin ihres (fIV33.3) - dr Lehreris (fIV33.3a) - ghört dr Lehrerin (fIV33.2) 	Possessivkonstruktionen
IV.34 Frage weggelassen		
IV.35 Frage weggelassen		
IV.36 Frage weggelassen		
IV.37 Frage weggelassen		
IV.38 Kann man in Ihrem Dialekt Männer/ Frauen mit <u>es</u> bezeichnen?	<ul style="list-style-type: none"> - Es ist 71 [Trudi] - Es ist 71 [Paul] 	es für Frauen und Männer

C. Liste mit der Zuordnung der SADS- zu den SDS-Orten

C.1 Berücksichtigte 378 Ortspunkte¹⁴

Gemeindenamen SDS	BFS_Code SDS	Gemeindename SADS	BFS_Nr SADS
Aarau (AG)	4001	Aarau (AG)	4001_1
Aarberg (BE)	301	Aarberg (BE)	301_1
Aarburg (AG)	4271	Aarburg (AG)	4271_1
Abländschen (BE)	843	Saanen (BE)	843_2
Adelboden (BE)	561	Adelboden (BE)	561_1
Aesch (LU)	1021	Aesch (LU)	1021_1
Aesch (BA)	2761	Aesch (BL)	2761_1
Aeschi (SO)	2511	Aeschi (SO)	2511_1
Agarn (WS)	6101	Agarn (VS)	6101_1
Alpnach (UW)	1401	Alpnach (OW)	1401_1
Alpthal (SZ)	1361	Alpthal (SZ)	1361_1
Altdorf (UR)	1201	Altdorf (UR)	1201_1
Altstätten (SG)	3251	Altstätten (SG)	3251_1
Amden (SG)	3311	Amden (SG)	3311_1
Amriswil (TG)	4461	Amriswil (TG)	4461_1
Andwil (SG)	3441	Andwil (SG)	3441_1
Appenzell (AP)	3101	Appenzell (AI)	3101_1
Arlesheim (BL)	2763	Muttenz (BL)	2770_1
<i>Auen (GL)</i>	<i>1613</i>	<i>Linthal (GL)</i>	<i>1613_1</i>
Ausserberg (WS)	6191	Ausserberg (VS)	6191_1
Avers (GR)	3681	Avers (GR)	3681_1
<i>Azmoos (SG)</i>	<i>3276</i>	<i>Wartau (SG)</i>	<i>3276_1</i>
Baldingen (AG)	4301	Baldingen (AG)	4301_1
Bäretswil (ZH)	111	Bäretswil (ZH)	111_1
Basadingen (TG)	4536	Basadingen-Schlattingen (TG)	4536_1
Basel (BA)	2701	Basel (BS)	2701_2
Bassersdorf (ZH)	52	Bassersdorf (ZH)	52_1
Bauma-Saland (ZH)	171	Bauma (ZH)	171_1
Belp (BE)	861	Belp (BE)	861_1
Benken (SG)	3312	Benken (SG)	3312_1
Bern (BE)	351	Bern (BE)	351_1
Berneck (SG)	3233	Berneck (SG)	3233_1
Betten (WS)	6171	Betten (VS)	6171_1
Bettingen (BA)	2702	Bettingen (BS)	2702_1
Bettlach (SO)	2543	Bettlach (SO)	2543_1
Binn (WS)	6054	Binn (VS)	6054_1
Birmenstorf (AG)	4024	Birmenstorf (AG)	4024_1

¹⁴ **fett:** BFS_Nr abweichend, *kursiv:* Nur Ortsname abweichend

Birwinken (TG)	4901	Birwinken (TG)	4901_1
Blatten (WS)	6192	Blatten (VS)	6192_1
Bleienbach (BE)	324	Bleienbach (BE)	324_1
Boltigen (BE)	791	Boltigen (BE)	791_1
Boswil (AG)	4228	Boswil (AG)	4228_1
Bottighofen (TG)	4643	Bottighofen (TG)	4643_1
Bremgarten-Widen-Zufikon (AG)	4063	Bremgarten (AG)	4063_1
Brienz (BE)	573	Brienz (BE)	573_1
Brugg (AG)	4095	Brugg (AG)	4095_1
<i>Brülisau (AP)</i>	<i>3103</i>	<i>Rüte (AI)</i>	<i>3103_1</i>
Brunnadern (SG)	3371	Brunnadern (SG)	3371_1
Bubikon (ZH)	112	Rapperswil-Jona (SG)	3336_1
Buckten (BA)	2843	Buckten (BL)	2843_1
Bülach (ZH)	53	Bülach (ZH)	53_1
Buochs (UW)	1502	Buochs (NW)	1502_1
Bürchen (WS)	6193	Bürchen (VS)	6193_1
Büren (BE)	383	Büren an der Aare (BE)	383_1
Busswil (BE)	384	Busswil bei Büren (BE)	384_1
Chur-Masans (GR)	3901	Chur (GR)	3901_1
Churwalden (GR)	3911	Churwalden (GR)	3911_1
Conters (GR)	3881	Küblis (GR)	3882_1
Dagmersellen (LU)	1125	Dagmersellen (LU)	1125_1
Davos-Frauenkirch (GR)	3851	Davos (GR)	3851_1
Densbüren (AG)	4004	Densbüren (AG)	4004_1
Diepoldsau (SG)	3234	Diepoldsau (SG)	3234_1
Düdingen (FR)	2293	Düdingen (FR)	2293_1
Ebnat-Kappel (SG)	3352	Ebnat-Kappel (SG)	3352_1
Egg-Esslingen (ZH)	192	Egg (ZH)	192_1
Eggiwil (BE)	901	Röthenbach im Emmental (BE)	904_1
Eggiwil (BE)	901	Trub (BE)	908_2
Eglisau (ZH)	55	Eglisau (ZH)	55_1
Einsiedeln (SZ)	1301	Einsiedeln (SZ)	1301_1
Elfingen (AG)	4097	Elfingen (AG)	4097_1
Elgg (ZH)	217	Elgg (ZH)	217_1
Ellikon (ZH)	218	Ellikon an der Thur (ZH)	218_1
Elm (GL)	1605	Elm (GL)	1605_1
Embrach (ZH)	56	Rorbach (ZH)	68_1
Engelberg (UW)	1402	Engelberg (OW)	1402_1
Engi (GL)	1606	Engi (GL)	1606_1
Entlebuch (LU)	1002	Entlebuch (LU)	1002_1
Erlach (BE)	492	Erlach (BE)	492_1
Ermatingen (TG)	4646	Ermatingen (TG)	4646_1
Ernen (WS)	6056	Fiesch (VS)	6057_1
Erschwil (SO)	2615	Erschwil (SO)	2615_1
Eschenbach (SG)	3332	Eschenbach (SG)	3332_1

Eschenbach (LU)	1026	Eschenbach (LU)	1026_1
Escholzmatt (LU)	1003	Escholzmatt (LU)	1003_1
Ettingen (BA)	2768	Ettingen (BL)	2768_1
Fällanden (ZH)	193	Fällanden (ZH)	193_1
<i>Fankhaus (BE)</i>	908	<i>Trub (BE)</i>	908_1
<i>Faulensee (BE)</i>	768	<i>Spiez (BE)</i>	768_1
Ferden (WS)	6195	Ferden (VS)	6195_1
Feschel (WS)	6117	Guttet-Feschel (VS)	6117_1
Fischingen (TG)	4726	Fischingen (TG)	4726_1
Flaach (ZH)	28	Flaach (ZH)	28_1
Fläsch (GR)	3951	Fläsch (GR)	3951_1
Flawil (SG)	3402	Flawil (SG)	3402_1
Flühli (LU)	1004	Flühli (LU)	1004_1
Flums (SG)	3292	Flums (SG)	3292_1
Frauenfeld (TG)	4566	Frauenfeld (TG)	4566_2
Frauenkappelen (BE)	663	Frauenkappelen (BE)	663_1
Freiburg (FR)	2196	Fribourg (FR)	2196_1
Frick (AG)	4163	Frick (AG)	4163_1
Frutigen (BE)	563	Frutigen (BE)	563_1
Gadmen (BE)	781	Gadmen (BE)	781_1
Gais (AP)	3022	Gais (AR)	3022_1
Galgenen (SZ)	1342	Lachen (SZ)	1344_1
Gamsen (WS)	6002	Brig-Glis (VS)	6002_1
Gänsbrunnen (SO)	2423	Welschenrohr (SO)	2429_1
Gelterkinden (BA)	2846	Gelterkinden (BL)	2846_1
Giffers (FR)	2294	Giffers (FR)	2294_1
<i>Gimmelwald (BE)</i>	584	<i>Lauterbrunnen (BE)</i>	584_2
Giswil (UW)	1403	Giswil (OW)	1403_1
Glarus (GL)	1609	Glarus (GL)	1609_1
<i>Göscheneralp (UR)</i>	1208	<i>Göschenen (UR)</i>	1208_1
Grabs (SG)	3273	Grabs (SG)	3273_1
Grafenried (BE)	539	Grafenried (BE)	539_1
Grindelwald (BE)	576	Grindelwald (BE)	576_1
Grosswangen (LU)	1086	Grosswangen (LU)	1086_1
<i>Grit (ZH)</i>	115	<i>Gossau (ZH)</i>	115_1
Gsteig (BE)	841	Gsteig (BE)	841_1
Guggisberg (BE)	852	Guggisberg (BE)	852_1
<i>Gündlischwand (BE)</i>	584	<i>Lauterbrunnen (BE)</i>	584_3
Gurmels (FR)	2262	Gurmels (FR)	2262_1
Gurtellen (UR)	1209	Gurtellen (UR)	1209_1
Guttannen (BE)	782	Guttannen (BE)	782_1
Habkern (BE)	579	Habkern (BE)	579_1
Habstetten (BE)	352	Ittigen (BE)	362_1
Hägglingen (AG)	4068	Hägglingen (AG)	4068_1
Hallau (SH)	2971	Hallau (SH)	2971_1

Haslen (AP)	3104	Schlatt-Haslen (AI)	3104_1
Hedingen (ZH)	5	Hedingen (ZH)	5_1
Heiden (AP)	3032	Heiden (AR)	3032_1
Heimberg (BE)	928	Steffisburg (BE)	939_1
Heitenried (FR)	2296	Heitenried (FR)	2296_1
Hemberg (SG)	3372	Wattwil (SG)	3377_2
Herbetswil (SO)	2424	Aedermannsdorf (SO)	2421_1
Herisau (AP)	3001	Herisau (AR)	3001_1
Hölstein (BA)	2886	Hölstein (BL)	2886_1
Homburg (TG)	4816	Homburg (TG)	4816_1
Horw (LU)	1058	Horw (LU)	1058_1
Hospental (UR)	1210	Andermatt (UR)	1202_1
Hünenberg (ZG)	1703	Hünenberg (ZG)	1703_1
Hütten (ZH)	134	Hütten (ZH)	134_1
Huttwil (BE)	954	Huttwil (BE)	954_1
Hüttwilen (TG)	4821	Hüttwilen (TG)	4821_1
Inden (WS)	6109	Inden (VS)	6109_1
Innerthal (SZ)	1343	Innerthal (SZ)	1343_1
Innertkirchen (BE)	784	Innertkirchen (BE)	784_1
Ins (BE)	496	Ins (BE)	496_1
Iseltwald (BE)	582	Iseltwald (BE)	582_1
Isenthal (UR)	1211	Isenthal (UR)	1211_1
Jaun (FR)	2138	Jaun (FR)	2138_1
Jenins (GR)	3952	Jenins (GR)	3952_1
Kaiserstuhl-Fisibach (AG)	4308	Kaiserstuhl (AG)	4308_1
Kaisten (AG)	4169	Kaisten (AG)	4169_1
Kandersteg (BE)	565	Kandersteg (BE)	565_1
Kappel (ZH)	6	Kappel am Albis (ZH)	6_1
Kesswil (TG)	4426	Kesswil (TG)	4426_1
Kiental (BE)	567	Reichenbach im Kandertal (BE)	567_1
Kilchberg (ZH)	135	Thalwil (ZH)	141_1
Kirchberg (SG)	3392	Kirchberg (SG)	3392_1
Kirchleerau (AG)	4275	Kirchleerau (AG)	4275_1
Kleinandelfingen (ZH)	33	Andelfingen (ZH)	30_1
Kleinlützel (SO)	2619	Kleinlützel (SO)	2619_1
Klosters (GR)	3871	Klosters-Serneus (GR)	3871_1
Konolfingen (BE)	612	Konolfingen (BE)	612_1
Krauchthal (BE)	414	Krauchthal (BE)	414_1
Krinau (SG)	3373	Krinau (SG)	3373_1
Küsnacht (ZH)	154	Küsnacht (ZH)	154_1
Küssnacht (SZ)	1331	Küssnacht (SZ)	1331_1
Langenbruck (BA)	2888	Langenbruck (BL)	2888_1
Langenthal (BE)	329	Langenthal (BE)	329_1
Langnau (ZH)	136	Langnau am Albis (ZH)	136_1
Langnau (BE)	902	Langnau im Emmental (BE)	902_1

Langwies (GR)	3924	Langwies (GR)	3924_1
Lauerz (SZ)	1365	Brunnen (SZ)	1364_1
Laufen (BE)	2787	Laufen (BL)	2787_1
Laupen (BE)	667	Laupen (BE)	667_1
Lauterbrunnen (BE)	584	Lauterbrunnen (BE)	584_1
Leibstadt (AG)	4311	Leibstadt (AG)	4311_1
Leissigen (BE)	585	Leissigen (BE)	585_1
Lenzburg (AG)	4201	Lenzburg (AG)	4201_1
Liesberg (BE)	2788	Liesberg (BL)	2788_1
Liestal (BA)	2829	Liestal (BL)	2829_1
Ligerz (BE)	740	Ligerz (BE)	740_1
Lohn (SH)	2917	Bibern (SH)	2913_1
Luchsingen (GL)	1614	Luchsingen (GL)	1614_2
Lungern (UW)	1405	Lungern (OW)	1405_1
Lupfig (AG)	4104	Lupfig (AG)	4104_1
<i>Lustdorf (TG)</i>	<i>4611</i>	<i>Thundorf (TG)</i>	<i>4611_1</i>
Lützelflüh (BE)	955	Lützelflüh (BE)	955_1
Luzern (LU)	1061	Luzern (LU)	1061_1
Magden (AG)	4253	Magden (AG)	4253_1
Maisprach (BA)	2853	Maisprach (BL)	2853_1
Malans (GR)	3954	Malans (GR)	3954_1
Malters (LU)	1062	Malters (LU)	1062_1
Mammern (TG)	4826	Mammern (TG)	4826_1
Marbach (LU)	1006	Marbach (LU)	1006_1
Marthalen (ZH)	35	Marthalen (ZH)	35_1
Maur (ZH)	195	Maur (ZH)	195_1
Meikirch (BE)	307	Meikirch (BE)	307_1
Meilen (ZH)	156	Meilen (ZH)	156_1
Meiringen (BE)	785	Meiringen (BE)	785_1
Melchnau (BE)	333	Melchnau (BE)	333_1
<i>Melchtal (UW)</i>	<i>1404</i>	<i>Kerns (OW)</i>	<i>1404_1</i>
Mels (SG)	3293	Mels (SG)	3293_1
Menziken (AG)	4139	Menziken (AG)	4139_1
Menzingen (ZG)	1704	Menzingen (ZG)	1704_1
Merenschwand (AG)	4234	Merenschwand (AG)	4234_1
Merishausen (SH)	2936	Merishausen (SH)	2936_1
Merlingen (BE)	938	Spiez (BE)	768_2
Messen (SO)	2457	Diessbach bei Büren (BE)	385_1
Mettmenstetten (ZH)	9	Rifferswil (ZH)	12_1
Mogelsberg (SG)	3406	Degersheim (SG)	3401_1
Möhlin (AG)	4254	Möhlin (AG)	4254_1
Mollis (GL)	1617	Mollis (GL)	1617_1
Mörel (WS)	6179	Mörel (VS)	6179_1
Mörschwil (SG)	3214	Mörschwil (SG)	3214_1
Mosnang (SG)	3394	Mosnang (SG)	3394_1

Muhen (AG)	4009	Muhen (AG)	4009_1
Mühledorf (SO)	2458	Bibern (SO)	2444_1
Mümliswil (SO)	2428	Mümliswil-Ramiswil (SO)	2428_1
Münchenbuchsee (BE)	546	Münchenbuchsee (BE)	546_1
Muotathal (SZ)	1367	Muotathal (SZ)	1367_1
<i>Murg (SG)</i>	<i>3295</i>	<i>Quarten (SG)</i>	<i>3295_1</i>
Murten (FR)	2275	Murten (FR)	2275_1
Mutten (GR)	3503	Mutten (GR)	3503_1
Näfels (GL)	1619	Näfels (GL)	1619_1
Neftenbach (ZH)	223	Neftenbach (ZH)	223_1
Neudorf (LU)	1092	Römerswil (LU)	1039_1
Neuenegg (BE)	670	Neuenegg (BE)	670_1
Neuenkirch (LU)	1093	Neuenkirch (LU)	1093_1
Niederbipp (BE)	981	Niederbipp (BE)	981_1
Niedergestein (WS)	6198	Steg (VS)	6200_1
Niederrohrdorf (AG)	4035	Niederrohrdorf (AG)	4035_1
Niederwichtrach (BE)	632	Wichtrach (BE)	621_1
Nunningen (SO)	2621	Nunningen (SO)	2621_1
Oberägeri (ZG)	1706	Oberägeri (ZG)	1706_1
Oberhof (AG)	4173	Oberhof (AG)	4173_1
Oberiberg (SZ)	1368	Oberiberg (SZ)	1368_1
Obermumpf (AG)	4256	Stein (AG)	4260_1
<i>Oberried (BE)</i>	<i>792</i>	<i>Lenk (BE)</i>	<i>792_1</i>
Oberriet (SG)	3254	Oberriet (SG)	3254_1
Obersaxen (GR)	3612	Obersaxen (GR)	3612_1
Oberwald (WS)	6066	Oberwald (VS)	6066_1
Obstalden (GL)	1624	Obstalden (GL)	1624_1
Pfäfers (SG)	3294	Pfäfers (SG)	3294_1
Pfäffikon-Irgenhausen (ZH)	177	Pfäffikon (ZH)	177_1
Pfaffnau (LU)	1139	Pfaffnau (LU)	1139_1
Pieterlen (BE)	392	Pieterlen (BE)	392_1
Plaffeien (FR)	2299	Plaffeien (FR)	2299_1
Pratteln (BA)	2831	Pratteln (BL)	2831_1
Rafz (ZH)	67	Rafz (ZH)	67_1
Ramsen (SH)	2963	Ramsen (SH)	2963_1
Randa (WS)	6287	Randa (VS)	6287_1
Reckingen (WS)	6075	Reckingen-Gluringen (VS)	6067_1
Regensberg (ZH)	95	Regensberg (ZH)	95_1
Reichenbach im Kandertal (BE)	567	Reichenbach im Kandertal (BE)	567_2
Reutigen (BE)	767	Reutigen (BE)	767_1
Rheineck (SG)	3235	Rheineck (SG)	3235_1
Ricken (SG)	3331	Wattwil (SG)	3377_1
Rickenbach (SO)	2582	Rickenbach (SO)	2582_1
<i>Riken (AG)</i>	<i>4279</i>	<i>Murgenthal (AG)</i>	<i>4279_1</i>
Risch (ZG)	1707	Risch (ZG)	1707_2

Rodersdorf (SO)	2479	Metzerlen-Mariastein (SO)	2477_1
Roggenburg (BE)	2790	Roggenburg (BL)	2790_1
Roggwil (TG)	4431	Roggwil (TG)	4431_1
Rorschach-Rorschacherberg (SG)	3215	Romanshorn (TG)	4436_1
Rüdlingen (SH)	2938	Buchberg (SH)	2933_1
Rüeggisberg (BE)	880	Rüeggisberg (BE)	880_1
Rümlang (ZH)	97	Rümlang (ZH)	97_1
Ruswil (LU)	1098	Ruswil (LU)	1098_1
Saas-Grund (WS)	6291	Saas-Grund (VS)	6291_1
Safien (GR)	3651	Safien (GR)	3651_1
Salgesch (WS)	6113	Salgesch (VS)	6113_1
Sarnen-Kägiswil (UW)	1407	Sarnen (OW)	1407_1
Schaffhausen (SH)	2939	Schaffhausen (SH)	2939_1
Schangnau (BE)	906	Schangnau (BE)	906_1
Schänis (SG)	3315	Schänis (SG)	3315_1
Schiers (GR)	3962	Schiers (GR)	3962_1
Schleitheim (SH)	2952	Schleitheim (SH)	2952_1
Schmitten (GR)	3514	Wiesen (GR)	3523_1
Schnottwil (SO)	2461	Schnottwil (SO)	2461_1
Schönenbuch (BA)	2774	Schönenbuch (BL)	2774_1
Schöpfheim (LU)	1008	Schöpfheim (LU)	1008_1
Schwanden (GL)	1627	Schwanden (GL)	1627_1
<i>Schwarzenburg (BE)</i>	<i>854</i>	<i>Wahlern (BE)</i>	<i>854_1</i>
Schwarzseeal (FR)	2299	Plaffeien (FR)	2299_2
Schwyz (SZ)	1372	Schwyz (SZ)	1372_1
Seengen (AG)	4208	Boniswil (AG)	4192_1
Seftigen (BE)	883	Seftigen (BE)	883_1
Sempach (LU)	1102	Sempach (LU)	1102_1
Sennwald (SG)	3274	Sennwald (SG)	3274_1
Sevelen (SG)	3275	Sevelen (SG)	3275_1
Siblingen (SH)	2953	Gächlingen (SH)	2901_1
Siglistorf (AG)	4319	Siglistorf (AG)	4319_1
Signau (BE)	907	Signau (BE)	907_1
Silenen (UR)	1216	Silenen (UR)	1216_1
Simplon Dorf (WS)	6009	Simplon (VS)	6009_1
Sitterdorf (TG)	4511	Zihlschlacht-Sitterdorf (TG)	4511_1
St. Antönien (GR)	3893	St. Antönien (GR)	3893_1
St. Gallen (SG)	3203	St. Gallen (SG)	3203_1
St. Niklaus (WS)	6292	St. Niklaus (VS)	6292_1
St. Peter (GR)	3931	Arosa (GR)	3921_1
St. Stephan (BE)	793	St. Stephan (BE)	793_1
Stadel (ZH)	100	Stadel (ZH)	100_1
Stallikon (ZH)	13	Stallikon (ZH)	13_1
Stans-Oberdorf (UW)	1509	Stans (NW)	1509_1
Stein (SG)	3356	Stein (SG)	3356_1

Stein (SH)	2964	Stein am Rhein (SH)	2964_1
Sternenberg (ZH)	179	Sternenberg (ZH)	179_1
Stüsslingen (SO)	2499	Stüsslingen (SO)	2499_1
Sufers (GR)	3695	Splügen (GR)	3692_1
Sulgen (TG)	4506	Kradolf-Schönenberg (TG)	4501_1
Sumiswald (BE)	957	Sumiswald (BE)	957_1
Sursee (LU)	1103	Sursee (LU)	1103_1
Tafers (FR)	2306	Tafers (FR)	2306_1
Tamins (GR)	3733	Tamins (GR)	3733_1
Täuffelen (BE)	751	Täuffelen (BE)	751_1
Teufen (AP)	3024	Bühler (AR)	3021_1
Teufenthal (AG)	4145	Teufenthal (AG)	4145_1
Thun (BE)	942	Thun (BE)	942_1
Thusis (GR)	3668	Thusis (GR)	3668_1
Triengen (LU)	1104	Triengen (LU)	1104_1
Trimmis (GR)	3945	Trimmis (GR)	3945_1
Trogen (AP)	3025	Trogen (AR)	3025_1
<i>Tscherlach (SG)</i>	<i>3298</i>	<i>Walenstadt (SG)</i>	<i>3298_1</i>
Tuggen (SZ)	1347	Tuggen (SZ)	1347_1
Turbenthal (ZH)	228	Turbenthal (ZH)	228_1
Tüscherz (BE)	752	Tüscherz-Alfermée (BE)	752_1
Ueberstorf (FR)	2308	Ueberstorf (FR)	2308_1
<i>Unterillnau (ZH)</i>	<i>174</i>	<i>Illnau-Effretikon (ZH)</i>	<i>174_1</i>
Unterschächen (UR)	1219	Unterschächen (UR)	1219_1
Unterseen (BE)	593	Interlaken (BE)	581_1
Unterstammheim (ZH)	42	Unterstammheim (ZH)	42_1
Untervaz (GR)	3946	Untervaz (GR)	3946_1
Urdorf (ZH)	250	Urdorf (ZH)	250_1
Urnäsch (AP)	3006	Urnäsch (AR)	3006_1
Ursenbach (BE)	344	Ursenbach (BE)	344_1
Uster-Nänikon (ZH)	198	Uster (ZH)	198_1
Utzenstorf (BE)	552	Utzenstorf (BE)	552_1
Vals (GR)	3603	Vals (GR)	3603_1
<i>Vättis (SG)</i>	<i>3294</i>	<i>Pfäfers (SG)</i>	<i>3294_3</i>
Villigen (AG)	4121	Villigen (AG)	4121_1
Visp (WS)	6297	Visp (VS)	6297_1
Visperterminen (WS)	6298	Visperterminen (VS)	6298_1
Volketswil (ZH)	199	Uster (ZH)	198_2
Vorderes Diemtigtal (BE)	762	Diemtigen (BE)	762_1
Wädenswil (ZH)	142	Wädenswil (ZH)	142_1
Walchwil (ZG)	1710	Walchwil (ZG)	1710_1
Wald (ZH)	120	Wald (ZH)	120_1
Waldstatt (AP)	3007	Waldstatt (AR)	3007_1
Walkringen (BE)	626	Rubigen (BE)	623_1
Wangen (BE)	992	Wangen an der Aare (BE)	992_1

Wängi (TG)	4781	Wängi (TG)	4781_1
Wegenstetten (AG)	4262	Wegenstetten (AG)	4262_1
Weggis (LU)	1069	Weggis (LU)	1069_1
Weinfelden (TG)	4946	Weinfelden (TG)	4946_1
<i>Weisstannen (SG)</i>	3293	<i>Mels (SG)</i>	3293_2
Wengi (BE)	394	Wengi (BE)	394_1
Wil (SG)	3425	Wil (SG)	3425_1
Wilchingen (SH)	2974	Wilchingen (SH)	2974_1
Wildhaus (SG)	3357	Wildhaus (SG)	3357_1
Willisau (LU)	1151	Willisau (LU)	1149_1
Winterthur (ZH)	230	Winterthur (ZH)	230_1
Wolfenschiessen (UW)	1511	Wolfenschiessen (NW)	1511_1
Wolhusen (LU)	1107	Wolhusen (LU)	1107_1
Wollerau-Freienbach (SZ)	1323	Wollerau (SZ)	1323_1
Worb (BE)	627	Worb (BE)	627_1
Würenlos (AG)	4048	Würenlos (AG)	4048_1
Wynigen (BE)	424	Wynigen (BE)	424_1
Zell (LU)	1150	Zell (LU)	1150_1
Zermatt (WS)	6300	Zermatt (VS)	6300_1
Ziefen (BA)	2834	Ziefen (BL)	2834_1
Zofingen (AG)	4289	Zofingen (AG)	4289_1
Zug (ZG)	1711	Zug (ZG)	1711_1
Zunzgen (BA)	2869	Zunzgen (BL)	2869_1
Zürich (ZH)	261	Zürich (ZH)	261_1
Zuzwil (BE)	557	Rapperswil (BE)	310_1
Zweisimmen (BE)	794	Zweisimmen (BE)	794_1

C.2 Nicht berücksichtigte SADS-Ortspunkte

Gemeindename SADS	BFS_Nr SADS
Davos (GR)	3851_2
Ermatingen (TG)	4646_2
Niederweningen (ZH)	91_1
Pfäfers (SG)	3294_2
Solothurn (SO)	2601_1

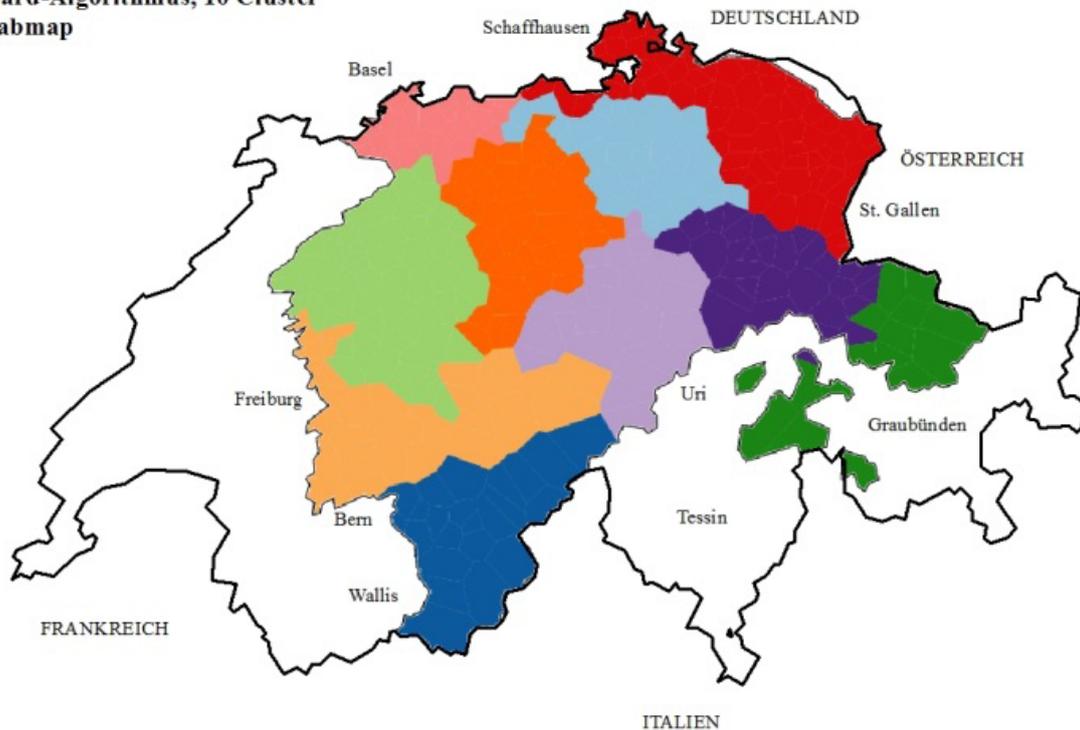
D. Liste mit SDS-Karten mit mehr als 20 Mehrfachvarianten

Karten mit mehr als 20 Mehrfachvarianten	Anzahl Mehrfachvarianten	Kommentare zur Auswahl einer Variante für VDM
1-015-Bett	29	
1-060-durch	74	
1-061	25	
1-095-Schnee	63	
1-112	33	
1-156	39	
1-165	55	
2-027-Besen	35	
2-170	20	
2-198	41	
3-007-gebracht	40	
3-008-gelaufen	81	
3-032-2pl-reg	28	
3-033-3pl-reg	36	
3-040f-1sg-kurz	182	cnt_0 rausgestrichen. Wenn -ne und -n vorkamen (19x), -ne gelassen, -n rausgelöscht.
3-044-1pl-kurz	37	
3-044-2pl-kurz	38	
3-044-3pl-kurz	37	
3-046-haben-pp	43	
3-051-1.3. pers_sg_konj_sein	48	
3-055-tun	48	
3-058ff-pl-vokal	94	
3-075-anfangen-23sg	33	
3-078-anfangen-inf	58	
3-080-liegen-pl	34	
3-081-liegen-23sg	39	
3-090-geben-pl	47	
3-091-nehmen-pl	51	
3-096-sehen-g	21	
3-102-kommen-pl	31	
3-112-wollen-2sg	68	
3-113-wollen-3sg	156	
3-131-das	50	
3-134-der-M-Sg	147	
3-142f-ein-N-Sg	119	Linke Variante bei «vor Substantiv» gewählt (nicht «vor Adjektiv» → wäre in ZH, SG und GL abweichend gewesen)
3-144-eine-F-Sg	72	
3-145-ein-M-Sg	89	
3-154-dimin-el	20	
3-162-in	22	
3-163-ung	49	Bei Unterschied in den Wörtern: Zeitung vor Rechnung bevorzugt Cnt_0 bevorzugt. *siehe Bemerkung Tabellenende
3-164ff-pl-e	563	
3-187-ene	24	
3-195f-ich	213	haupttonig, isoliert berücksichtigt
3-195f-mich	150	
3-216-ihr-F	20	
3-253-adj-Pl	198	flektiert bevorzugt
3-260-ihr_es	132	
3263_lass_ihn_gehen	40	
3-265_regnen	22	

4-155-ein-wenig	126	Karte vor Tabelle bevorzugt. «chli», «chlei», «chlai» gemäss linker Spalte, Tabelle 156
4-168-nicht-mehr	24	
4-176-nur	70	nume vor ume bevorzugt
4-180-und	39	
4-183-wie	131	wi und we vor wie bevorzugt
5-112-gruss	124	gxtag vor gxetag gemäss Legende
5-115-danke	225	
5-118-lueg	35	
5-119-ja	122	ja vor jo
6-021-heute	31	hütte vor hüt
6-022ff-manchmal	297	2. Prio Karte S 23, mängmal vor mängisch, öppe vor öppedie
6-026-immer	51	
6-027-jeweils	54	
6-029-fr1	153	
6-030-bald	30	
6-098-hier	38	hie bevorzugt
6-101-dort	70	
6-110-hinauf	21	
Karten ohne Total-Feld («cnt_Total»)		
2-071f	37	cnt_k bevorzugt*
2-093	18	cnt_ä bevorzugt*
3-164ff	21	cnt_0 bevorzugt*
		*Jeweils kleinere Anzahl = 0 gesetzt, damit die Variante ausgewählt wird, die öfter vorkommt. Wenn beide Varianten gleich oft vorkommen → Zufällige Auswahl (Anz. Fälle in Klammer).

E. Liste mit je zehn bestimmenden Faktoren pro Cluster

SDS- und SADS-Datensatz
Ward-Algorithmus, 10 Cluster
Gabmap



Die Hintergrundfarbe in der Tabelle entspricht derjenigen des betreffenden Clusters auf der Karte in Abbildung 23 auf Seite 68, die hier nochmals abgebildet ist. Variablen, die in mehreren Clustern als einer der ersten zehn bestimmenden Faktoren auftreten, sind mit der ihrer Auftretenshäufigkeit entsprechenden Anzahl «*» markiert.

Differenz	innerhalb Cluster	zwischen Clustern	Variable
-2.029546	-1.369854	0.659692	1134 t[ie]f: aobd. iu **
-1.99416	-1.460525	0.533635	6106 Richtungsadv.: hinaus
-1.875776	-0.645976	1.2298	3003 Präfix des stark. Verbs: Part. Perfkt
-1.829864	-1.332902	0.496962	6112 Richtungsadv.: hinab **
-1.758333	-1.220289	0.538044	6111 Richtungsadv.: herab **
-1.730125	-0.553526	1.176599	2003 Schn[a]bel: mhd. a off. Sil.
-1.647388	-1.351646	0.295742	6108 Richtungsadv.: hinein
-1.641931	-0.883376	0.758555	5115 Partikel: Dankformel
-1.605048	-1.560834	0.044214	1067 S[a]men: mhd. â
-1.596401	-0.098675	1.497726	2165 [T]ag: mhd. t (germ. d) im Anlt
-1.92811	-1.508274	0.419836	5112 Partikel: Grussformeln **
-1.765072	-0.471907	1.293165	3106 mögen: 1. Pers. Sg.
-1.676308	-1.121763	0.554545	3044 End. d. Kurzverb.: 3. Pers. Pl. Ind. Präs. ***
-1.571463	-0.661161	0.910302	3159 Bern[erin]: Subst. f. Derivation
-1.561006	-1.159214	0.401792	3050 sein: 1.2.3. Pers. Pl. Ind. Präs

-1.558664	-0.704881	0.853783	3204 Personalpr.: ihr (2. Pers. Pl.)
-1.536045	-1.11482	0.421225	3044 End. d. Kurzverb.: 2. Pers. Pl. Ind. Präs.
-1.526052	-0.886144	0.639908	1054 Br[ü]cke: Uml./ n. Uml. mhd. u v. ck, gg
-1.522302	-1.10691	0.415392	6029 Zeitadverb: früher (Lautung)
-1.522085	-1.048631	0.473454	3044 End. d. Kurzverb.: 1. Pers. Pl. Ind. Präs.
-2.00428	-1.962975	0.041305	4176 nur **
-1.899279	-1.537222	0.362057	3008 gelaufen: Part. Perfekt (f.)
-1.741631	-1.508415	0.233216	3051 sein: 1.3. Pers. Sg. Konj. Präs.
-1.581083	-0.956671	0.624412	3080 liegen: 1.-3. Pl.
-1.544825	-0.926892	0.617933	1060 [ü]ber:Umlt/ Numlt. mhd. u
-1.544737	-1.121763	0.422974	3044 End. d. Kurzverb.: 3. Pers. Pl. Ind. Präs. ***
-1.529528	-1.263242	0.266286	1152 b[au]en: mhd. û in Hiatt
-1.52891	-1.40055	0.12836	5119 Partikel: ja
-1.52808	-1.350639	0.177441	1148 schn[e]ien: mhd. î in Hiatt
-1.490053	-1.33621	0.153843	1156 n[eu]-e,-er,-es: ü in Hiatt
-2.652561	-0.53456	2.118001	3154 Vög[el]chen: Subst. Diminutiv
-2.518746	-2.39809	0.120656	6027 Zeitadverb: jeweils ****
-2.456436	-0.477304	1.979132	2144 Läu[s]e: mhd. s > sch **
-2.377317	-1.998795	0.378522	6101 Ortsadverb: dort
-2.100354	-1.121763	0.978591	3044 End. d. Kurzverb.: 3. Pers. Pl. Ind. Präs. ***
-2.08624	-0.582073	1.504167	1156 n[eu]-e,-er,-es: ü in Hiatt (Ent.)
-2.008183	-0.85264	1.155543	2119 f[ir]nd[en]: mhd. -nd-
-1.95301	-0.771521	1.181489	3018 fl[ie]gen, l[ü]gen: SSVok
-1.9509	-0.651106	1.299794	1160 Epf[e]ll: e > ö (Rund)
-1.898883	-1.252294	0.646589	1056 dr[ü]cken: Uml./ n. Uml. mhd. u v. ck, gg ****
-2.049373	-1.733885	0.315488	6027 Zeitadverb: jeweils ****
-1.765178	-1.450984	0.314194	3057 gehen (Vok): 2./ 3. Pers. Sg. Ind. Präs.
-1.668266	-1.484021	0.184245	6110 Richtungsadv.: hinauf **
-1.649843	-1.465597	0.184246	6109 Richtungsadv.: herauf
-1.609056	-1.604953	0.004103	3075 (an)f[an]gen: 2./3. Pers. Sg. Ind. Präs. ***
-1.489981	-0.858249	0.631732	2138 ge[rn]: mhd. -rn **
-1.484469	-0.983701	0.500768	6111 Richtungsadv.: herab **
-1.397043	-0.939125	0.457918	6112 Richtungsadv.: hinab **
-1.385794	-0.83256	0.553234	3263 lass ihn gehen: erstarrte Infinitiv-Partikeln
-1.354812	-1.343248	0.011564	1109 G[e]i[ss]: mhd. ei
-1.408248	-1.287258	0.12099	3075 (an)f[an]gen: 2./ 3. Pers. Sg. Ind. Präs. ***
-1.385774	-1.230341	0.155433	3074 (an)f[an]gen: 1. Pers. Sg. Ind. Präs.
-1.331499	-1.092751	0.238748	2027 B[e]sen: Dehnung alt. Kürze in off. Silbe, mhd. ë
-1.107456	-1.06538	0.042076	1056 dr[ü]cken: Uml./n. Uml. mhd. u v. ck, gg ****
-0.843223	-0.82171	0.021513	1015 B[e]tt: mhd.e (Primärumlaut)
-0.804193	-0.764246	0.039947	1055 R[ü]cken: Uml./nicht Uml. von mhd. u vor gg **
-0.795628	-0.708204	0.087424	2077 r[ei]ten: Kürz. alt. Läng vor t, k
-0.725827	-0.670458	0.055369	1052 F[ü]chse: mhd. ü
-0.698097	-0.534199	0.163898	1048 Schl[i]tten: mhd. i
-0.696006	-0.447991	0.248015	6027 Zeitadverb: jeweils ****

-2.268966	-2.062026	0.20694	3047 haben: Pl. Ind. Präs.
-1.436937	-0.886982	0.549955	3207 Personalpr.: uns **
-1.379095	-0.429647	0.949448	3108 dürfen: 3. Pers. Pl.
-1.376883	-1.135101	0.241782	3134 der:best. Art. Nom. Sg. m
-1.361142	-1.110644	0.250498	1129 räuchern: mhd. öu
-1.345012	-1.05652	0.288492	1121 [Au]gen: mhd. ou
-1.324623	-1.283684	0.040939	4181 sonst **
-1.305124	-0.937213	0.367911	3016 sch[ie]ssen: SSVok (2. Kl. stark. V.)
-1.287422	-1.270055	0.017367	1055 R[ü]cken: Uml./ nicht Uml. von mhd. u vor gg **
-1.258583	-1.043377	0.215206	3028 kostet: Endung Normalv.: 3. Pers. Sg. Ind. Präs.
-1.605122	-0.936313	0.668809	2138 ge[rn]: mhd. -rn **
-1.371664	-1.213549	0.158115	3218 Possessivpr.: unser/ euer (pl.)
-1.314707	-0.852237	0.46247	3218 Possessivpr.: unser/ euer (f.)
-1.262129	-1.252294	0.009835	1056 drücken: Uml./ n. Uml. mhd. u v. ck, gg ****
-1.260533	-0.257108	1.003425	3217 Possessivpr.: unser (Dat)
-1.242295	-0.37679	0.865505	3054 tun: Part. Perfekt
-1.234954	-0.048791	1.186163	2095 drü[ck]en: urd. kk hint.Vok i. Inlt u. Auslt
-1.216026	-0.638999	0.577027	2097 tr[ink]en: germ. -nk(-)
-1.068242	-0.907047	0.161195	3091 nehmen: 1.-3. Pl.
-1.052351	-0.852329	0.200022	4176 nur **
-2.409103	-2.303769	0.105334	6027 Zeitadverb: jeweils ****
-2.033422	-0.656496	1.376926	3266 wirst/ kommst: Umschr. inchoativen Funkt.
-1.973816	-0.174672	1.799144	2144 Läu[s]e: mhd. s > sch **
-1.95181	-1.560414	0.391396	6110 Richtungsadv.: hinauf **
-1.948365	-1.630336	0.318029	4181 sonst **
-1.625031	-0.797895	0.827136	2113 wa[chs]t: mhd. -hs-
-1.504952	-1.097996	0.406956	3265 es kommt regnen: erstarrte Infinitiv-Partikeln
-1.476278	-1.450594	0.025684	3112 wollen: 2. Pers. Sg.
-1.420571	-1.028315	0.392256	1164 Br[i]lle: i > ü (Rund)
-1.395873	-0.574549	0.821324	3207 Personalpr.: uns **
-1.705452	-0.8012	0.904252	3262 gehen lassen: Wortstellung
-1.658839	-0.943049	0.71579	3048 haben: 2./ 3. Pers. Sg. Ind. Präs.
-1.651259	-1.540238	0.111021	5112 Partikel: Grussformeln **
-1.647754	-1.25387	0.393884	1134 t[ie]f: aobd. iu **
-1.647171	-0.943759	0.703412	201 Verdoppelung lassen: flicken lassen
-1.553694	-1.404112	0.149582	3209 Personalpr.: euch
-1.534006	-0.71742	0.816586	3142 ein (I):unbest. Art. Nom. Akk. Sg. Neutr
-1.466754	-0.999176	0.467578	3145 ein: unbest. Art. Nom. Akk. Sg. m
-1.448928	-1.178136	0.270792	3075 (an)fangen: 2./ 3.Pers. Sg. Ind. Präs. ***
-1.439487	-1.252294	0.187193	1056 dr[ü]cken: Uml./ n. Uml. mhd. u v. ck, gg ****

F. Inhalt der CD

Name Ordner	Inhalt	Abschnitt
<i>01_Arbeit</i>	Digitale Arbeit als PDF (und .rtf)	
<i>02_Abbildungen</i>	Alle Abbildungen aus der Arbeit im jpg.-Format	
<i>03_Python-Skripts</i> a. <i>SADS_Gabmap</i> b. <i>SADS_VDM</i> c. <i>SDS_Gabmap</i> d. <i>SDS_VDM</i>	a. Alle SADS-Dateien und Python-Skripts für die Daten-Aufbereitung für Gabmap b. Alle SADS-Dateien und Python-Skripts für die Daten-Aufbereitung für VDM c. Alle SDS-Dateien und Python-Skripts für die Daten-Aufbereitung für Gabmap d. Alle SDS-Dateien und Python-Skripts für die Daten-Aufbereitung für VDM	4.2.3
<i>04_VDM-Projekte</i>	Die fertigen VDM-Dateien und Projekte	4.3
<i>05_Gabmap-Dateien</i>	Die fertigen Dateien für die Verwendung in Gabmap	4.4

Persönliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und die den verwendeten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Zürich, 30. April 2014

Sandra Kellerhals