

UNIVERSITÄT ZÜRICH

GEO 511 – MASTERARBEIT

---

**Einfluss von scharfen und unscharfen Grenzen  
auf syntaktische Dialektunterschiede in der  
deutschen Schweiz**

---

CHRISTIAN D. SIEBER

Matrikelnummer 11-747-730

Betreuung: Dr. Curdin Derungs

Co-Betreuung: Prof. Dr. Elvira Glaser

Deutsches Seminar

Schönbergstrasse 9

8001 Zürich

eglaser@ds.uzh.ch

Verantwortliches Fakultätsmitglied: Prof. Dr. Robert Weibel

GEOGRAPHISCHE INFORMATIONSWISSENSCHAFT (GIS)

Abteilung Geographie

Eingereicht am 30. Juni 2017



*„The finality of language is mobility, not immobility.“*

(Jan Blommaert, 2010, S. xiv)



# Zusammenfassung

Sprachen werden von einer Vielzahl von Faktoren beeinflusst. Es besteht die linguistische Theorie, dass politische, religiöse und wirtschaftliche Grenzen sowie geographische Distanzen als Kontakt- und somit als Sprachbarrieren wirken. Ihre effektive Bedeutung ist in der linguistischen Forschung jedoch noch umstritten.

In der vorliegenden Arbeit wird zum einen der Einfluss diverser scharfer und unscharfer Grenzen auf die syntaktischen Dialektunterschiede zwischen Ortschaften untersucht. Zum anderen wird mit der linearen, multiplen Regressionsanalyse eine Methode vorgestellt, mit der alle untersuchten Grenzarten gleichzeitig in ein einziges Modell eingebunden werden können. Die Regressions-Modellierung der dialektischen Unterschiede zwischen Ortschaften soll Erkenntnisse über den Einfluss von unterschiedlichen Grenzarten liefern und die Grundlage für eine zukünftige, flächenhafte Modellierung von Dialektdaten bilden. Die damit gelegte Basis soll Ausgangspunkt für eine angepasste spezifische Interpolationsmethode für Sprachdaten sein.

Die Resultate zeigen, dass sowohl die untersuchten scharfen Grenzarten als auch die untersuchten geographischen Distanzen einen starken Zusammenhang mit den sprachlichen Unterschieden der Ortschaften aufweisen. Des Weiteren zeigt sich, dass mit der linearen, multiplen Regressionsanalyse eine geeignete Methode besteht, sprachliche Unterschiede mit mehreren erklärenden Faktoren zu beschreiben.



# Danksagung

Vor acht Jahren habe ich beschlossen, meinen ursprünglich gelernten Beruf aufzugeben, meine damalige Arbeitsstelle zu kündigen und an der technischen Berufsmittelschule die Berufsmaturität zu erlangen. Damals wusste ich noch nicht, wohin mich der zweite Bildungsweg führen würde. Heute weiss ich es.

Die vorliegende Arbeit stellt den Abschluss meines erfolgreichen Studiums der Geographie mit Vertiefung in Geographischer Informationswissenschaft dar.

Besonders bedanken möchte ich mich bei meinem Supervisor Dr. Curdin Derungs für seine wertvolle Unterstützung mit Ideen, Verbesserungsvorschlägen, Hilfestellungen bei methodischen und technischen Problemen sowie seinen aufmunternden Worten während dem letzten Jahr. Ich hätte mir keinen besseren Betreuer wünschen können. Ebenso gilt mein Dank Prof. Dr. Robert Weibel, Prof. Dr. Elvira Glaser, Dr. Philipp Stöckle und (baldiger Dr.) Péter Jeszenszky für ihre hilfreichen Ideen, ihre Zeit und dem Teilen ihres Wissens.

Anne Wegmann, Thomas Lüscher, Jessica Zanetti und Martina Schmocker danke ich besonders für das zeitintensive Gegenlesen und Korrigieren der Arbeit während ihren freien Wochenenden.

Meinen Mitbewohnern danke ich für ihre Geduld, welche sie mit mir in Zeiten von Prüfungs- und Masterarbeits-Stress hatten. Besonderer Dank gilt Christoph Moser für die starken Espresso, feinen Sandwiches und die leckeren Pesto Rosso Spaghetti während den letzten paar Wochen vor dem Abgabetermin, Franziska Schmid für ihren hilfreichen Einfall mit den euklidischen Reisezeiten sowie Sabrina Gnehm für ihre stets stoisch-ruhige Art und ihr herzliches Lächeln, das die Sonne scheinen lässt.

Meinen Bandkollegen Samuel Thalmann und Julian Kronbach danke ich für ihr Verständnis bezüglich meiner Abwesenheit in den Proben während den letzten Wochen.

Andrea Brea danke ich besonders für ihre aufmunternden und lieben Worte, dem last-minute-Gegenlesen trotz schlimmer Erkältung und für die erholsamen Pausen von der Masterarbeit. Sie ermöglichten mir, in den letzten, stressigsten Wochen neue Energie zu schöpfen.

Meinen Studienfreunden Annabelle Jaggi, Sandra Burkhalter, Jessica Zanetti, Anne Wegmann und Fabiola Kälin danke ich für ihre Unterstützung während den letzten Jahren.

Meinen Eltern gilt der grösste Dank. Ohne ihre unendlich grosse Unterstützung und Liebe während all diesen Jahren würde ich nicht da sein, wo ich jetzt bin. Danke!

# Inhaltsverzeichnis

<b>Zusammenfassung</b>	<b>iii</b>
<b>Danksagung</b>	<b>v</b>
<b>Abkürzungsverzeichnis</b>	<b>xi</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Relevanz für die Geoinformations- und Sprachwissenschaften . . . . .	2
1.2 Das Ziel und die Motivation dieser Arbeit . . . . .	3
<b>2 Forschungskontext</b>	<b>5</b>
2.1 Bisherige Erkenntnisse über Sprachveränderungen . . . . .	5
<b>3 Forschungslücken und Ziele</b>	<b>11</b>
3.1 FF 1: Haben politische, religiöse und wirtschaftliche Grenzen einen Einfluss auf sprachliche Unterschiede? . . . . .	11
3.2 FF 2: Haben geographische Distanzen einen Einfluss auf sprachliche Unterschiede? . . . . .	12
3.3 FF 3: Können syntaktische Unterschiede in einem Modell mit mehreren scharfen und unscharfen Grenzen erklärt werden? . . . . .	12
<b>4 Daten</b>	<b>15</b>
4.1 Hintergrundwissen . . . . .	15
4.2 Verwendete Daten . . . . .	15
4.2.1 Die syntaktische Distanz . . . . .	18
<b>5 Methodik</b>	<b>21</b>
5.1 Unschärfe Grenzen: Distanzmasse . . . . .	23
5.1.1 Die Erreichbarkeit . . . . .	23
5.1.2 Die Einflussbereichsmasse . . . . .	30
5.2 Scharfe Grenzen . . . . .	31
5.2.1 Kantongrenzen . . . . .	31
5.2.2 Konfessionsgebietsgrenzen . . . . .	33
5.2.3 Arbeitsmarktregionsgrenzen . . . . .	34

5.3	Verwendete Analysemethoden . . . . .	36
5.3.1	Räumlicher Effekt und Skalenproblem (MAUP) . . . . .	36
5.3.2	Korrelationsanalysen und Dominanzkarten . . . . .	42
5.3.3	Mittelwerttests . . . . .	43
5.3.4	Lineare, multiple Regressionsmodelle . . . . .	46
<b>6</b>	<b>Resultate und Interpretationen</b>	<b>53</b>
6.1	Unschärfe Grenzen: Distanzmasse . . . . .	53
6.1.1	Topographie-Einfluss auf die syntaktische Distanz . . . . .	53
6.1.2	Korrelationsanalysen: Boxplots . . . . .	56
6.1.3	Korrelationsanalysen: Korrelationskarten . . . . .	57
6.1.4	Korrelationsanalysen: Dominanzkarten . . . . .	63
6.2	Scharfe Grenzen: Kantone, Konfessionsgebiete und Arbeitsmarktregionen . . . . .	65
6.2.1	Teststatistiken der Mittelwerttests . . . . .	65
6.3	Regressionsmodelle mit scharfen und unsharp Grenzen . . . . .	68
<b>7</b>	<b>Diskussion</b>	<b>73</b>
7.1	FF 1: Haben politische, religiöse und wirtschaftliche Grenzen einen Einfluss auf sprachliche Unterschiede? . . . . .	73
7.2	FF 2: Haben geographische Distanzen einen Einfluss auf sprachliche Unterschiede? . . . . .	75
7.3	FF 3: Können syntaktische Unterschiede in einem Modell mit mehreren geographischen und demographischen Faktoren erklärt werden? . . . . .	78
7.4	Die abschliessende Relevanz der Resultate . . . . .	79
<b>8</b>	<b>Limitierungen</b>	<b>81</b>
8.1	Limitierungen in Bezug auf die Modellannahmen . . . . .	81
8.2	Weitere Limitierungen . . . . .	82
<b>9</b>	<b>Konklusion</b>	<b>85</b>
	<b>Anhang</b>	<b>87</b>

# Abbildungsverzeichnis

4.1	383 Erhebungsorte des SADS . . . . .	16
4.2	Beispielfragen aus dem ersten SADS-Fragebogen . . . . .	17
4.3	Syntaktische Distanz-Berechnungsbeispiel . . . . .	19
5.1	Übersicht der Methodik . . . . .	22
5.2	Least Cost Path Visualisierung . . . . .	24
5.3	DHM25 und Convex Hull mit 8km Buffer um SADS-Orte . . . . .	25
5.4	Toblers Hiking Funktion . . . . .	25
5.5	Least Cost Path Visualisierung zwischen St. Antönien und Klosters . . . . .	27
5.6	Multidimensionale Skalierung (MDS) der SADS-Orte . . . . .	29
5.7	Kantons Grenzen der Schweiz . . . . .	32
5.8	Konfessionskarte der Schweiz und die SADS-Erhebungsorte . . . . .	34
5.9	Arbeitsmarktregionen der Schweiz . . . . .	35
5.10	MAUP: Skalenproblem . . . . .	37
5.11	MAUP: Problem in der Kartographie . . . . .	37
5.12	MAUP: Zonenproblem . . . . .	38
5.13	Voronoi-Polygone . . . . .	39
5.14	Voronoi-Polygone der SADS-Orte . . . . .	39
5.15	Visualisierte Voronoi-Nachbarschaften . . . . .	40
5.16	Unterschiedliche Nb-Skalen und ihre Daten . . . . .	41
5.17	Regressionsgerade mit Methode der kleinsten Quadrate . . . . .	47
5.18	Regressionsplot der 50 nächsten Nachbarn von Küblis . . . . .	48
5.19	Regressionsplot der Ortschaft Küblis: Kantons Grenzen . . . . .	50
6.1	Lokaler Vergleich der syntaktischen Distanz zum Topographie-Einfluss auf die Reisezeiten . . . . .	55
6.2	Boxplots der Korrelationswerte . . . . .	56
6.3	Korrelationskarten Voronoi- vs. 50-Nb-Skala . . . . .	59
6.4	Topographie der Deutschschweiz und die Einwohnerzahlen der SADS-Orte . . . . .	61
6.5	Abweichung der syntaktischen Distanz vom Durchschnitt sowie die GI-LCPs der 50-Nb-Skala . . . . .	61
6.6	Vergleich mit den Dialektregionen von Sandra Kellerhals . . . . .	62

6.7	Dominanzkarten . . . . .	64
6.8	$R^2_{kor}$ -Werte der Regressionsmodelle mit und ohne Einbezug der scharfen Grenzen. . . . .	70
9.1	Lokaler Vergleich der syntaktischen Distanz zum Topographie-Einfluss auf die Reisezeiten . . . . .	88
9.2	erklärende Graphik für die Boxplot-Darstellung . . . . .	89
9.3	Korrelationskarten 5- vs. 10-Nb-Skala . . . . .	90
9.4	Korrelationskarten 15- vs. 20-Nb-Skala . . . . .	91
9.5	Korrelationskarten 25- vs. 30-Nb-Skala . . . . .	92
9.6	Q-Q Plots der syntaktischen Distanzen . . . . .	93
9.7	Q-Q Plots der syntaktischen Distanzen . . . . .	94

# Tabellenverzeichnis

5.1	Asymmetrische LCP Reisezeiten . . . . .	26
5.2	Symmetrische LCP Reisezeiten . . . . .	27
5.3	Euklidische Reisezeiten . . . . .	28
5.4	Durchschnittlicher Topographie-Einfluss auf Reisezeiten . . . . .	28
5.5	Mögliche acht Fälle der scharfen Grenzen, die zwischen einem Ortspaar liegen können. . . . .	49
6.1	Resultate der Mittelwerttests für die Kantonsgrenzen . . . . .	66
6.2	Resultate der Mittelwerttests für die Konfessionsgebiete . . . . .	66
6.3	Resultate der Mittelwerttests für die Arbeitsmarktregionen . . . . .	67
6.4	$R_{korr}^2$ -Werte der Regressionsmodelle mit Einbezug der scharfen Grenzen . . . . .	71
6.5	$R_{korr}^2$ -Werte der Regressionsmodelle ohne Einbezug der scharfen Grenzen . . . . .	72

# Abkürzungsverzeichnis

<b>DHM</b> . . . . .	Digitales Höhenmodell
<b>ED</b> . . . . .	euklidische Distanz (hier jedoch für <i>euklidische Reisezeit</i> verwendet)
<b>FDP</b> . . . . .	Fundamental Dialectological Postulate
<b>FF</b> . . . . .	Forschungsfrage
<b>GI</b> . . . . .	Gravitations-Index
<b>GI-ED</b> . . . . .	Gravitations-Index mit euklidischer Reisezeit als Distanzmass
<b>GI-LCP</b> . . . . .	Gravitations-Index mit Least Cost Path Distanz als Distanzmass
<b>GIS</b> . . . . .	Geographische Informationssysteme
<b>GIScience</b> . . . . .	engl. Synonym für <i>Geographische Informationswissenschaft</i>
<b>LCP</b> . . . . .	Least Cost Path
<b>MAUP</b> . . . . .	Modifiable Areal Unit Problem
<b>MDS</b> . . . . .	Multidimensionale Skalierung
<b>SADS</b> . . . . .	Syntaktischer Atlas der deutschen Schweiz
<b>SDS</b> . . . . .	Sprachatlas der deutschen Schweiz



# Kapitel 1

## Einleitung

Sprache bewegt und berührt uns Menschen. Sie bringt uns zum Nachdenken, zum Lachen oder verletzt uns. Sie trägt zum gegenseitigen Verständnis bei oder führt zu Missverständnissen. Sie kann uns im Falle einer Fremdsprache oder eines Dialekts ausschliessen, uns aber auch Türen in fremde Kulturen und Welten öffnen. Unser eigener Dialekt stiftet Identität, sowohl persönlich als auch kulturell und bildet eine wichtige Grundlage für ein Zusammengehörigkeitsgefühl (z.B. Bartholy, 1992; Esser, 1983; Lameli, 2013; Reichmann, 2000). Auch in der Schweiz, welche geprägt ist von ihren vier Amtssprachen Deutsch, Französisch, Italienisch und Rätoromanisch, ist dieses Zusammengehörigkeitsgefühl wichtig. Die ersten drei erwähnten Sprachen werden in klar abgegrenzten Gebieten gesprochen, wo grösstenteils keine Mehrsprachigkeit besteht. Der deutschsprachige Landesteil unterscheidet sich aber insofern von den anderen, dass die mündliche Kommunikation polylektal (jeder spricht seinen Dialekt) erfolgt und das Standarddeutsche sich nur auf die Schrift<sup>1</sup> beschränkt (Glaser und Bart, 2015). Dass in der Deutschschweiz des Weiteren ein grosses öffentliches Interesse an Dialekten besteht, zeigen Radiosendungen<sup>2</sup>, Fernsehberichte<sup>3</sup>, die grosse Anzahl freiwilliger Teilnahmen an Sprachstudien<sup>4</sup> oder der kommerzielle Erfolg des «Kleinen Sprachatlas der Deutschen Schweiz» (Christen et al., 2010). Letzterer ist bereits in der 6. Auflage erhältlich und veranschaulicht die Vielfalt der schweizerischen Dialektlandschaft. Abgesehen von der alltäglichen Dialekt-Relevanz gibt es auch einige wichtige Kernforschungsfragen in den Sprachwissenschaften. Dazu gehören zum Beispiel die Fragen, was genau an Dialekten identitätsstiftend ist oder wie sich die Dialekte zukünftig entwickeln (z.B.

---

<sup>1</sup> Obwohl bei der jungen Generation auch die schriftliche Kommunikation in Dialekt verbreitet ist, ist der Grossteil der Schriftlichkeit (z.B. in den Medien, Kommunikation mit den Behörden, usw.) weiterhin in Standarddeutsch.

<sup>2</sup> z.B. die Sendungen «Schnabelweid», «Schwiiz und dütlich» oder «Auf den Spuren eures Namens» auf Radio SRF über die Herkunft und Bedeutung diverser schweizerdeutscher Begriffe, Flur- sowie Familiennamen (SRF Radio 2016: o.S., Zugriff 8.9.2016)

<sup>3</sup> z.B. die Episode der Sendung «Einstein» auf SRF über Herkunft, Wirkung und Bedeutung deutschschweizer Dialekte (SRF Fernsehen 2016: o.S., Zugriff 8.9. 2016)

<sup>4</sup> z.B. der Erfolg der «Dialäkt Äpp» (o.J., o.S., Zugriff 8.9.2016) oder dem «Chochichäschtl-Orakel» (o.J., o.S., Zugriff 8.9.2016), beides Apps resp. Websites, auf denen man seine Dialektherkunft bestimmen kann und gleichzeitig seine eingegebenen Daten an Wissenschaftler übermittelt.

ob innerhalb der Deutschschweiz eine Angleichung der Dialekte zu einem «Einheitsdialekt» stattfindet). Eine weitere Kernforschungsfrage ist, wie die Dialekte und Sprachen über den geographischen Raum verteilt sind und was für eine Rolle dabei der geographische Raum spielt. Diese letzte Frage stellt die Grundlage der vorliegenden Arbeit dar.

## 1.1 Relevanz für die Geoinformations- und Sprachwissenschaften

Obwohl die räumliche Analyse von Daten ein Kerngebiet der Geoinformationswissenschaften ist, wurde die Analyse von Sprachdaten in der GIScience bislang wenig thematisiert. Zwar schrieben Lee und Kretzschmar (1993, S. 541) in einem Artikel:

*„It is somewhat surprising that up to this point linguistic researchers have not maintained closer relations with other geographers except for cartographic assistance.“*

Doch änderte sich bis heute nicht viel an der Tatsache, dass geostatistische Analysen von Sprachdaten in der GIScience ein Nischendasein führen. Für die Geographie im Allgemeinen und die Geoinformationswissenschaften im Spezifischen sind Sprachdaten aber eine interessante Thematik. Dies zeigt sich darin, dass Sprachphänomene oft kontinuierlich über den Raum verteilt sind, aber auch scharfe Grenzen aufweisen können. Sprachen zeugen von den soziokulturellen und geographischen Einflüssen, denen sie unterworfen sind (z.B. Politik, Religion, Wirtschaft oder Topographie) und können beispielsweise, wie Lameli (2013) in seiner Arbeit «Strukturen im Sprachraum» aufzeigte, als Approximation von Kulturräumen dienen. Solche Approximationen von räumlichen Gliederungen können wiederum in Modellen anderer Wissenschaftszweigen, wie z.B. in der Wirtschaftsgeographie, zur Analyse von Binnenmigration genutzt werden (vgl. Lameli, 2013). Umgekehrtes gilt für die linguistische Forschung: zwar werden Geographische Informationssysteme (GIS) von Sprachwissenschaftlern vermehrt verwendet, dies aber meist nur für flächenhafte Visualisierungen von Sprachdaten und ohne die geostatistischen Methoden der Geoinformationswissenschaften oder das Potential von GIS vollumfänglich zu nutzen (z.B. Glaser, 2014; Pickl et al., 2014). Die GIScience und die daraus entwickelten GIS bieten den Linguisten viel mehr als nur die Darstellung von anschaulichen Karten. Die Geoinformationswissenschaften legen mit ihrer Forschung die Grundlage für die räumliche Datenanalyse in allen Wissenschaftszweigen. Dabei sind es weniger die ständig weiter entwickelten, vielfältigen Methoden der GIScience, welche den Wissenschaftlern die Grenzen des Möglichen aufzeigen, sondern vielmehr der eigene Einfallsreichtum und das Wissen der Forschenden über die vorhandenen geostatistischen Methoden. Genau so wie Linguisten von Geoinformationswissenschaftlern profitieren können, ist es auch umgekehrt. Ohne Hintergrundwissen über Sprachdaten und tiefgreifenden Kenntnissen über Zusammenhänge der Sprachveränderungen ergibt geolinguistische Forschung wenig Sinn. Vor diesem Hintergrund bietet sich eine interdisziplinäre Zusammenarbeit zwischen den Geoinformationswissenschaften und den Sprachwissenschaften an.

## 1.2 Das Ziel und die Motivation dieser Arbeit

Das Ziel dieser Masterarbeit ist es, erstmalig mehrere scharfe und unscharfe Grenzen in ein einziges Regressionsmodell einzubinden. Mit diesem Modell sollen die Dialektunterschiede zwischen Ortschaften möglichst exakt erklärt werden. Als scharfe Grenzen dienen politische, religiöse und wirtschaftliche Grenzen; als unscharfe Grenzen werden verschiedene Distanzmasse verwendet. Die Untersuchung der einzelnen scharfen und unscharfen Grenzen sowie die anschließende Regressions-Modellierung der Dialektunterschiede soll Erkenntnisse über den Einfluss von unterschiedlichen Grenzarten liefern. Zudem soll das Regressionsmodell die Grundlage für eine zukünftige, flächenhafte Modellierung von Dialektdaten bilden und Ausgangspunkt für eine angepasste, spezifische Interpolationsmethode von Sprachdaten sein.

Präzise Modelle und die daraus erstellten Karten, welche die Dialektregionen und die Ursachen der Dialektunterschiede exakt beschreiben, sind nicht nur für die Sprachwissenschaften von Nutzen, sondern auch für andere Wissenschaftszweige, wie beispielsweise den Wirtschaftswissenschaften. Solide, verlässliche Karten helfen, räumliche Prozesse und Strukturen innerhalb der Dialektverteilung aufzudecken. Sie erweitern das Verständnis für vergangene und zukünftige Entwicklungen der Dialekte und dienen dazu, zugrundeliegende Prozesse, die für die Dialektunterschiede verantwortlich sind, zu erkennen. Als Approximationen von Kulturregionen helfen Dialektgebiete, wirtschaftlichen Austausch und regionale Arbeitsmigration zu verstehen. So sind Dialektregionen gemäss Falck et al. (2012) Abbilder von lokaler Kultur und beeinflussen damit als kulturelle Barrieren den wirtschaftlichen Transfer der Regionen. Gebiete mit stark unterschiedlichen Dialekten unterliegen laut Falck et al. (2012) dem wirtschaftlich bedeutenden Effekt, dass ihre Dialektunterschiede regionale Migrationsströme negativ beeinflussen. Aus diesen Gründen ergibt sich die Wichtigkeit eines Modells, welches die Ursachen der Dialektunterschiede aufdeckt und beschreibt. Das Wissen über die Ursachen der Dialektunterschiede sowie die exakte Visualisierung der Dialektregionen helfen dabei, die negativen Folgen zu bekämpfen und die positiven Auswirkungen zu nutzen.



# Kapitel 2

## Forschungskontext

### 2.1 Bisherige Erkenntnisse über Sprachveränderungen

Sprachen und Dialekte sind nichts Statisches, sondern verändern und vermischen sich mit der Zeit. Welche Faktoren und Mechanismen dafür verantwortlich sind, ist noch immer nicht klar (z.B. Bowerman und Evans, 2015; Lucas, 2015). Es besteht aber die linguistische Hypothese, dass sich Sprache über den Kontakt zu anderen Menschen verändert und anpasst (z.B. Bowerman, 2013; Trudgill, 2010). Dabei ist jedoch unklar, ob externe oder interne Faktoren stärker für Sprachveränderungen verantwortlich sind. So verändern sich Dialekte gemäss Auer et al. (2011) hauptsächlich durch den Kontakt mit anderen Sprechern und weniger durch endogene Ursachen. Für Thomason (2001) sind hingegen externe und interne Faktoren gleichermaßen dafür verantwortlich. Klar ist aber auch für Thomason (2001), dass der Kontakt zwischen Menschen die treibende Kraft der Sprachveränderung ist.

Viele Linguisten beschäftigen sich deshalb mit den Auswirkungen von Kontakt auf Dialekt- und Sprachveränderungen. Trudgill (2010) beschäftigt sich zum Beispiel mit der Frage, in was für einem Ausmass der Kontakt von Sprechern einer spezifischen Sprachvariante zu Sprechern einer anderen Variante diese beeinflusst. Er kommt zum Schluss, dass Gesellschaften mit häufigem Kontakt zu anderen Gesellschaften eine höhere Wahrscheinlichkeit für eine komplexe Sprache aufweisen als Gesellschaften, die isolierter leben. Neben dem Komplexer-Werden einer Sprache kann durch Kontakt aber auch genau das Gegenteil, nämlich eine Simplifizierung von Sprache, eintreffen. Dies ist zum Beispiel bei Pidgin oder kreolischen Sprachen der Fall, welche als Beihilfssprachen für gegenseitige Verständigung zur Zeit der Kolonisierung entstanden sind (Trudgill, 2010) (vgl. auch Siegel, 2008). Trudgill zeigt in seiner Arbeit zwar die Gründe für eine allfällige Veränderung von Sprachen auf, geht dabei aber nicht auf die Faktoren ein, welche Kontakt zwischen Menschen überhaupt ermöglichen oder behindern. Bowerman (2013) untersuchte u.a. die Frage, ob Sprachkontakt durch verwandte Sprachen vereinfacht wird und kommt zum Schluss, dass viele Behauptungen über den Einfluss von Sprachverwandtschaft bei kontakt-bedingter Sprachveränderung mit anderen Faktoren als der Sprachverwandtschaft zusammenhängen, wie beispielsweise geographischer Nähe.

Dass Geographie einen Einfluss auf Sprache hat, konnte z.B. Nichols (2013) aufzeigen. In ihrer Arbeit *«The vertical archipelago: Adding the third dimension to linguistic geography»* untersuchte sie die nordostkaukasische-Sprachfamilie des östlichen Kaukasusgebirges um Südrussland, nördliches Georgien sowie Aserbajdschan. Nichols entdeckte, dass Höhenunterschiede in diesen Gebirgsregionen einen grossen Einfluss auf strukturelle Eigenschaften der Sprachen und die soziolinguistische Dimension der Bewohner haben. Höher gelegene Regionen sind isolierter und ihre wirtschaftlichen Ressourcen oft knapp, weshalb deren Bewohner mit Menschen aus tiefer gelegenen Regionen Handel betreiben müssen. Der Kontakt mit Flachlandbewohnern auf deren Märkten führt dazu, dass Bergbewohner oftmals zweisprachig sind und neben ihrer eigenen auch die Sprache der Flachländer sprechen (müssen). Im Gegensatz dazu beherrschen die Flachländer die Sprache der Bergbewohner oftmals nicht (Nichols, 2013).

Dass ausser Höhenunterschieden auch andere geographische Faktoren die Kontakthäufigkeit und -art bestimmen, wurde in zahlreichen Studien aufgezeigt. Neben wirtschaftlichen Gründen, wie sie Nichols (2013) beschrieb, wird vermutet, dass Einflüsse wie Reisedistanz, soziale und historische (bspw. religiöse) Permeabilität, administrative Grenzen oder Topographie (bspw. Berge, Seen und Flüsse) als Kontaktbarrieren und somit linguistisch gesehen auch als Sprachbarrieren wirken. So untersucht zum Beispiel Pickl (2013) den Einfluss von Flüssen und politischen (u.a. historischen) Grenzen als vermutete Sprachbarrieren. Gemäss Pickl (2013, S. 146) sind Flüsse *«(...) Verkehrshindernisse, die je nach Breite, Tiefe, Fließgeschwindigkeiten etc. unterschiedlich schwer zu überwinden sind.»* und somit Kontaktbarrieren darstellen. Politische Grenzen können laut Pickl (2013) wegen Passier- und Zollkontrollen ebenfalls starke Verkehrshindernisse bedeuten, aber auch als Verwaltungsgrenzen mit z.B. unterschiedlichen Schulsystemen oder Vorschriften in der Berufsausübung Kommunikationshindernisse darstellen. In seiner Arbeit konnte Pickl aufzeigen, dass die Flüsse *Lech* und *Donau* eine signifikante Trennwirkung auf die Dialekte im deutschen Bayerisch-Schwaben haben. Bei den politisch-historischen Grenzen kam er zum Schluss, dass vor allem ältere Grenzen einen Einfluss auf die Dialektunterschiede haben. Dies insbesondere dann, wenn sie von Konfessionsunterschieden (katholisch/protestantisch) begleitet werden. Den Einfluss von politischen Grenzen untersuchten auch Kürschner und Gooskens (2011). Sie erforschten, ob die Nähe zur Landesgrenze für das Verstehen der Sprache eines Nachbarlandes eine Rolle spielt. Dazu untersuchten sie drei Fallbeispiele: Das Verstehen des Niederdeutschen in den Niederlanden und Dänemark, das Verstehen des Schwedischen in Norwegen und Dänemark sowie das Verstehen des Dänischen in Norwegen und Schweden. Kürschner und Gooskens kamen dabei zu sehr unterschiedlichen Resultaten, wobei sie dies dem Studiendesign und den *« ... daraus resultierenden methodischen Schwierigkeiten ... »* anlasteten (Kürschner und Gooskens, 2011, S. 182). Die linguistische Untersuchung von scharfen Grenzen auf ihren Einfluss auf die Dialekte ist aber keineswegs neu. Bereits 1898 stellte Haag (1898) fest, dass sprachliche Grenzen oft mit politischen zusammenfallen (zit. in Pickl, 2013). Wie Pickl (2013) führte Haag dies auf ihre Wirkung als Kontaktbarrieren zurück und vermutete, dass die ältesten Grenzen auch die grössten Spuren in der Dialektlandschaft des Neckar- und Donaulandes aufzeigen würden. Besonders gross ist laut Haag der Einfluss von politischen Grenzen, wenn diese mit topographischen (bspw. Berge oder grosse Flüsse) und konfessionellen Grenzen zusammentreffen

(Haag, 1898 zit. in Pickl, 2013).

Neben den zahlreichen Studien zu scharfen Grenzen ist auch der Einfluss der euklidischen Distanzen, den Reisezeiten und der damit zusammenhängenden Topographie auf Dialektunterschiede oft untersucht worden. So versuchte Gooskens (2004) aufzuzeigen, welchen Einfluss Reisezeiten auf Dialektunterschiede in Norwegen haben. Dazu verwendete sie das Korrelationsmass von Pearson und analysierte damit den Zusammenhang von den Dialektunterschieden und alten resp. neuen Reisezeiten zwischen 15 Ortschaften. Die alten Reisezeiten berechnete Gooskens für das Jahr 1900 aufgrund alter Dampfschiff- und Zugfahrpläne sowie der Hinzunahme von Schätzungen über die Reisezeiten zu Pferd. Die modernen Reisezeiten berechnete Gooskens über die Strassennetzdistancen der Ortschaften und den daraus geschätzten Reisezeiten mit dem Auto. Sie konnte aufdecken, dass Orte, die reisezeitmässig nah beieinander liegen, auch eine grössere Wahrscheinlichkeit für eine linguistische Nähe aufweisen. Dabei wiesen die alten Reisezeiten eine höhere Korrelation auf ( $r = 0.76$ ) als die modernen Reisezeiten ( $r = 0.54$ ). Vergleichbare Analysen machten Jeszenszky und Weibel (2014) für die Schweiz. Sie untersuchten, ob euklidische Distanzen sowie Reisezeiten mit dem Auto aus den Jahren 1950 und 2000 mit den syntaktischen Unterschieden innerhalb der Deutschschweiz korrelieren und ob deren Korrelationsunterschiede signifikant sind. Dabei führten sie ihre Analysen – anders als Gooskens (2004) – auf einem globalen *und* einem lokalen Massstab durch. Global bedeutet dabei, dass die Korrelation mit allen vorhandenen Datenpunkten berechnet wurde. Gemäss Jeszenszky und Weibel (2014) ist das Problem bei globalen Analysen zum einen, dass man dabei kleinräumige Unterschiede - wie sie bei Dialekten üblich sind - nicht beachtet. Zum anderen besteht bei einer globalen Untersuchung die Gefahr des ökologischen Fehlschlusses<sup>1</sup>. Aus diesem Grund führten Jeszenszky und Weibel (2014) neben globalen auch lokale Korrelationsanalysen durch. Sie konnten zum einen für den globalen als auch den lokalen Massstab hohe Korrelationen und signifikante Unterschiede zwischen den drei Distanzmassen feststellen. Zum anderen zeigte sich auch, dass die Korrelationswerte stark vom lokalen oder globalen Untersuchungsmassstab abhängig sind (Jeszenszky und Weibel, 2014). Auch Heeringa und Nerbonne untersuchten den Zusammenhang zwischen geographischer Nähe und linguistischer Distanz. In ihrer Arbeit *«Dialect areas and dialect continua»* (Heeringa und Nerbonne, 2001) untersuchten sie mittels *einfachen* Regressionsanalysen<sup>2</sup>, wie viel von der phonologischen Variation Niederländischer Dialekte (d.h. andere Aussprache von Wörtern) mit euklidischer Distanz zu erklären ist. Sie konnten aufzeigen, dass zwischen der phonologischen Distanz und dem Logarithmus der euklidischen Distanz eine starke Korrelation besteht und damit 81% der phonologischen Variation erklärt werden kann.

Ähnlich wie Heeringa und Nerbonne untersuchte auch Szmrecsanyi (2011) den Zusammenhang zwischen den syntaktischen Distanzen und räumlicher Nähe von Ortschaften, und zwar innerhalb Grossbritanniens. Neben den euklidischen Distanzen sowie den Reisezeiten per Auto (welche Szmrecsanyi mit Google Maps berechnete) deckte er vor allem bei Trudgills Gravitations-Index<sup>3</sup> starke Korrelationen auf. Gemäss Szmrecsanyi ist der Zusammenhang zwi-

<sup>1</sup> Ein ökologischer Fehlschluss besteht, wenn man vom Kollektiv auf ein Individuum schliesst.

<sup>2</sup> Zur Erklärung von Regressionsmodellen siehe Abschnitt Unterabschnitt 5.3.4

<sup>3</sup> Zur Erklärung des Gravitations-Indexes siehe Unterabschnitt 5.1.2

schen geographischer Nähe und der Ähnlichkeit von Dialekten aber nicht immer gegeben. In seiner Arbeit «*Geography is overrated*» (Szmrecsanyi, 2012) widersprach er Waldo Toblers erstem Gesetz der Geographie «*Everything is related to everything else, but near things are more related than distant things.*» (Tobler, 1970) und schrieb, dass dieses für Dialekte nicht gelte. Toblers erstes Gesetz der Geographie wurde erst 2007 von Nerbonne und Kleiweg (2007) unter dem Namen *Fundamental Dialectological Postulate*<sup>4</sup> (FDP) für die Linguistik adaptiert. Gemäss Szmrecsanyi (2012) ist geographische Nähe aber ein schlechter Prädikator für syntaktische Ähnlichkeit. Mittels *einfachen* linearen Regressionsanalysen konnte er für Trudgills Gravitations-Index nur ca. 24% der syntaktischen Varianz zwischen den Ortschaften erklären. Für die euklidischen Distanzen und die Reisezeiten per Auto waren es gar nur 4% resp. 8%. Szmrecsanyi (2012, S. 226) schreibt dazu: „*It is fair to say that the FDP has failed (...)*“. Szmrecsanyi hat für seine abweichenden Resultate im Vergleich zu anderen Studien, welche klare Zusammenhänge zwischen Dialektähnlichkeit und geographischer Distanz aufzeigen konnten, drei Erklärungen: Erstens, aber wenig wahrscheinlich, könne es sein, dass sich die syntaktische Variabilität der britischen Dialekte von anderen Dialekten unterscheide. Zweitens - und wahrscheinlicher - sei es möglich, dass sich die (von Szmrecsanyi untersuchte), *syntaktische* Variabilität stark von der *lexikalischen* oder *phonologischen* Variabilität unterscheide, welche andere Studien untersuchten (z.B. in der weiter oben beschriebenen Studie von Heeringa und Nerbonne, 2001). Drittens, und laut Szmrecsanyi auch der wahrscheinlichste Grund für seine diskrepanten Ergebnisse, ist, dass seine verwendeten Sprachdaten nicht vorgängig aufgrund geographischer Verteilung ausgesucht wurden. Dies stände im Gegensatz zu den meisten anderen Sprachdaten, welche Linguisten verwenden würden (Szmrecsanyi, 2012).

Analog zu Szmrecsanyi verwendete auch Auer et al. (2011) Regressionsmodelle um Dialektvarietäten zu untersuchen. Im Gegensatz zu Szmrecsanyi, welcher mit der räumlichen Distanz jeweils nur *eine* erklärende Variable in seine Regressionsmodelle einfliessen liess, verwendete Auer aber *mehrere* erklärende Variablen pro Regressionsmodell. Neben der euklidischen Distanz waren dies zum Beispiel das Geschlecht der Probanden, die morphologische Komplexität einer Wortform und weitere linguistische Faktoren. Ähnlich wie bei Szmrecsanyi (2012) führte bei Auer et al. (2011) die Zunahme der euklidischen Distanzen zu einer Abnahme der von den Regressionsmodellen erklärten Dialektunterschiede. Im Gegensatz zu Szmrecsanyi verzichtete Auer aber auf mögliche Erklärungen weshalb die räumliche Distanz ein schlechter Prädikator für sprachliche Unterschiede sei. Den Ansatz von Auer et al. (2011), *ein* Modell mit *mehreren* erklärenden Variablen zur Erklärung sprachlicher Unterschiede zu erstellen, brachte bereits Gooskens (2004, S. 9) auf. Im Ausblick ihrer Arbeit «*Norwegian dialect distances geographically explained*» schrieb sie:

---

<sup>4</sup> «*Geographically proximate varieties tend to be more similar than distant ones.*»

---

*„It would be instructive to incorporate the above-mentioned geographic, demographic and attitudinal factors into a model for predicting linguistic distances between dialects. This might lead to a greater understanding of mechanisms involved in dialectal variation.“*

Da die Sprache von diversen Faktoren beeinflusst wird, kann gemäss Gooskens Aussage ein Modell, das mehrere erklärende geographische und demographische Variablen gleichzeitig mit einbezieht, zu einem verbessertem Verständnis der Sprachunterschiede führen. Dieser Ansatz soll in der vorliegenden Arbeit weiterverfolgt werden. Im nachfolgenden Kapitel werden die zugehörigen Forschungsfragen vorgestellt.



## Kapitel 3

# Forschungslücken und Ziele

Aufgrund der in Abschnitt 2 beschriebenen Arbeiten kristallisieren sich folgende Forschungsfragen heraus. Die Beschreibung der Forschungslücke fasst nochmals die wichtigsten Punkte zusammen, die zur jeweiligen Forschungsfrage (FF) führen. Zusätzlich sind noch die daraus folgenden drei Ziele dieser Arbeit formuliert, die aus der Beantwortung der Forschungsfragen hervorgehen.

### 3.1 FF 1: Haben politische, religiöse und wirtschaftliche Grenzen einen Einfluss auf sprachliche Unterschiede?

**Forschungslücke:** Es ist noch immer unklar, welche Faktoren für die Veränderungen von Sprachen verantwortlich sind (Bowerman und Evans, 2015). Der Einfluss von diskreten Grenzen als vermutete Sprachbarrieren ist in der Linguistik im Allgemeinen und in der Schweiz im Speziellen noch ungenügend erforscht (vgl. dazu die Arbeit von Pickl, 2013). Bisherige Forschung basierte hauptsächlich auf phonologischen oder lexikalischen Daten, die grammatikalische Variation der Dialekte wurde bis anhin vernachlässigt (Glaser, 2013). Da die Grammatik die sich am langsamsten verändernde Ebene einer Sprache ist (Thomason, 2001), kann man annehmen, dass bei einem nachweislichen Einfluss von Grenzen auf die Syntax diese soziokulturell auch wirklich wichtig sind. Weiter hilft die Untersuchung über den Einfluss von vermutete Sprachbarrieren auf syntaktische Unterschiede die Evolution der Dialekte besser zu verstehen und leistet einen wichtigen Beitrag zur aktuellen Diskussion über die Veränderung von Sprache durch Kontakt.

**Erstes Ziel** der Masterarbeit ist es, den Einfluss von diskreten Grenzen als vermuteten Sprachbarrieren auf die Verteilung schweizerdeutscher Dialekte zu untersuchen und mit den Erkenntnissen einen Beitrag innerhalb der linguistischen Forschung zu leisten.

## 3.2 FF 2: Haben geographische Distanzen einen Einfluss auf sprachliche Unterschiede?

**Forschungslücke:** Ob und wie stark die geographische Distanz einen Einfluss auf Dialektunterschiede hat, darüber herrscht bis heute eine Debatte in der geolinguistischen Forschung. Das von Waldo Tobler (1970) proklamierte erste Gesetz der Geographie «*Everything is related to everything else, but near things are more related than distant things.*», wurde erst 2007 von Nerbonne und Kleiweg (2007) unter dem Namen *Fundamental Dialectological Postulate* für die Linguistik adaptiert. Dieses wurde aber von Benedikt Szendrői (2012) in Frage gestellt. In seiner Studie über den Einfluss von geographischer Distanz auf die morphosyntaktische Variabilität britischer Dialekte kam er zum Ergebnis, dass diese ein schlechter Prädiktor sei.

Obwohl bereits Jeszenszky und Weibel (2014) mit denselben Daten, welche auch in der vorliegenden Arbeit verwendet werden, Zusammenhänge von geographischer Distanz und Dialektunterschiede aufzeigen konnten, lässt sich eine nochmalige Untersuchung mit folgenden drei Punkten rechtfertigen. Erstens hat der Autor dieser Arbeit keinen Zugriff auf die von Jeszenszky und Weibel (2014) berechneten Distanzmasse. Zweitens wird in der vorliegenden Arbeit ein einfacher zu berechnendes Mass für die Reisezeit-Distanzen verwendet (vgl. Unterabschnitt 5.1.1). Wenn sprachliche Unterschiede mit den hier verwendeten, einfach zu berechnenden Wanderzeiten ähnlich oder besser erklärt werden können wie mit den komplizierteren, von Jeszenszky und Weibel (2014) berechneten historischen resp. neuzeitlichen Autoreisezeiten, stellt sich die Frage nach dem Sinn von komplexen Reisezeitberechnungen. Drittens sollen die hier verwendeten Distanzmasse zusammen mit den scharfen Grenzen als Folgeziel in ein grösseres Modell miteinbezogen werden. Die Berechnung eigener Distanzmasse geschieht deshalb auch aus Gründen der Konsistenz.

**Zweites Ziel** der Arbeit ist es, den Einfluss des geographischen Raumes in Form verschiedener Distanzmasse auf die sprachlichen Unterschiede zu untersuchen.

## 3.3 FF 3: Können syntaktische Unterschiede in einem Modell mit mehreren scharfen und unscharfen Grenzen erklärt werden?

**Forschungslücke:** Die präzise Modellierung und die daraus folgenden flächenhaften Visualisierungen der Dialektunterschiede sind sowohl für die Sprachwissenschaften als auch für andere Wissenschaftszweige, wie z.B. die Wirtschaftsgeographie, wichtig. Solide, verlässliche Karten helfen, räumliche Prozesse und zugrunde liegende Strukturen innerhalb der Dialektverteilung zu identifizieren. Modelle, welche Dialektunterschiede erklären, können aber auch behilflich sein beim Verstehen von Mechanismen des wirtschaftlichen Austauschs und der regionalen Arbeitsmigration (vgl. Kapitel 1).

Zwar wurden, wie in Kapitel 2 beschrieben, schon einige Versuche unternommen um Dialekt-

unterschiede mit Modellen zu erklären, doch beschränkten sich die bisher verwendeten Regressionsmodelle auf jeweils *ein* geographisches Distanzmass. Auch der Einfluss von unterschiedlichen scharfen Grenzen auf Dialektunterschiede wurde schon untersucht. Die unterschiedlichen Forschungsergebnisse und die geringe Anzahl bisheriger Studien führen aber zu dem, dass die für die Sprachunterschiede verantwortlichen Faktoren noch immer unklar sind. Zum anderen wurden bis anhin nur Regressionsmodelle verwendet, welche die Dialektunterschiede mit *einem* erklärendem Faktor zu erklären versuchten. Obwohl bereits von Gooskens (2004) vorgeschlagen (vgl. letzter Abschnitt von Kapitel 2), wurde noch nie ein Regressionsmodell erstellt, welches gleichzeitig mehrere scharfe und unscharfe geographische Grenzen einbindet. Als scharfe Grenzen dienen in der vorliegenden Arbeit politische, religiöse und wirtschaftliche Grenzen; als unscharfe Grenzen werden verschiedene Distanzmasse verwendet. Mit einem Multikomponenten-Modell sollen die Dialektunterschiede zwischen Ortschaften möglichst exakt erklärt werden.

**Drittes Ziel** der Masterarbeit ist es, ein Modell zu erstellen, das die syntaktischen Unterschiede zwischen Ortschaften möglichst gut erklärt. Dazu sollen erstmalig mehrere scharfe und unscharfe Grenzen in einem Modell zusammengefasst werden. Durch die damit gewonnenen Erkenntnisse über den Einfluss von Grenzen und strukturellen Kontakthindernissen auf die syntaktischen Unterschiede wird es später möglich sein, mit einer durch das Modell angepassten Interpolation die sprachliche Realität der Länder möglichst exakt zu erklären und flächenhaft abzubilden.



# Kapitel 4

## Daten

### 4.1 Hintergrundwissen

In der Linguistik werden traditionell folgende Ebenen innerhalb einer Sprache unterschieden: Phonetik, Lexik, Morphologie und Syntax. Die Phonetik oder auch die Phonologie beschäftigt sich mit den Lauten einer Sprache und untersucht, wie diese erzeugt und ausgesprochen werden (Lyons, 1968). Die Lexik ist die Lehre des Wortschatzes und zeigt sich im Alltag darin, welche Wörter Sprechende verwenden, um Dinge zu bezeichnen. Die Morphologie studiert die interne Struktur von Wörtern und deren Bildung, wohingegen die Syntax sich mit den Regeln der Wortkombination und deren Reihenfolge innerhalb von Sätzen auseinandersetzt (Lyons, 1968; O’Grady, Dobrovolsky und Katamba, 1997). Die vorliegende Arbeit verwendet als Datensatz ebensolche Syntax-Daten. Zwei Gründe sprechen für die Benutzung von Syntax-Daten. Erstens basierten bisherige linguistische Studien hauptsächlich auf phonologischen oder lexikalischen Sprachdaten. Syntax-Untersuchungen wurden bisher nur wenige durchgeführt (Glaser, 2013). Zweitens ergibt es Sinn, dass wenn man den Einfluss des Raumes auf Sprachunterschiede untersuchen möchte, möglichst «zeitstabile» Sprachdaten verwendet. Syntax-Daten eignen sich deshalb besonders gut, weil die Syntax von allen Sprachebenen diejenige ist, welche sich im Verlauf der Zeit am langsamsten verändert und somit die «stabilste» aller Sprachebenen darstellt (neben bspw. der Phonetik oder Lexik) (vgl. Thomason, 2001). Bei einem nachweislichen Einfluss von Grenzen auf die Syntaxunterschiede kann man dadurch auch annehmen, dass die Grenzen soziokulturell wirklich wichtig sind.

### 4.2 Verwendete Daten

Als Datensatz dient der Syntaktische Atlas der Deutschen Schweiz (SADS). Im Rahmen eines Nationalfondprojekts wurde im Jahr 2000 eine gross angelegte schriftliche Befragung gestartet, welche 3’187 Personen aus 383 deutschschweizer Gemeinden umfasste. Die 383 Gemeinden entsprechen etwa 25% aller deutschsprachigen Gemeinden in der Schweiz (Abb. 4.1). Die Probanden und mindestens einer ihrer Elternteile mussten im Ort aufgewachsen sein. Die teilnehmenden

Personen, welche aus allen sozialen Schichten, Alters- und Berufsgruppen stammen, beantworteten insgesamt 118 Fragen. Es handelt sich dabei um Übersetzungs- (vom Standarddeutschen), Ergänzungs- und Ankreuzfragen, wodurch 54 syntaktische Variablen der deutschschweizer Syntax erfasst wurden. Die 118 Fragen wurden in vier Fragebögen aufgeteilt und halbjährlich in den Jahren 2000-2002 an die Probanden verschickt. Abbildung 4.2 zeigt je ein Beispiel einer Übersetzungs-, Ergänzungs- und Ankreuzfrage aus dem ersten der insgesamt vier Fragebögen, welche die Probanden schriftlich und von Hand ausgefüllt haben. Von den 3'187 Personen, welche den ersten Fragebogen ausfüllten, haben 2'771 Probanden alle vier zurückgeschickt. Für die Validierung der Daten und zum Ausgleichen der Unsicherheit der schriftlichen Befragung wurden an einzelnen Orten zum einen mündliche Befragungen durchgeführt, zum anderen wurden an jedem der Erhebungsorte mehrere Personen befragt. Die 383 Untersuchungsorte sind mit durchschnittlich fünf bis sechs Probanden vertreten<sup>1</sup>, wobei die Anzahl Teilnehmenden pro Ort zwischen drei und 26 Personen schwankt. An 90% der Orte haben fünf bis zehn Probanden alle vier Fragebögen ausgefüllt. Von den 118 Fragen sind zurzeit 68 bereinigt und zur Analyse bereit. Diese räumlich referenzierten, nominalen Sprachdaten erlauben die Erforschung von syntaktischen Eigenheiten schweizerdeutscher Dialekte sowie deren räumliche Analyse (Bucheli und Glaser, 2002; Glaser und Bart, 2015; Universität Zürich, 2016b).

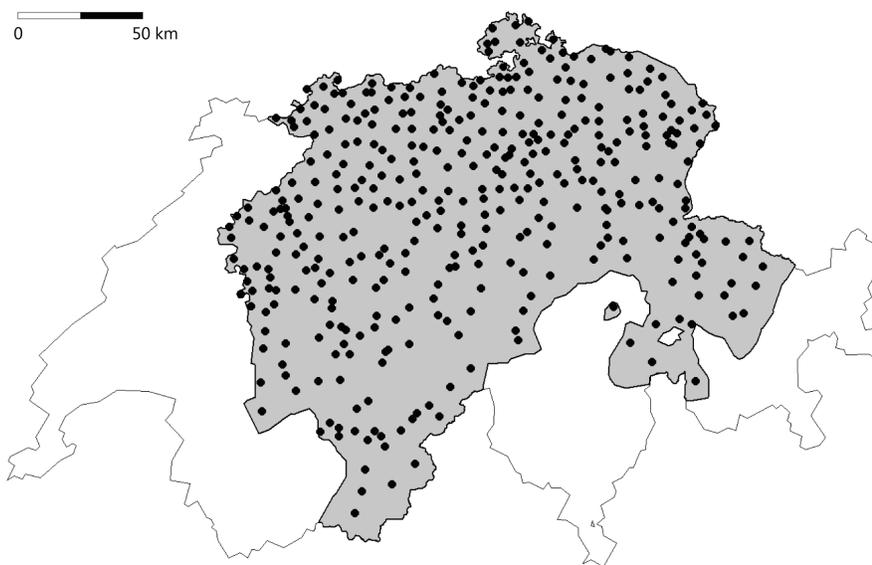


Abbildung 4.1: Alle 383 Orte in der deutschsprachigen Schweiz (grau schraffiert), an denen die Daten des SADS erhoben wurden.

---

<sup>1</sup>Es handelt sich dabei um den Median.

1. Sie müssen an einer Bahnstation ohne Schalter ein Billett kaufen. Vor dem Automaten merken Sie, dass Sie zu wenig Kleingeld haben. Sie sprechen deshalb eine Passantin an:

→ Bitte übersetzen Sie den folgenden Satz in Ihren Dialekt und schreiben Sie ihn so auf, wie Sie ihn sagen würden:

Entschuldigung, ich habe zu wenig Kleingeld, um ein Billett zu lösen.

\_\_\_\_\_

\_\_\_\_\_

---

4. Sie rufen Ihre Nachbarin an, um ihr das Neueste zu erzählen. Der Sohn nimmt ab. Sie sagen ihm, dass Sie mit seiner Mutter sprechen wollen. Er antwortet:

→ Vervollständigen Sie den Antwortsatz; er soll Auskunft darüber geben, wo die Nachbarin ist („einkaufen“):

Oh, si isch nid da, si isch \_\_\_\_\_

---

11. Nach einem anstrengenden Tag machen Sie es sich auf dem Sofa bequem, hören Ihre Lieblingsmusik und wollen den spannenden Roman „Wüstenblume“ zu Ende lesen. Da klingelt das Telefon, Ihre Mutter ist dran und will, dass Sie vorbeikommen. Sie antworten ihr ausweichend:

→ Welche der folgenden Sätze können Sie in Ihrem Dialekt sagen ("ja"), welche sind nicht möglich ("nein")?

	ja	nein	
1)	<input type="checkbox"/>	<input type="checkbox"/>	Aber jetzt bin i grad aneghocket <b>für</b> es Buech <b>z</b> läse.
2)	<input type="checkbox"/>	<input type="checkbox"/>	Aber jetzt bin i grad aneghocket <b>für</b> es Buech läse.
3)	<input type="checkbox"/>	<input type="checkbox"/>	Aber jetzt bin i grad aneghocket es Buech <b>z</b> läse.
4)	<input type="checkbox"/>	<input type="checkbox"/>	Aber jetzt bin i grad aneghocket <b>zum</b> es Buech läse.
5)	<input type="checkbox"/>	<input type="checkbox"/>	Aber jetzt bin i grad aneghocket <b>zum</b> es Buech <b>z</b> läse.
6)	<input type="checkbox"/>	<input type="checkbox"/>	Aber jetzt bin i grad aneghocket <b>um</b> es Buech <b>z</b> läse.
7)	<input type="checkbox"/>	<input type="checkbox"/>	Aber jetzt bin i grad aneghocket <b>für zum</b> es Buech <b>z</b> läse.
8)	<input type="checkbox"/>	<input type="checkbox"/>	Aber jetzt bin i grad aneghocket <b>für zum</b> es Buech läse.

→ Welche Variante ist für Sie die natürlichste?  
Nr. \_\_\_\_

→ Würden Sie den Satz normalerweise in einer Form sagen, die nicht aufgeführt ist?  
 ja  nein

→ Wenn "ja": Bitte notieren Sie hier den Satz so, wie Sie ihn normalerweise sagen würden:

\_\_\_\_\_

\_\_\_\_\_

Abbildung 4.2: Beispielhaft je eine Übersetzungs-, Ergänzungs- und Ankreuzfrage aus dem ersten von insgesamt vier Fragebögen der SADS-Erhebung, wie sie die Probanden erhalten haben.

### 4.2.1 Die syntaktische Distanz

In sprachwissenschaftlichen Studien werden Sprachcharakteristika mittels kategorialen Daten erfasst. Damit die Unterschiede zwischen Sprachen und Dialekten gemessen werden können, wurde in der dialektometrischen Forschung bereits eine Vielzahl von Messmethoden entwickelt und bspw. auch aus der Mathematik adaptiert. Die Herausforderung bei der Entwicklung eines Masses für die sprachlichen Unterschiede ist stets, aus den kategorialen Sprachdaten ein numerisches Distanzmass berechnen zu können. Beispiele dafür sind die Hamming- und die Levenshtein-Distanz (Levenshtein, 1966), wobei letztere bspw. in Arbeiten von Nerbonne et al. (1996) und Gooskens (2004) verwendet wurde. Die (Un-)ähnlichkeiten der Syntax werden laut Jeszenszky et al. (angenommen) traditionell mit der *Hamming-Distanz* (Hamming, 1950) oder dem *gewichteten Identitätswert* von Goebel (1982) berechnet<sup>2</sup>. Diese Masse bestimmen die syntaktische Differenz zweier Erhebungsorte bezogen auf je *einer* syntaktischen Variante pro Erhebungsort. Da im SADS jeweils *mehrere* syntaktische Varianten pro Erhebungsort erfasst sind, eignen sich die bekannten Distanzmasse nicht für die Daten des SADS (Jeszenszky et al., angenommen). Jeszenszky et al. berechneten deshalb die syntaktische Distanz zwischen jeweils zweier Ortschaften in mehreren Schritten: In einem ersten Schritt berechneten sie für jede Variante das Verhältnis zwischen der Anzahl Nennungen einer Variante und der Anzahl Probanden im jeweiligen Ort. *Varianten* sind die von Linguisten zusammengefassten Varietäten der deutschschweizer Syntax, die aus den Antworten der Probanden des SADS stammen. Für die beispielhafte erste Frage aus Abbildung 4.2 wären dies die zwei Varianten *...für es Billet z'lööse.* und *...um es Billet z'lööse.* In einem zweiten Schritt berechneten Jeszenszky et al. von jeder Variante die Differenzen der berechneten Verhältnisse zwischen den zwei Orten. Drittens und letztens addierten sie die im zweiten Schritt berechneten Differenzen aller Varianten von allen Fragen zusammen. Das Resultat ist die syntaktische Distanz zwischen den zwei betrachteten Ortschaften. Abbildung 4.3 zeigt eine Beispielberechnung der syntaktischen Distanz zwischen den Orten Klosters und Flühli mit zwei Fragen aus dem SADS. Wenn bspw. in Klosters neun Probanden an der SADS-Erhebung teilgenommen haben, dann haben in unserem Beispiel die Übersetzungsfrage Q1.01 „*Ich habe zu wenig Kleingeld um ein Billet zu lösen.*“ sechs der neun Probanden die Variante „*...für es Billet (z)lööse.*“ hingeschrieben und drei der neun Probanden die Variante „*...zum es Billet (z)lööse.*“. Das Verhältnis der beiden Varianten der Frage Q1.01 beträgt für Klosters also 0,66 zu 0,33. Genauso werden auch die Verhältnisse der Varianten für Flühli berechnet und danach die Differenzen der Varianten-Verhältnisse beider Orte ermittelt. Jeszenszky et al. berechneten auf diese Art die Differenzen der Varianten-Verhältnisse für 60 der zurzeit 68 bereinigten Fragen des SADS und summierten diese anschliessend für jedes der 383 SADS-Ortspaare auf<sup>3</sup>. Die aufsummierten Differenzen ergeben die syntaktischen Distanzen der SADS-Erhebungsorte und sind somit ein Mass für deren syntaktische (Un-)ähnlichkeit. Je grösser die syntaktische Distanz zwischen zwei Ortschaften ist, desto weiter sind diese syntaktisch

<sup>2</sup> vgl. z.B. Nerbonne und Wiersma, 2006; Spruit, 2006; Spruit, Heeringa und Nerbonne, 2009

<sup>3</sup>  $383^2 = 146'689$  Ortspaare, wobei die jeweiligen berechneten syntaktischen Distanzen zwischen einem Ortspaar *A-B* und *B-A* symmetrisch sind.

voneinander „entfernt“. Je kleiner die syntaktische Distanz, desto syntaktisch ähnlicher sind sich die Dialekte zweier Ortschaften. In dieser Arbeit werden die von Jeszenszky et al. (angenommen) berechneten syntaktischen Distanzen zwischen den SADS-Erhebungsorten verwendet.

1. Nimm zwei Erhebungsorte und berechne das Verhältnis der Varianten für eine gegebene Variable. Berechne dann die Differenz der jeweiligen Varianten zwischen den Erhebungsorten.	<p><b>Q1.01</b> Ich habe zu wenig Kleingeld, <b>um</b> ein Billet <b>zu</b> lösen.</p> <p style="text-align: center;"><i>... für es Billet (z) lööse.    ... zum es Billet (z) lööse.</i></p> <table border="0" style="width: 100%;"> <tr> <td style="width: 60%;">Erhebungsort I. „Klosters“</td> <td style="width: 20%; text-align: center;">0.66*</td> <td style="width: 20%; text-align: center;">0.33</td> </tr> <tr> <td>Erhebungsort II. „Flühli“</td> <td style="text-align: center;">0.2</td> <td style="text-align: center;">0.8</td> </tr> <tr style="background-color: #e0e0e0;"> <td>Differenz:</td> <td style="text-align: center;">0.46</td> <td style="text-align: center;">0.47</td> </tr> </table> <p style="text-align: center;">- - - - -</p>	Erhebungsort I. „Klosters“	0.66*	0.33	Erhebungsort II. „Flühli“	0.2	0.8	Differenz:	0.46	0.47
Erhebungsort I. „Klosters“	0.66*	0.33								
Erhebungsort II. „Flühli“	0.2	0.8								
Differenz:	0.46	0.47								
2. Führe obigen Schritt für alle Variablen durch.	<p><b>Q1.03</b> Ich habe den Fritz <b>kommen hören</b>.</p> <p style="text-align: center;"><i>Ich ha de Fritz ghöört choo. Ich ha de Fritz choo ghöört.</i></p> <table border="0" style="width: 100%;"> <tr> <td style="width: 60%;">Erhebungsort I. „Klosters“</td> <td style="width: 20%; text-align: center;">0.2</td> <td style="width: 20%; text-align: center;">0.8</td> </tr> <tr> <td>Erhebungsort II. „Flühli“</td> <td style="text-align: center;">0.4</td> <td style="text-align: center;">0.6</td> </tr> <tr style="background-color: #e0e0e0;"> <td>Differenz:</td> <td style="text-align: center;">0.2</td> <td style="text-align: center;">0.2</td> </tr> </table>	Erhebungsort I. „Klosters“	0.2	0.8	Erhebungsort II. „Flühli“	0.4	0.6	Differenz:	0.2	0.2
Erhebungsort I. „Klosters“	0.2	0.8								
Erhebungsort II. „Flühli“	0.4	0.6								
Differenz:	0.2	0.2								
3. Addiere alle erhaltenen Differenzen der beiden Erhebungsorte um die „syntaktische Distanz“ zwischen ihnen zu erhalten.	<hr style="border: 1px solid black;"/> <p><b>∑ Differenz:</b>    0.46 + 0.47 + 0.2 + 0.2 = <b><u>1.33</u></b>    (= syntaktische Distanz zwischen Klosters und Flühli)</p>									

Abbildung 4.3: Beispielhafte Berechnung der syntaktischen Distanz zwischen Klosters und Flühli mit zwei der insgesamt 68 verwendeten SADS-Variablen. Bildquelle: modifiziert nach Jeszenszky et al. (angenommen).



# Kapitel 5

## Methodik

Der methodische Teil dieser Arbeit unterteilt sich in zwei Hauptteile. **Im ersten Hauptteil** werden in den Abschnitten 5.1 und 5.2 die Berechnungen der verschiedenen *unscharfen* Grenzen resp. die verwendeten *scharfen* Grenzen vorgestellt. Unscharfe Grenzen zwischen Ortschaften ergeben sich zum einen durch die Erreichbarkeit, welche von der Topographie und der Distanz beeinflusst wird (Unterabschnitt 5.1.1), zum anderen durch den Einflussbereich der Orte. Dieser ist abhängig von der Bevölkerungszahl der Orte und der Entfernung zwischen ihnen (Unterabschnitt 5.1.2). Politische, religiöse und wirtschaftliche Grenzen zwischen den Ortschaften dienen als scharfe, diskrete Grenzen. Als politische Grenzen werden die Kantonsgrenzen untersucht. Sie prägen die föderalistisch aufgebaute Schweiz seit ihren Ursprüngen (Unterabschnitt 5.2.1). Die Konfessionsgrenzen zwischen katholischen und protestantischen Gebieten bilden die religiösen Grenzen der Schweiz. Sie haben ihre Ursprünge in der Reformation der katholischen Kirche, welche das gesellschaftliche Zusammenleben der Schweiz ebenfalls stark prägte (Unterabschnitt 5.2.2). Als wirtschaftliche Grenzen werden die Arbeitsmarktregionsgrenzen behandelt, die ein Abbild der wirtschaftlichen Verknüpfungen innerhalb der Schweiz darstellen. Zwar sind sie eine rein raumplanerische Konstruktion, haben ihren Ursprung aber in statistisch erfassten Pendlereinzugsgebieten der Schweiz (Unterabschnitt 5.2.3). **Im zweiten Hauptteil** 5.3 folgt die Beschreibung der verwendeten Analysemethoden, mit denen die Einflüsse der Grenzen auf die syntaktischen Unterschiede der Dialekte untersucht werden. Aus zwei Gründen erfolgen alle Analysen der unscharfen und scharfen Grenzen mit acht verschiedenen Masstabsebenen: Zum einen um den Einfluss des MAUP (Modifiable Areal Unit Problem) zu berücksichtigen (vgl. Openshaw und Taylor, 1979), zum anderen um den Raumeinfluss auf die syntaktischen Distanzen besser analysieren zu können (Unterabschnitt 5.3.1). Unterabschnitt 5.3.2 zeigt auf, wie mittels Korrelationsanalysen und Dominanzkarten der Einfluss der *unscharfen* Grenzen auf die syntaktischen Distanzen der Ortschaften analysiert wird. Die *scharfen* Grenzen werden mit statistischen Mittelwertstests sowie Cohens  $d$ , einem Mass der Effektstärke, untersucht (Unterabschnitt 5.3.3). Dies soll aufzeigen, ob sich die syntaktischen Distanzen von Ortspaaren, zwischen denen eine der untersuchten *scharfen* Grenzen liegt, von Ortspaaren, zwischen denen keine Grenze liegt, statistisch signifikant unterscheiden. Unterabschnitt 5.3.4 zeigt mittels li-

nearen Regressionsanalysen schliesslich eine Möglichkeit auf, wie die scharfen und unscharfen Grenzen zusammengefasst und zur Modellierung der syntaktischen Distanz verwendet werden können. Dies alles mit dem Ziel, den Einfluss von Grenzen auf die syntaktischen Unterschiede zwischen Ortschaften zu untersuchen und die Vorarbeit für eine Kriging-Interpolation zu leisten. Abbildung 5.1 zeigt eine Übersicht des methodischen Aufbaus dieser Arbeit.

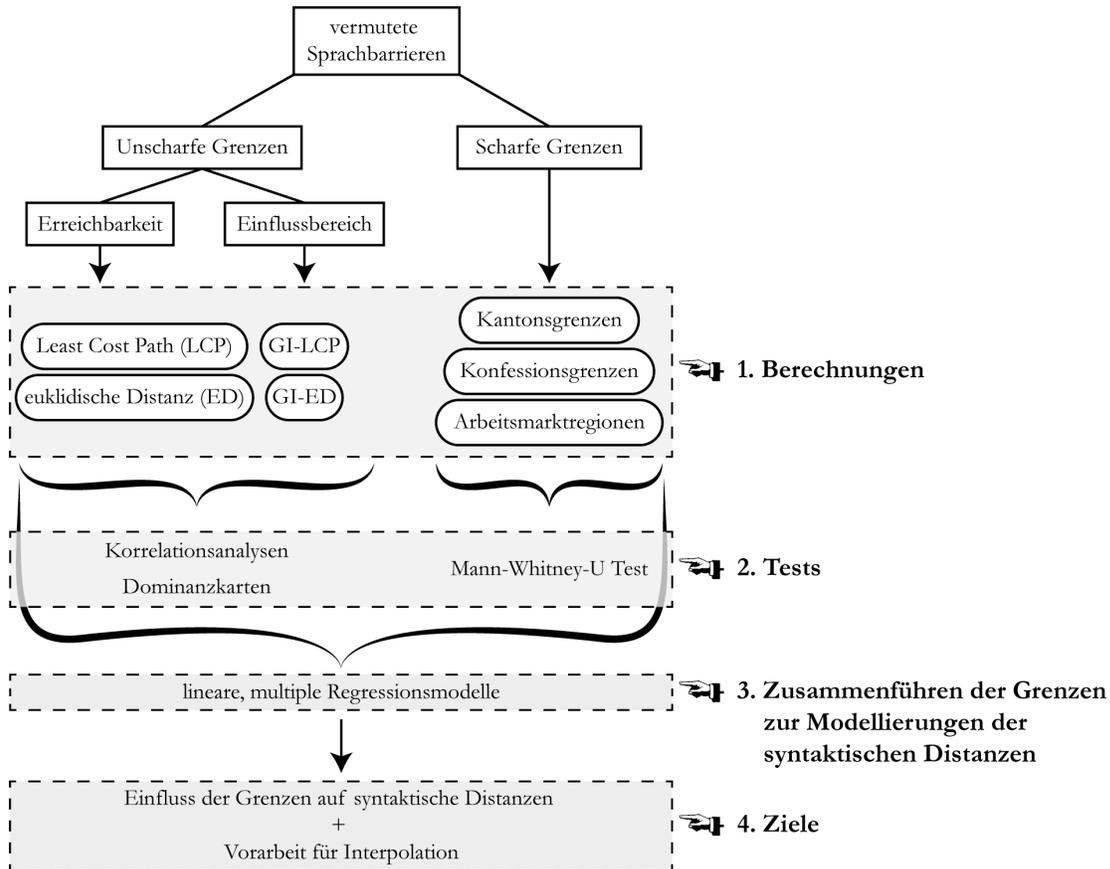


Abbildung 5.1: Übersicht über den methodischen Aufbau dieser Arbeit.

## 5.1 Unscharfe Grenzen: Distanzmasse

Bei den unscharfen Grenzen wird die Erreichbarkeit und der gegenseitige Einflussbereich von Ortschaften untersucht. In nachfolgendem Unterabschnitt 5.1.1 werden zwei unterschiedliche Erreichbarkeitsmasse vorgestellt. In Unterabschnitt 5.1.2 folgt die detaillierte Beschreibung der zwei Einflussbereichsmasse.

### 5.1.1 Die Erreichbarkeit

Es besteht die linguistische Hypothese, dass sich Sprache über den Kontakt zu anderen Menschen verändert und anpasst (vgl. Abschnitt 2.1). Wie oft Bewohner von verschiedenen Ortschaften in Kontakt zueinander stehen und sich somit sprachlich beeinflussen, ist abhängig von der Erreichbarkeit der Orte. Da sich die Syntax einer Sprache nur sehr langsam verändert und die heutige Sprache ein Abbild aus Einflüssen der Vergangenheit ist (vgl. Croft, 2000), bietet es sich an, mit Wanderzeiten die Erreichbarkeit zwischen Ortschaften zu approximieren. Die Fortbewegung zu Fuss stellt das ursprünglichste Fortbewegungsmittel der Menschheit dar, weshalb sich Wanderzeiten als Erreichbarkeitsmass besonders gut eignen. Für die Berechnung der Wanderzeiten zwischen den Ortschaften wird angenommen, dass die Menschen stets den schnellsten Weg zwischen zwei Orten wählen. Zwar ist es gemäss Herzog (2013) unwahrscheinlich, dass Menschen aus dem Altertum stets den schnellstmöglichen Weg gewählt haben, doch kann man durchaus davon ausgehen, dass sie häufig begangene Pfade – wie Handelsrouten oder Wege, um Nachbarorte zu besuchen – zeitlich optimierten. Für die Berechnung der Wanderzeiten zwischen den 383 Ortschaften des SADS wird zuerst mit einem digitalen Höhenmodell (DHM) der Schweiz und *Toblers Hiking* Funktion (Tobler, 1993), welche die Wandergeschwindigkeit in Abhängigkeit zur Neigung des Geländes stellt, eine Kostenoberfläche der Schweiz erstellt (Abschnitt 5.1.1.1). Aus dieser Kostenoberfläche wird mithilfe eines Algorithmus', der automatisiert die schnellsten Pfade findet, die *Least Cost Paths* (LCP, engl. für «kostengünstigster Pfad») berechnet (Abschnitt 5.1.1.2). Ein noch einfacheres Mass für die Erreichbarkeit ist die **euklidische Distanz** zwischen Ortschaften (= Luftlinie) (Abschnitt 5.1.1.3). Sie dient in dieser Arbeit als Vergleichsmass zu den LCP-Distanzen und zur Berechnung eines Topographie-Einflussfaktors auf die Reisezeit, welcher für eine erste Visualisierung über den Einfluss der Topographie auf die syntaktischen Distanzen dient (Abschnitt 5.1.1.4). Nachfolgend wird die Berechnung der zwei Erreichbarkeitsmasse im Detail vorgestellt.

Um die Reisezeiten zwischen den SADS-Orten berechnen zu können, muss zuerst eine Kostenoberfläche für die Schweiz erstellt werden. Eine Kostenoberfläche ist ein Raster, bei dem die Werte einzelner Rasterzellen den Kosten entsprechen, die für das Erreichen dieser Zellen aufgewendet werden müssen. Die Kosten des Wegs von einem Ort *A* nach einem Ort *B* entsprechen den akkumulierten Kosten, die bei der Durchquerung der Rasterzellen von *A* nach *B* aufgewendet werden. Abbildung 5.2 zeigt beispielhaft den Unterschied zwischen dem euklidischen und dem LCP-Reiseweg. Die Kostenoberfläche besteht in diesem Beispiel aus der Geländehö-

he. Ein spezieller Algorithmus<sup>1</sup> berechnet auf Grundlage der Kostenoberfläche automatisch den Weg der tiefsten akkumulierten Kosten (engl.: Least Cost Path).

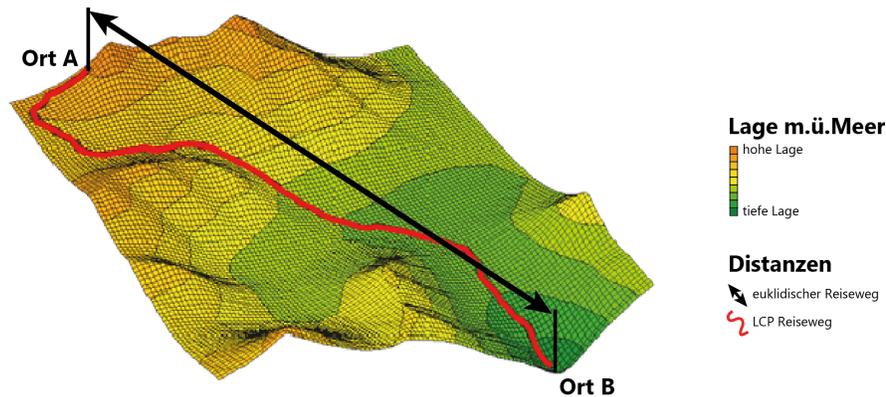


Abbildung 5.2: LCP- und euklidischer Reiseweg auf Grundlage einer Kostenoberfläche. Bildquelle: modifiziert nach Berry (2007)

#### 5.1.1.1 Kostenoberfläche mit Toblers Hiking Funktion

Die in dieser Arbeit verwendete Kostenoberfläche für die Schweiz wird mit *Toblers Hiking Funktion* und einem digitalen Höhenmodell (DHM) der Schweiz berechnet. Dazu wird zuerst das DHM mit einer Auflösung von  $25 \times 25m$  auf eine Auflösung von  $100 \times 100m$  skaliert. In einem zweiten Schritt wird das DHM mit einer 8km-Bufferzone um die *Convex Hull* der SADS-Erhebungsorte zugeschnitten<sup>2</sup> (Abb. 5.3). Die Convex Hull-Fläche entspricht dabei der kleinstmöglichen konvexen Fläche, die alle Ortspunkte umfasst (Preparata und Shamos, 2012). Beide Schritte dienen dazu, bessere Prozessierungsgeschwindigkeiten bei den späteren Berechnungen der Kostenoberfläche und den LCPs zu erhalten. Als dritter und letzter Schritt wird die Kostenoberfläche mit dem zugeschnittenen DHM und Toblers Hiking Funktion auf Grundlage des modifizierten *R-Codes* von van Etten (2012, 2015) berechnet.

<sup>1</sup>z.B. «*Dijkstra*» oder «*A\**»

<sup>2</sup>Würde man anstelle der Convex Hull die Grenze der Deutschschweiz verwenden um das DHM zuzuschneiden, so würde der Algorithmus, der später die LCPs zwischen den Orten berechnet, Abkürzungen über die Grenze und somit Gebiete ohne Höhendaten verwenden. Als Folge wären die berechneten Pfade unbrauchbar.

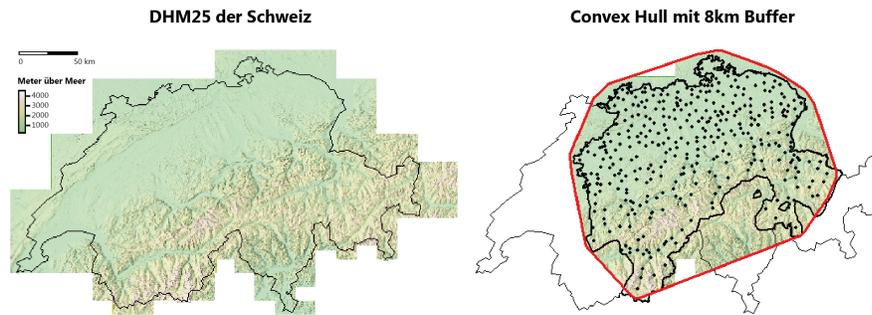


Abbildung 5.3: Das digitale Höhenmodell der Schweiz (linke Seite), welches mit einem 8km Buffer (rote Linie) um die Convex Hull der SADS-Orte zugeschnitten wurde (rechte Seite). Die schwarzen Punkte entsprechen der Lage der SADS-Orte.

**Toblers Hiking Funktion** (Formel 5.1) ist eine von Waldo Tobler (1993) erstellte exponentielle Funktion, die er aus empirischen Daten von Eduard Imhof (1950) herleitete. Sie setzt die Wandergeschwindigkeit in Abhängigkeit zum Gefälle des Geländes:

$$v = \frac{6e^{-3,5|m+0,05|}}{3,6} \quad (5.1)$$

$v$  ist die Gehgeschwindigkeit in Meter pro Sekunde,  $m$  ist die Hangneigung in Grad. D.h.  $m$  ist der Quotient der Höhendifferenz  $\Delta h$  geteilt durch die Distanz  $\Delta x$  zwischen zwei Wegpunkten und entspricht dem Tangens vom Neigungswinkel  $\theta$  des Geländes:

$$m = \frac{\Delta h}{\Delta x} = \tan(\theta) \quad (5.2)$$

Waldo Tobler erstellte die Funktion ursprünglich in  $km/h$ . Da die Masseinheit des verwendeten Höhenmodells sowie die Koordinaten der SADS-Orte der Einheit *Meter* entsprechen, wird die Funktion durch 3,6 geteilt um so die Geschwindigkeit in  $m/s$  zu erhalten. Abbildung 5.4 zeigt Toblers Hiking Funktion.

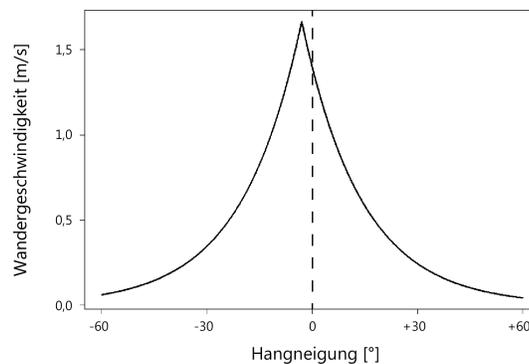


Abbildung 5.4: Die exponentielle Hiking Funktion von Waldo Tobler (1993) zeigt die Abhängigkeit der Geschwindigkeit von der Neigung des Geländes auf.

Es gilt zu beachten, dass die maximale Wandergeschwindigkeit von  $1,67 \text{ m/s}$  ( $\approx 6 \text{ km/h}$ ) nicht im ebenen Gelände ( $m = 0^\circ$ ), sondern bei leichtem Gefälle ( $m = -2,86^\circ$ ) erreicht wird. Die Geschwindigkeit im flachen Terrain beträgt nach Toblers Hiking Funktion  $1,39 \text{ m/s}$  ( $\approx 5 \text{ km/h}$ ).

### 5.1.1.2 LCP Reisezeiten

Aus der in Abschnitt 5.1.1.1 beschriebenen Kostenoberflächen werden nun zwischen allen 383 SADS-Orten paarweise die schnellsten Reisezeiten zu Fuss berechnet. Diese LCP Reisezeiten wurden mit der  $R$ -Funktion *costDistance* (van Etten, 2015) berechnet. Die Funktion *costDistance* verwendet zur Berechnung der LCPs den *Dijkstra*-Algorithmus (Dijkstra, 1959), welcher zwischen zwei Orten  $A$  und  $B$  den Weg berechnet, der am wenigsten «kostet», d.h. in kürzester Reisezeit von  $A$  nach  $B$  führt. Tabelle 5.1 zeigt die berechneten LCPs zwischen drei der insgesamt 383 SADS-Ortschaften.

Tabelle 5.1: Asymmetrische LCP Reisezeiten (in Sekunden)

	Klosters	St. Antönien	Küblis
Klosters	-	12588	6448
St. Antönien	12055	-	5890
Küblis	7854	7720	-

Es zeigt sich, dass die Reisezeit von Ort  $A$  nach  $B$  nicht dieselbe sein muss wie die Reisezeit von  $B$  nach  $A$ . Dies ist deshalb der Fall, weil der Dijkstra-Algorithmus für den Hin- und Rückweg nicht zwingend denselben Pfad berechnet. Der Grund dafür liegt in der asymmetrischen Hiking Funktion von Tobler, mit der die Kostenoberfläche berechnet wurde (vgl. Abb. 5.4): Bei leichter Hangneigung ist die Wandergeschwindigkeit höher als geradeaus. Abbildung 5.5 zeigt beispielhaft die Unterschiede des Hin- und Rückwegs zwischen Klosters und St. Antönien. Der Hin- und Rückweg (rote resp. blaue Linie) unterscheiden sich leicht wegen der asymmetrischen Hiking-Funktion von Tobler.

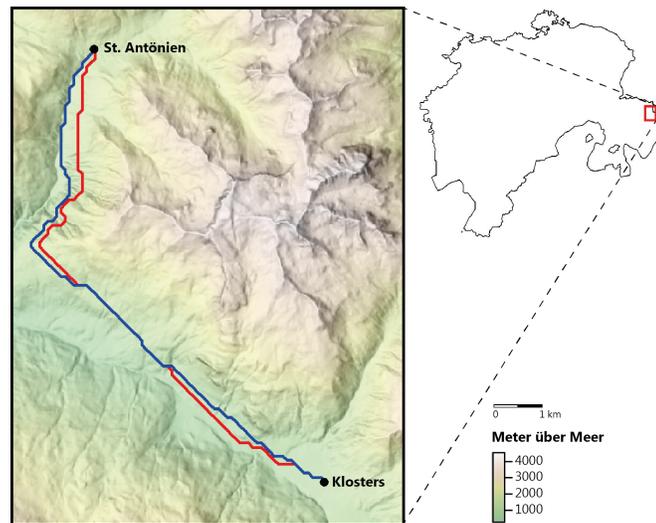


Abbildung 5.5: Beispiel zweier LCP-Reisewege zwischen St. Antönien und Klosters. Zwecks besserer Optik wurde für die Darstellung das DHM mit der Auflösung  $25 \times 25m$  verwendet.

Da die syntaktischen Unterschiede zwischen den Ortschaften symmetrisch sind (vgl. Abschnitt 4.2.1), wird vereinfacht angenommen, dass der Kontakt der Ortschaften untereinander ebenso symmetrisch erfolgt. Aus diesem Grund werden aus den asymmetrischen die durchschnittlichen Reisezeiten zwischen den Ortspaaen berechnet. Tabelle 5.2 zeigt denselben Tabellenausschnitt wie Tabelle 5.1 mit den durchschnittlichen LCP Reisezeiten.

Tabelle 5.2: Symmetrische LCP Reisezeiten (in Sekunden)

	Klosters	St. Antönien	Küblis
Klosters	-	12322	7151
St. Antönien	12322	-	6805
Küblis	7151	6805	-

### 5.1.1.3 Euklidische Reisezeiten (ED)

Die euklidischen Distanzen zwischen Ortschaften lassen sich mit ihren Koordinaten<sup>3</sup> und dem Satz des Pythagoras berechnen. Eine Korrektur der Erdkrümmung ist nicht notwendig, da diese innerhalb der Deutschschweiz einen zu vernachlässigenden Effekt auf die Distanzen hat. Damit die euklidischen Distanzen mit den LCP Reisezeiten verglichen werden können, werden sie anschliessend durch  $1,39 \text{ m/s}$  dividiert. Die  $1,39 \text{ m/s}$  entsprechen der Gehgeschwindigkeit bei flachem Terrain (vgl. Toblers Hiking Funktion in Abb. 5.4). Damit ergeben sich aus den euklidischen Distanzen euklidische Reisezeiten, die mit den LCPs verglichen werden können (vgl. Tabelle 5.3). Da der Begriff der euklidischen Reisezeiten nicht gebräuchlich ist, wird nachfolgend

<sup>3</sup> Die Ortschaften werden wegen der grossen Distanzen vereinfacht als Punkt und nicht als Fläche dargestellt.

die Abkürzung *ED* (für *euklidische Distanzen*) verwendet.

Tabelle 5.3: Euklidische Reisezeiten (in Sekunden)

	Klosters	St. Antönien	Küblis
Klosters	0	8267	6329
St. Antönien	8267	0	4800
Küblis	6329	4800	0

#### 5.1.1.4 Topographie-Einflussfaktor auf die Reisezeit

Für eine erste Visualisierung des Topographie-Einflusses auf die syntaktischen Distanzen wird ein sogenannter *Topographie-Einflussfaktor* berechnet (vgl. Kapitel 6). Dazu werden die durchschnittlichen LCP-Reisezeiten durch die euklidischen Reisezeiten der Ortschaften dividiert, wodurch man für jedes Ortspaar einen Einflussfaktor der Topographie auf die Reisezeit erhält. Der Vorteil eines solchen Quotienten ist, dass man den Einfluss der Topographie zwischen Ortschaften kondensiert in jeweils einer Zahl erhält. Ein Faktor kleiner als 1 würde bedeuten, dass die Reisezeit durch den Topographie-Einfluss kürzer wird (bei leichtem Gefälle des Geländes, vgl. Abb.5.4). Ein Faktor grösser als 1 bedeutet, dass die Reisezeit wegen der Topographie länger wird. Ein Faktor = 1 bedeutet, dass die Topographie keinen Einfluss auf die Reisezeit zwischen zwei Orten hat. Tabelle 5.4 zeigt wieder denselben Ausschnitt wie in den Tabellen 5.1 und 5.3. Zwischen den Ortschaften Klosters, St. Antönien und Küblis hat die Topographie in diesem Beispiel einen verlängernden Einfluss auf die Wanderzeiten. Die Wanderzeit zwischen Küblis und St. Antönien ist demnach stärker von der Topographie beeinflusst als zwischen Küblis und Klosters.

Tabelle 5.4: Durchschnittliche Einflussfaktoren der Topographie auf die Reisezeiten

	Klosters	St. Antönien	Küblis
Klosters	-	1.49	1.13
St. Antönien	1.49	-	1.42
Küblis	1.13	1.42	-

Um den Effekt der Topographie auf die Reisezeiten zu veranschaulichen, werden die beiden berechneten ED- und LCP-Distanzen der Ortspaare (vgl. Tabelle 5.2 resp. 5.3) mittels multidimensionaler Skalierung (MDS<sup>4</sup>) zurück in Koordinaten transformiert. Diese Koordinaten können danach kartographisch visualisiert und somit die Topographie-Einflüsse auf die Reisezeiten veranschaulicht werden (vgl. Abb. 5.6). Die linke Karte zeigt die originale Lage der SADS-Ortschaften und die Grenze der Deutschschweiz an. Die mittlere Karte stellt mit Pfeilen die Verschiebung sowie die relative Lage der SADS-Orte nach der MDS-Transformation dar.

<sup>4</sup>für eine detaillierte Beschreibung der MDS siehe die Publikation von John C. Gower (1966)

Vereinfacht gesagt zeigt die mittlere Karte die Lage der Ortschaften an, wenn Berge und Hügel der Deutschschweiz flach gewalzt würden. Durch den Einfluss der Topographie rückt das Mittelland im Verhältnis zu den Alpengebieten reisezeitmässig zusammen und die Alpen-Orte im Gegensatz dazu auseinander. Die rechte Karte zeigt nochmals die transformierte Lage der Orte und die Grenze der Deutschschweiz mit einem darüber gelegten Ellipsoid-Raster. Dieses veranschaulicht noch deutlicher die Stauchung im Flachland resp. das Aufblähen der Deutschschweiz in den Alpenregionen.

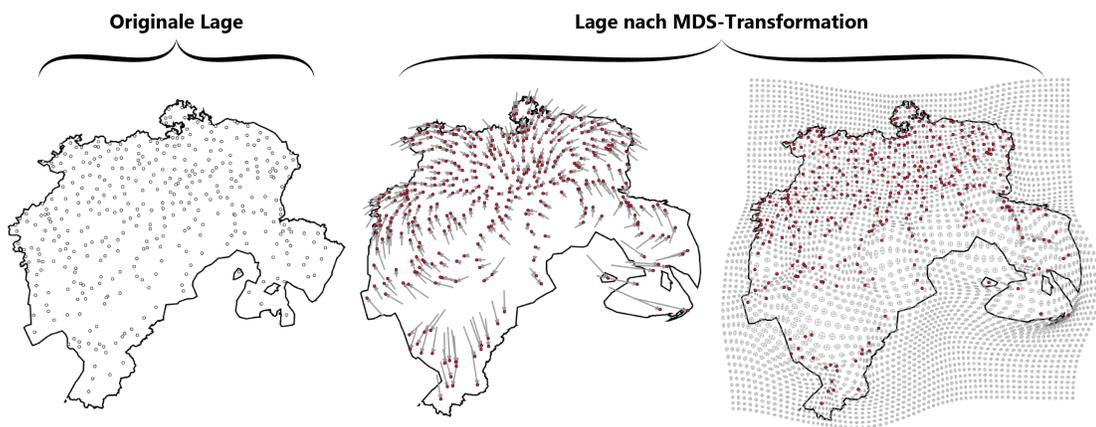


Abbildung 5.6: Konkrete und relative Lage der SADS-Orte auf Grundlage einer MDS-Transformierung der beiden Erreichbarkeitsmasse ED und LCP. Die Karten zeigen so den Einfluss der Topographie auf die relativen Distanzen der Ortschaften.

### 5.1.2 Die Einflussbereichsmasse

Wie im Abschnitt 2.1 beschrieben, verändert sich Sprache durch den Kontakt zu anderen Menschen. Neben der Erreichbarkeit hat gemäss dem Soziolinguisten Peter Trudgill (1974) auch die Bevölkerungszahl der Ortschaften einen Einfluss auf die gegenseitige Veränderung resp. Angleichung ihrer dort üblicherweise gesprochenen Dialekte. Nachfolgend werden die Berechnungen zweier auf Trudgills Theorie aufbauender Einflussbereichsmasse im Detail beschrieben.

Der Einflussbereich von Ortschaften wird mittels Gravitations-Index (GI) von Peter Trudgill (1974) untersucht. Dieser besagt, dass der gegenseitigen Einfluss von Ortschaften von deren Einwohnerzahlen und der Distanz zwischen ihnen abhängig ist. Je höher die Einwohnerzahl zumindest eines Ortes und je geringer die Distanz zweier Ortschaften, desto stärker soll der gegenseitige Einfluss auf die dort gesprochenen Dialekte sein. Dies äussert sich wiederum in einer geringeren sprachlichen Distanz. Trudgill erstellte einen asymmetrischen Gravitations-Index, bei dem bei einem paarweisen Ortsvergleich der Einfluss des Ortes mit der höheren Einwohnerzahl grösser ist als des Ortes mit der kleineren Einwohnerzahl (vgl. Formel 5.3):

$$GI_{A \rightarrow B} = \frac{Population_A \times Population_B}{Distanz_{A \leftrightarrow B}} \times \frac{Population_A}{Population_A + Population_B} \quad (5.3)$$

Der  $GI_{A \rightarrow B}$  ist der Wert für den Einfluss des Ortes  $A$  auf Ort  $B$ .  $Population_A$  ist die Bevölkerungszahl von Ort  $A$ ,  $Population_B$  ist die Bevölkerungszahl von Ort  $B$ . Die  $Distanz_{A \leftrightarrow B}$  ist die Distanz zwischen den Orten  $A$  und  $B$ . Der rechte Teil der Multiplikation ist dabei der Richtungsfaktor der Gleichung, welcher bestimmt, wie gross der Einfluss des Ortes  $A$  auf  $B$  ist, abhängig vom Grössenverhältnis der Einwohnerzahlen beider Ortschaften. Da für die späteren Korrelationsanalysen die GIs mit den symmetrischen syntaktischen Distanzen der Ortschaften verglichen werden, fällt der Richtungsfaktor weg. Die Folge ist ein symmetrischer  $GI_{A \leftrightarrow B}$ :

$$GI_{A \leftrightarrow B} = \frac{Population_A \times Population_B}{Distanz_{A \leftrightarrow B}} \quad (5.4)$$

Für die nachfolgenden Analysen werden für alle Ortspaare jeweils zwei unterschiedliche Gravitations-Indizes berechnet. Diese unterscheiden sich durch das verwendete Distanzmass voneinander (vgl. Formel 5.5 und 5.6):

$$GI-ED_{A \leftrightarrow B} = \frac{Population_A \times Population_B}{euklid. \text{ Reisezeit}_{A \leftrightarrow B}} \quad (5.5)$$

$$GI-LCP_{A \leftrightarrow B} = \frac{Population_A \times Population_B}{LCP \text{ Reisezeit}_{A \leftrightarrow B}} \quad (5.6)$$

Für die GI-EDs werden als Distanzmasse die euklidischen Reisezeiten zwischen den Ortschaften verwendet, für die GI-LCPs hingegen die LCP-Reisezeiten.

## 5.2 Scharfe Grenzen

Politik, Religion und Wirtschaft stellen die strukturellen Grundpfeiler eines Landes dar. In einem sich abwechselnden, historischen Zusammenspiel ihrer Machtverhältnisse beeinflussen, trennen und vereinen sie die in einem Land lebende Gesellschaft (vgl. Kley, 2016; Schnyder, 2013; Schuler et al., 2005). Die politischen, religiösen und wirtschaftlichen Grenzen beeinflussen deshalb auch die Kontakthäufigkeit der Menschen, was sich wiederum auf die gesprochenen Sprachen und ihre Dialekte abfärbt. Aus diesen Gründen wird je eine scharfe Grenze aus Politik, Religion und Wirtschaft auf ihren Einfluss auf die sprachlichen Unterschiede zwischen Ortschaften hin untersucht:

- Mit den **Kantonsgrenzen** werden als erstes politische Grenzen betrachtet (Unterabschnitt 5.2.1). Durch den föderalistischen Aufbau der Eidgenossenschaft haben die Kantone von der Vergangenheit bis hin zur Gegenwart einen grossen Einfluss auf das Zusammenleben in der Schweiz (vgl. Kley, 2016).
- Als zweites wird der Einfluss von **Konfessionsgrenzen** auf die schweizerdeutsche Syntax untersucht (Unterabschnitt 5.2.2). Die Reformation der katholischen Kirche hat in der Schweiz zu einer religiösen Trennung und bewaffneten Konflikten geführt. Obwohl die konfessionellen Unterschiede in der heutigen Zeit kaum mehr spürbar sind, hatten sie in der Vergangenheit den wohl stärksten Einfluss auf die Menschen im damals noch losen Staatenbund der frühen Eidgenossenschaft (Schnyder, 2013).
- Drittens und letztens wird mit den neuzeitlichen **Arbeitsmarktregionen** der Schweiz der Einfluss von wirtschaftlichen Grenzen und Pendlern auf die deutschschweizer Syntax untersucht (Unterabschnitt 5.2.3). Obwohl seit dem späten 19. Jh. die steigenden Pendlerzahlen eine immer grössere Bedeutung für den interregionalen sprachlichen Austausch haben (Auer et al., 2011), ist ihr Einfluss auf die Dialekte noch immer nicht geklärt. (Auer, 2013).

In den nachfolgenden Unterabschnitten 5.2.1 bis 5.2.3 werden die verwendeten scharfen Grenzen und ihre jeweilige Bedeutung für die Schweiz kurz vorgestellt.

### 5.2.1 Kantonsgrenzen

Die Kantone und ihre Grenzen haben eine lange Geschichte in der föderalistischen Schweiz. Die heutige Schweiz hat ihren Ursprung im 14. Jh. als Geflecht von Bündnissen innerhalb des heiligen römischen Reichs. Die damalige Eidgenossenschaft war noch keine Einheit, sondern bestand aus lockeren Bündnissen zwischen Städten, Kantone sowie deren Untertanengebieten. Erst mit dem Einmarsch Napoléons in die Schweiz und der Gründung der helvetischen Republik zwischen 1798 und 1803 wurden die Untertanengebiete der alten Eidgenossenschaft befreit und der lose Staatenbund zu einem Zentralstaat nach französischem Vorbild umfunktioniert (Würgler, 2012). Nach der endgültigen Niederlage Napoléons bei Waterloo wurden am Wiener Kongress

sowie während den Verhandlungen des 2. Pariser Friedens zwischen 1814 und 1815 von den Siegermächten Österreich, Preussen, Russland und Grossbritannien die europäischen Grenzen neu gezogen. Dabei war die Schweiz als ehemaliger französischer Vasallenstaat Verhandlungsobjekt des Wiener Kongresses. Die Vertreter der Grossmächte strebten nach den Erfahrungen der vergangenen Kriege die Schaffung eines neutralen Pufferstaats zwischen den Grossmächten Frankreich und Österreich an. Der Schweiz wurde dabei von den Grossmächten immerwährende Neutralität garantiert und die bis heute unveränderte Landesgrenze sowie die Kantons Grenzen wurden festgelegt<sup>5</sup> (Jorio, 2015). Abbildung 5.7 zeigt die heutigen Kantons Grenzen der Schweiz.

Für alle verwendeten 383 SADS-Ortschaften sind die Kantonszugehörigkeiten im verwendeten Datensatz vorhanden, wobei allfällige Enklavengebiete oder frühere historische Kantonswechsel der Ortschaften nicht berücksichtigt werden.

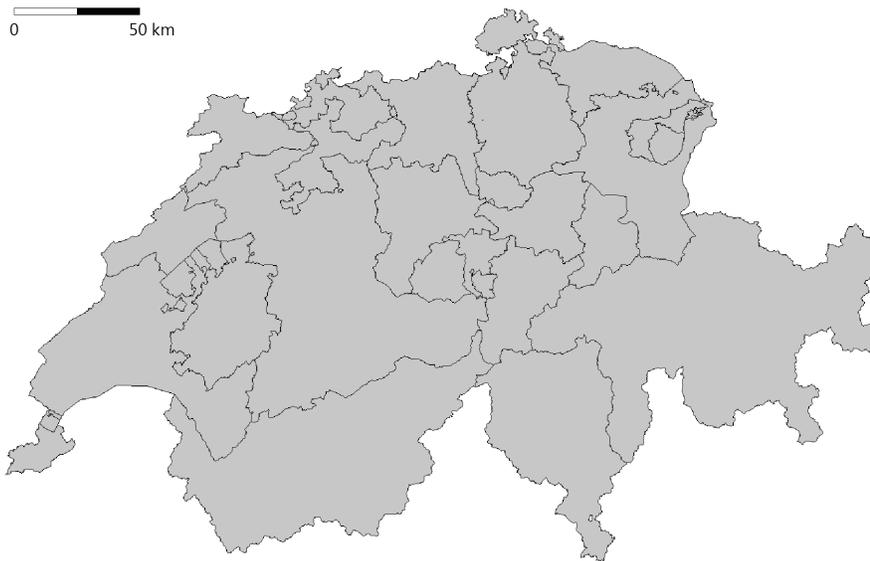


Abbildung 5.7: Die Kantons Grenzen der Schweiz. Datenquelle: Universität Zürich (2017b)

---

<sup>5</sup>mit Ausnahme des 1979 gebildeten Kanton Jura.

### 5.2.2 Konfessionsgebietsgrenzen

Die Konfessionsgrenzen zwischen katholischen und protestantischen Gebieten haben ihre Ursprünge im 16. Jh. in der Spaltung der katholischen Kirche. Die Reformation hatte nicht nur in der Eidgenossenschaft, sondern im gesamten Europa weitreichende politische, kulturelle und gesellschaftliche Auswirkungen. Zu Beginn der Reformation hatte die Kirchenspaltung noch wenig spürbaren Einfluss auf das Zusammenleben in der Schweiz. Die Bevölkerungsstrukturen wandelten sich durch protestantische Glaubensflüchtlinge hauptsächlich in Genf und in geringerem Ausmass auch in Zürich und Basel. Bis Mitte des 16. Jh. blieben die konfessionellen Grenzen aber fließend und die Anhänger der beiden Konfessionen teilten sich vielerorts auch eine Kirche. Mitte 16. Jh. wurden die Unterschiede der beiden Konfessionen exakter ausformuliert, was zu vermehrten konfessionellen Konflikten führte und 1597 im Kanton Appenzell beispielsweise auf eine Kantonsteilung hinauslief (Schnyder, 2013). Die Spannungen entluden sich schliesslich in mehreren bewaffneten Konflikten zwischen protestantischen und katholischen Ortschaften wie bspw. den beiden Villmergerkriegen 1656/1712 und endeten mit dem Sonderbundskrieg von 1847, welcher gleichzeitig auch der letzte militärische Konflikt innerhalb der Schweiz war (Lau, 2013, 2014; Schnyder, 2013). Gemäss Schnyder (2013) zählt die Reformation zu den prägendsten Ereignissen der Schweizer Geschichte und bedeutete für die Eidgenossenschaft eine Zerreißprobe.

Die verwendeten Konfessionsgrenzen stammen von der Konfessionskarte des *Sprachatlas' der Deutschen Schweiz* (Hotzenköcherle, 1962) und ursprünglich aus dem *Atlas der schweizerischen Volkskunde*<sup>6</sup>. Diese wurde eingescannt und mithilfe eines *LANDSAT*-Luftbildes der Schweiz (NASA, 1994) mit der Software *ArcGIS* (ESRI, 2017) georeferenziert. Da die Karte von Hand gezeichnet wurde, weist diese leichte Ungenauigkeiten auf. Zusammen mit den Mischgebieten hat dies zur Folge, dass nicht alle der 383 SADS-Erhebungsorte eindeutig einer der zwei Konfessionen (katholisch oder protestantisch) zugewiesen werden können. In der weiteren Analyse werden deshalb nur die 325 Orte berücksichtigt, deren Konfessionszugehörigkeit auf Grundlage der Konfessionskarte eindeutig ist. Von den verbleibenden 325 Orten sind 139 katholisch und 186 protestantisch. Abbildung 5.8 zeigt die mit dem *LANDSAT*-Luftbild verwendete georeferenzierte Konfessionskarte der Schweiz und die den Konfessionsgebieten zugewiesenen SADS-Ortschaften. Mit einem schwarzen Sternsymbol sind die Ortschaften gekennzeichnet, welche entweder in einem Mischgebiet oder direkt auf einer der Grenzen liegen und deshalb von der Analyse ausgeschlossen werden. Die unterschiedlichen Farben der Symbole verweisen auf die je sechs *zusammenhängenden* katholischen resp. protestantischen Gebiete. Die Zuteilung in separate, zusammenhängende Gebiete der jeweiligen Konfession wurde gemacht, damit bei den späteren Analysen mit unterschiedlichen Nb-Skalen (vgl. Abschnitt 5.3.1.1 für eine Erklärung der Nb-Skalen) nicht Ortspaare gleicher Konfession als zusammengehörig betrachtet werden, wenn zwischen diesen ein anderes Konfessionsgebiet liegt. Dieser Fall tritt zwar nicht bei den direkten Voronoi-Nachbarorten auf, jedoch bei den zusätzlichen untersuchten Nb-Skalen.

---

<sup>6</sup>Eine exakte Quellenangabe kann nicht gemacht werden, da der Autor des *Sprachatlas der Deutschen Schweiz* auf einen genauen Literaturverweis verzichtete.

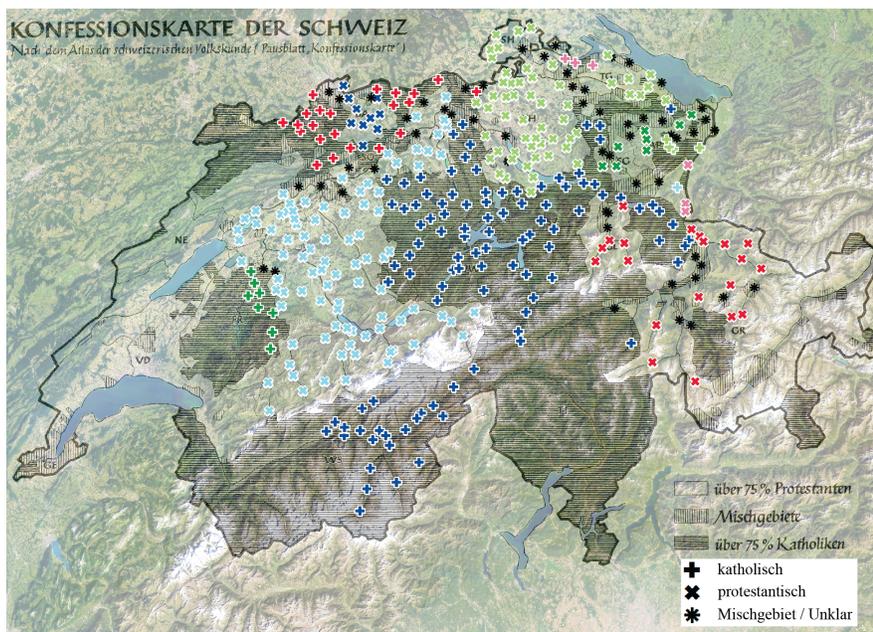


Abbildung 5.8: Alte Konfessionskarte der Schweiz mit allen 383 SADS-Erhebungsorten, welche mit einem LANDSAT-Satellitenbild georeferenziert wurde. Die Ortschaftssymbole einer Konfession mit der gleichen Farbe befinden sich im selben Konfessionsgebiet.

### 5.2.3 Arbeitsmarktregionsgrenzen

Der Einfluss von Pendler auf die Dialektlandschaft ist eines der grossen vernachlässigten Felder innerhalb der Geolinguistik. Ihre Auswirkungen auf städtische und ländliche Dialekte sind noch immer nicht verstanden (Auer, 2013). Die Gründe für die Vernachlässigung des Einflusses von Pendlern liegen zum einen im erst kürzlich vollzogenen Fokuswechsel der Dialektologen vom traditionellen Studienobjekt des seit seiner Geburt auf dem Land lebenden, alten Mannes hin zu einer heterogenen, mobileren städtischen Studiengruppe, welche alters-, geschlechts- und herkunftsgemischt ist und unterschiedliche Bildungshintergründe hat (Britain, 2012) (vgl. auch Szmrecsanyi, 2011). Zum anderen explodierte der Einfluss der Mobilität und der daraus folgende interregionale Kontakt von Ortschaften laut Britain (2012) auch erst mit dem Aufkommen von Pendlern des hoch mobilen tertiären Wirtschaftssektors und deren Bereitschaft, stets grössere Distanzen für eine Arbeitsstelle in Kauf zu nehmen<sup>7</sup>. Laut Auer et al. (2011) führen Pendlerbewegungen zu einem vermehrt interregionalen Dialektaustausch. Gemäss dem Bundesamt für Statistik (BFS) haben diese auch innerhalb der Schweiz in den letzten zwei Jahrzehnten deutlich zugenommen. Im Jahr 2015 waren neun von zehn Erwerbstätige in der Schweiz Pendler<sup>8</sup>, wobei 51% davon ihre Wohngemeinde und weitere 20% gar ihren Wohnkanton verliessen (Bundesamt

<sup>7</sup> vgl. dazu bspw. die Arbeiten von Champion, Coombes und Brown (2009), Green, Hogarth und Shackleton (1999) und Rouwendal und Rietveld (1994) über die Zunahme von Pendlerbewegungen.

<sup>8</sup> nach der Definition des Bundesamts für Statistik (BFS) sind Pendler Personen, die zum Aufsuchen ihres Arbeitsplatzes ihr Wohngebäude verlassen (Bundesamt für Statistik, 2016)

für Statistik, 2016).

Die 16 in dieser Arbeit verwendeten Arbeitsmarktregionen sind Aggregationen der kleinräumigen MS-Regionen (= «mobilité spatiale»), welche 1982 im Rahmen eines Forschungsprojekts des Nationalen Forschungsprogramms (NFP) entstanden sind (Schuler et al., 2005) und stützen sich auf Pendlereinzugsgebiete von 1960 ab. Als solches repräsentieren sie wirtschaftlich-raumplanerische Grenzen und sind ein Abbild der Schweizer Pendlerbewegungen. Die 383 SADS-Orte und die Arbeitsmarktregionen wurden mit der Software ArcGIS (ESRI, 2017) räumlich verschnitten und die Ortschaften somit den einzelnen Regionen zugewiesen.

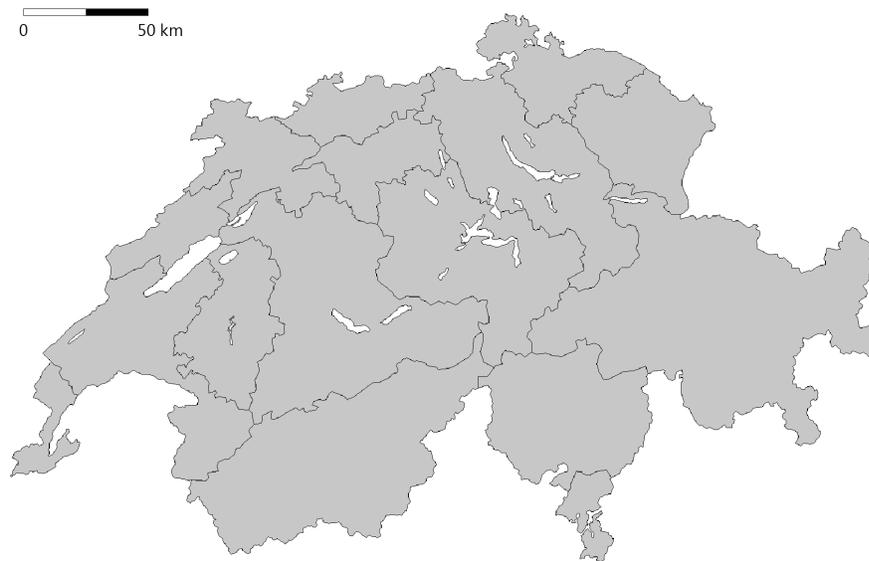


Abbildung 5.9: Die 16 Arbeitsmarktregionen der Schweiz.

## 5.3 Verwendete Analysemethoden

In den nachfolgenden Abschnitten 5.3.1 bis 5.3.4 werden die Analysemethoden vorgestellt, mit denen die Einflüsse unterschiedlicher Grenzen auf syntaktische Unterschiede der SADS-Erhebungsorte untersucht wurden. Abschnitt 5.3.1 beschreibt das *Modifiable Areal Unit Problem* (MAUP) und zeigt auf, wie in dieser Arbeit damit umgegangen wird (Abschnitt 5.3.1.1). Die Abschnitte 5.3.2 und 5.3.3 erklären die Methoden, die verwendet wurden, um den Einfluss unscharfer resp. scharfer Grenzen auf die syntaktischen Distanzen zu erforschen. Abschnitt 5.3.4 zeigt schliesslich eine Möglichkeit auf, wie man die (un-)scharfen Grenzen zusammenzufassen und zur Modellierung der syntaktischen Distanzen zu verwenden kann.

### 5.3.1 Räumlicher Effekt und Skalenproblem (MAUP)

Unsere Welt ist eingeteilt in von Menschen gezogene Grenzlinien. Die Gründe, wie und wo diese Grenzen verlaufen, sind vielschichtig. Unabhängig von den historischen oder topographischen Hintergründen der Grenzziehungen sind sie aber die Ursache für ein analytisches Problem, das in der räumlichen Datenanalyse als *Modifiable Areal Unit Problem* (MAUP) bekannt ist (vgl. Openshaw und Taylor, 1979). Das MAUP besteht aus zwei Teilen: dem Skalen-Problem und dem Zonen-Problem. Das Skalen-Problem definiert sich dadurch, dass die berechneten und dargestellten Resultate abhängig sind von der gewählten räumlichen Auflösung. Dies zeigt sich oftmals bei der Aggregation von Daten vieler kleiner Gebiete zu wenigen grossen Gebieten. Das MAUP ist nicht nur ein statistisches, sondern auch ein kartographisches Problem. Abbildung 5.10 zeigt die Visualisierungen der amerikanischen Präsidentschaftswahlen von 2012. Gemäss der auf Bezirksebene aggregierten Karte auf der linken Seite von Abbildung 5.10 wählte die Mehrheit der Amerikaner republikanisch. Das Problem einer solchen Visualisierung ist jedoch, dass sie nicht das Verhältnis der Einwohner zur Bezirksfläche zeigt. Eine mögliche kartographische Lösung für dieses Problem sind sogenannte anamorphe Karten, die das Verhältnis der Bevölkerung zur Bezirksfläche miteinbeziehen (siehe Abb. 5.11). Der zweite Teil des MAUP, das Zonen-Problem, beschreibt das Problem der sich ändernden Resultate wenn Grenzlinien unterschiedlich gezogen werden (Wong, 2009). Dass das Zonen-Problem relevante Auswirkungen im Alltag hat, zeigt beispielsweise die übliche Praxis der *Wahlkreisverschiebungen*<sup>9</sup> in den USA, bei dem die Grenzen der Wahlkreise so gezogen werden, dass sie dem eigenen Wahlsieg bei der nächsten Wahl dienen (Ingraham, 2017; Johnston, 2002; Massey, Rothwell und Domina, 2009). Abbildung 5.12 zeigt beispielhaft, was für Auswirkungen unterschiedlich gezogenen Wahlkreise auf eine Wahl haben können, wobei je nach Grenzziehung der Wahlkreise bei einer Majorzwahl entweder *Blau* die Mehrheit der Wahlkreise gewinnt oder *Rot*. Obwohl in der Vergangenheit einige Vorschläge gemacht wurden um das MAUP zu lösen, konnte sich keine davon als Universalösung durchsetzen<sup>10</sup>. Gemäss Wong (2009) ist es deshalb bei der räumlichen Datenanalyse

<sup>9</sup>engl. Gerrymandering

<sup>10</sup>z.B. Waldo Toblers Migrationsmodell (Tobler, 1989), die geographisch-gewichtete Regressionsanalyse von Fotheringham, Brunson und Charlton (2002) oder die Miteinbeziehung geostatistischer Variogramme, wie sie

am wichtigsten, das MAUP als solches zu erkennen und zu beschreiben.

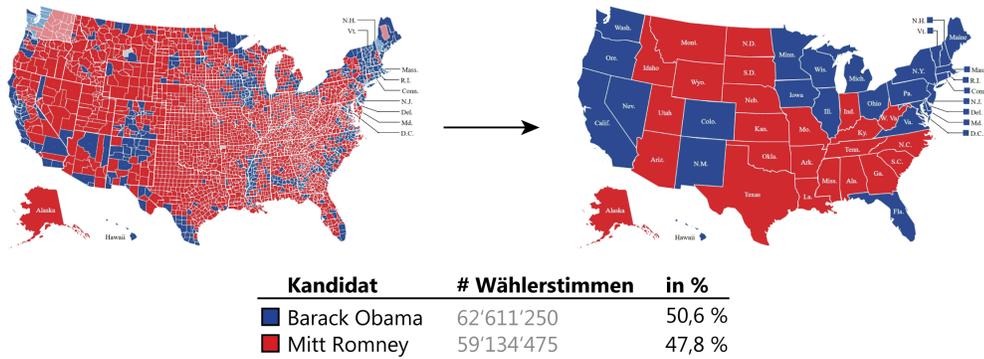


Abbildung 5.10: Das MAUP Skalen-Problem am Beispiel der amerikanischen Präsidentschaftswahl 2012. Je nach verwendetem Massstab für die Aggregation ergibt sich ein anderes Bild des Wahlergebnisses. Bildquelle: modifiziert nach Andrews et al. (2012).

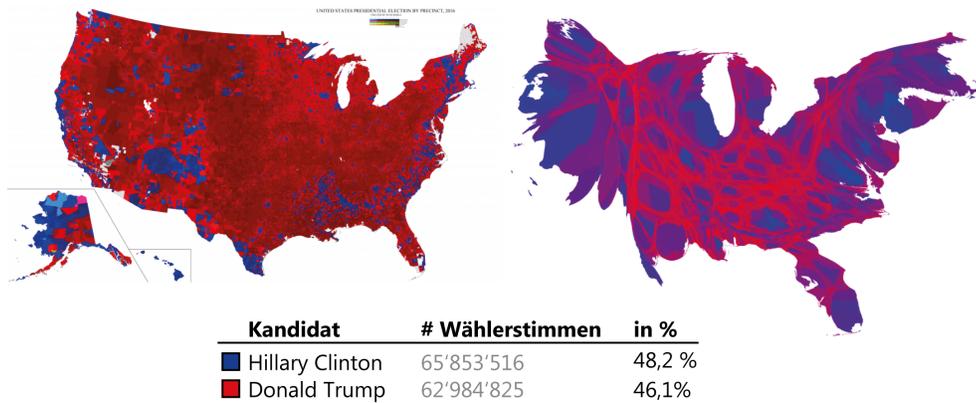


Abbildung 5.11: Das MAUP Skalen-Problem am Beispiel der amerikanischen Präsidentschaftswahl 2016. Obwohl die Karte auf der linken Seite dem Betrachter suggeriert, dass die Mehrheit der Amerikaner den republikanischen Kandidaten gewählt hat, zeigt sich bei Miteinbezug des Verhältnisses Einwohner/Bezirksfläche mit einer anamorphen Darstellung (rechte Seite) ein anderes Bild. Bildquelle: modifiziert nach Newman (2016).

Atkinson und Tate (2000) vorschlagen

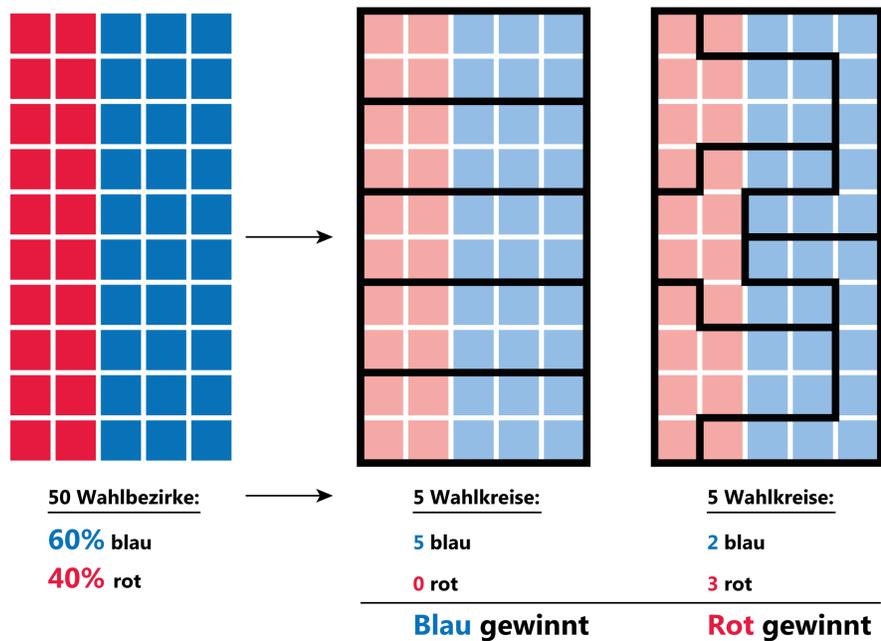


Abbildung 5.12: Das MAUP Zonen-Problem am Beispiel der Wahlkreisverschiebung, welches den Ausgang von Majorzwahlen entscheidend beeinflussen kann. Bildquelle: modifiziert nach Nass (2015).

### 5.3.1.1 Unterschiedliche Nachbarschaftsgrößen

In der Linguistik gibt es verschiedene Modelle und Theorien, welche die Veränderung von Dialekten zu erklären versuchen (vgl. Auer et al., 2011). Eines davon ist das *Wellenmodell*, welches davon ausgeht, dass Dialektveränderungen wellenförmig von einem Ursprungszentrum stattfinden. Die Veränderungen starten von einem Punkt und breiten sich über ihre Nachbarorte aus. Gemäss dieser Theorie ist die wichtigste Form des Kontakts diejenige der direkten Nachbarschaft. Gemäss Auer et al. (2011) hat sich die Kontaktrealität der Menschen seit dem späten 19. Jh. aber verändert. Die grössere Mobilität sowie moderne Kommunikationsmittel führen dazu, dass Menschen nicht nur Kontakt zu Dialekten aus ihren Nachbardörfern, sondern auch mit Dialekten aus anderen Landesteilen, der schriftlichen sowie mündlichen Standardsprache oder Fremdsprachen haben (vgl. Auer et al., 2011; Hernández-Campoy, 2003).

Um sowohl das Skalenproblem (MAUP) (vgl. Abschnitt 5.3.1) als auch die grössere Mobilität der Menschen zu berücksichtigen (vgl. Auer et al., 2011), sind in dieser Arbeit alle Untersuchungen mit variierenden Nachbarschaftsgrößen durchgeführt (nf. *Nb-Skalen*). Alle bei den Analysen verwendeten Werte sind für jede Ortschaft mit folgenden Nb-Skalen aggregiert: direkte Voronoi-Nachbarorte sowie die jeweils fünf, zehn, 15, 20, 25, 30 und 50 nächsten Ortschaften. Ein Voronoi-Polygon ist eine Fläche um einen Punkt, innerhalb der alle Stellen näher

an diesem Punkt liegen wie an jeglichen anderen Punkten ausserhalb der Fläche (vgl. Abb. 5.13)(GITTA, 2015). Die Grenzlinien eines Voronoi-Polygons eines Punktes werden dabei über die Mittelsenkrechten zu seinen Nachbarpunkten hin konstruiert. Abbildung 5.14 zeigt die verwendeten Voronoi-Polygone, mit denen die direkten, angrenzenden Nachbarschaften der Orte definiert wurden (nachfolgend auch Voronoi-Nachbarn genannt). Abbildung 5.15 stellt die aus den Voronoi-Polygone berechneten angrenzenden Voronoi-Nachbarschaften dar. Es gibt einige Sonderfälle wie bspw. Voronoi-Nachbarschaftsbeziehungen in Schaffhausen oder zwischen Oberraxen und Tamins, welche über die Landesgrenze resp. über die Grenze der Deutschschweiz gehen.

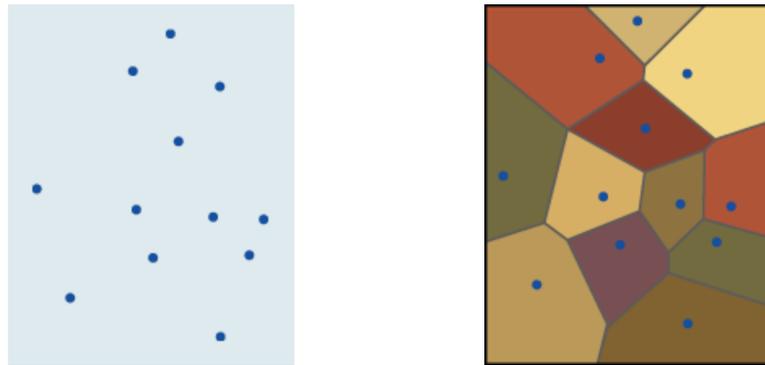


Abbildung 5.13: Datenpunkte (links) und ihre Voronoi-Polygone (rechts). Bildquelle: GITTA (2015)

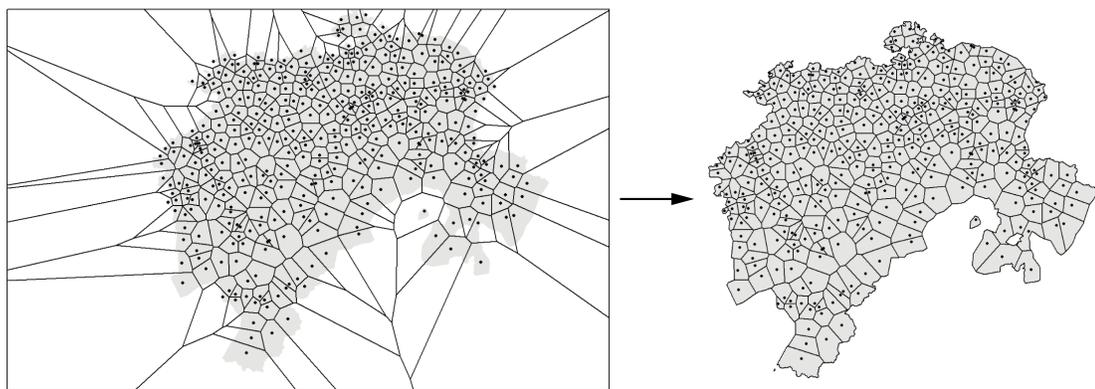


Abbildung 5.14: Die ursprünglich berechneten Voronoi-Polygone der SADS-Orte (links) und nach der Verschneidung mit der Deutschschweizer Grenze (rechts), wie sie für die weiteren Analysen verwendet werden.



Abbildung 5.15: Mit Verbindungslinien visualisierte Voronoi-Nachbarschaften zwischen den SADS-Orten (schwarze Punkte). Die Voronoi-Nachbarschaft zwischen Obersaxen (roter Punkt) und Tamins (blauer Punkt) kommt deshalb zustande, weil deren gemeinsame Voronoi-grenze ein wenig näher bei Tamins liegt als die Grenze der Deutschschweiz.

Obwohl anzunehmen ist, dass Orte keinen oder nur wenig direkt über die Landesgrenzen führenden Kontakt haben, liegen die berechneten Voronoi-Nachbarschaften in der Natur der Voronoi-Polygone und wurden deshalb nicht manuell korrigiert. Zu beachten ist zusätzlich, dass die jeweiligen Voronoi-Nachbarn eines Ortes nicht zwingend die reellen nächsten Nachbarschaftsorte sein müssen, da die 383 SADS-Orte nur ca. 25% aller Orte der Deutschschweiz sind (vgl. Abschnitt 4.2). Es handelt sich somit nur um eine Näherung der direkten Nachbarschaften. Die restlichen untersuchten Nb-Skalen der fünf, zehn, 15, 20, 25, 30 und 50 nächste Ortschaften wurden für jede der 383 SADS-Orte anhand der euklidischen Distanz bestimmt. Abbildung 5.16 zeigt am Beispiel der Ortschaft Küblis alle acht untersuchten Nb-Skalen sowie deren Daten der Voronoi-Nb-, 5-Nb, 10-Nb und 15-Nb-Skala. Für die Korrelationsberechnung zwischen den syntaktischen und euklidischen Distanzen von Küblis und seiner 5er-Nachbarschaft bedeutet das, dass der Korrelationswert zwischen den fünf euklidischen und den fünf syntaktischen Distanzen von Küblis zu seinen fünf nächsten Nachbarn berechnet wird. Alle Analysen werden wie am Beispiel von Küblis für jeden der 383 Erhebungsorte durchgeführt und die Resultate für alle acht Nb-Skalen berechnet.

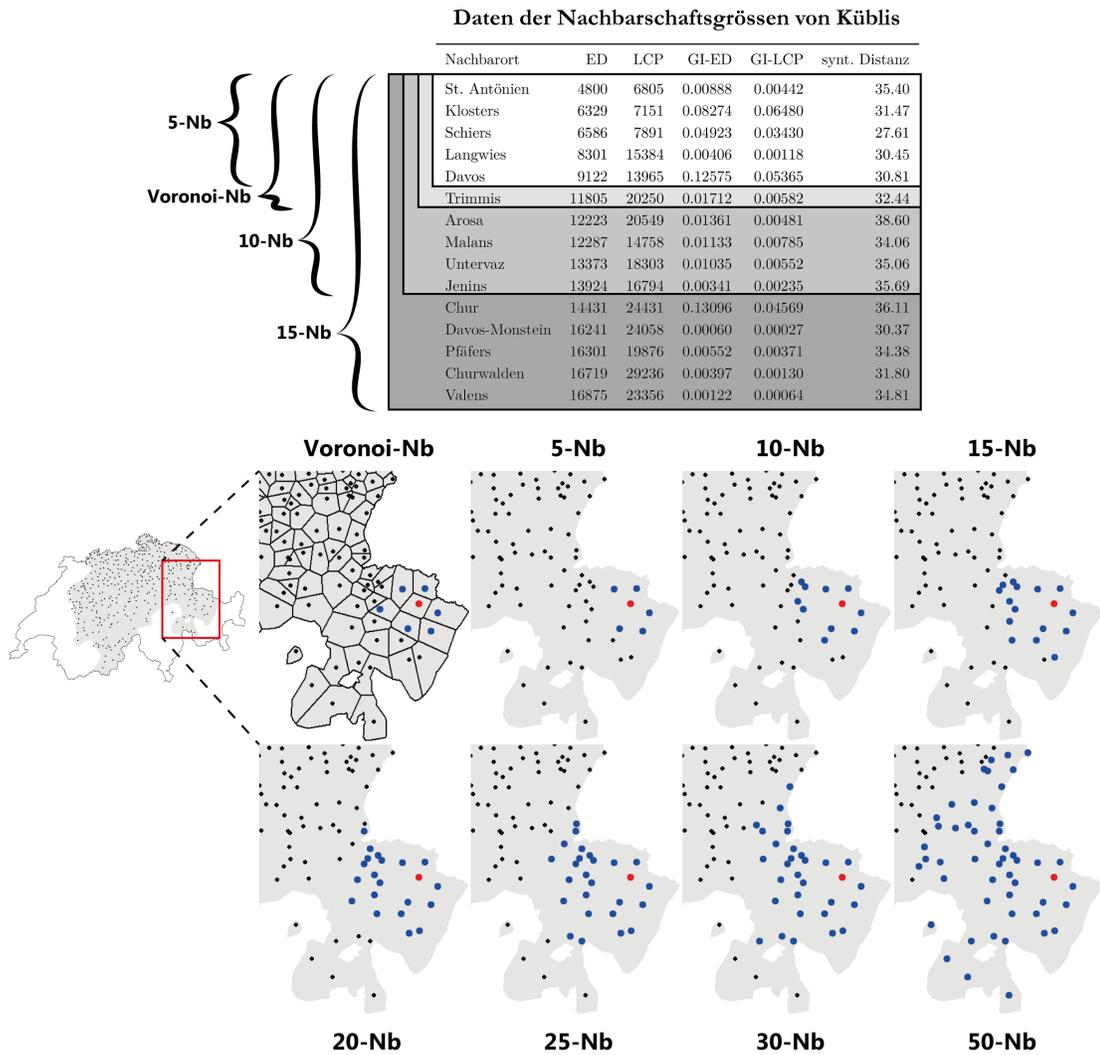


Abbildung 5.16: *Oberer Grafikteil:* Die Daten der Nachbarschaftsgrößen der Voronoi-Nb-, 5-Nb-, 10-Nb- und 15-Nb-Skala von Küblis aus betrachtet. *Unterer Grafikteil:* Die insgesamt acht Nb-Skalen am Beispiel der Gemeinde Küblis (roter Punkt). Die blauen Punkte sind die Nachbarorte von Küblis für die jeweilige Skala.

Eine Untersuchung der verschiedenen Nb-Skalen ist nicht nur wegen des MAUP wichtig, sondern auch, um den Effekt der geographischen Distanz auf die syntaktischen Distanzen besser untersuchen und verstehen zu können. Unterschiedliche Nb-Skalen ergeben eine grössere Varianz der räumlichen Distanzen zwischen den Ortschaften. Diese grössere Varianz erleichtert wiederum das Erkennen eines allfälligen räumlichen Effekts auf die syntaktischen Distanzen der Orte.

### 5.3.2 Korrelationsanalysen und Dominanzkarten

Mittels **Korrelationsanalysen** werden die in Abschnitt 5.1 beschriebenen Erreichbarkeits- resp. Einflussbereichsmasse auf einen allfälligen Zusammenhang (= Korrelation) mit den syntaktischen Distanzen der Ortschaften hin untersucht. Es wird erwartet, dass die beiden Erreichbarkeitsmasse ED und LCP positiv mit der syntaktischen Distanz korrelieren. D.h. je grösser die geographische Distanz, desto grösser sind die syntaktischen Unterschiede zwischen Ortschaften. Erwartet wird weiter, dass die LCPs stärker mit den syntaktischen Unterschieden korrelieren als die euklidischen Reisezeiten, da die LCP-Reisezeiten die Erreichbarkeit der Orte – und somit deren sprachlichen Austausch – besser abbilden als die ED-Reisezeiten. Für die Einflussbereichsmasse GI-ED und GI-LCP wird erwartet, dass diese negativ mit den syntaktischen Distanzen korrelieren. Je stärker also der Einfluss von zwei Ortschaften aufeinander, desto geringer ist ihre sprachliche Distanz. Weiter wird erwartet, dass die Korrelationen der syntaktischen Distanzen mit den beiden GIs stärker sind als mit den Erreichbarkeitsmassen ED und LCP. Dies deshalb, weil bei den GIs mit der Bevölkerungsanzahl ein weiterer Faktor miteinbezogen wird, der gemäss Trudgill (1974) die sprachliche Ähnlichkeit erklären kann. Ähnlich wie bei den Erreichbarkeitsmassen wird vermutet, dass der GI-LCP stärker korreliert als der GI-ED, weil im GI-LCP der Einfluss der Topographie miteinbezogen wird. Als Korrelationsmass wird *Kendalls*  $\tau$  (Kendall, 1938) verwendet. Kendalls  $\tau$  ist wie *Spearman's*  $\rho$  (Spearman, 1904) ein parameterfreies Rangkorrelationsmass, welches den Zusammenhang zweier Variablen beschreibt und keine Normalverteilung der Daten voraussetzt. Im Vergleich zu Spearman's  $\rho$  haben Ausreisser in den Daten einen geringeren Einfluss auf Kendalls  $\tau$ . Zusätzlich kann der Standardfehler berechnet werden (Howell, 2012). In dieser Arbeit wird mit Kendalls  $\tau$  untersucht, ob ein Zusammenhang zwischen den unscharfen Grenzen und den syntaktischen Distanzen der Ortschaften besteht. Um allfällige räumliche Muster sichtbar zu machen, werden für alle Distanz- und Einflussbereichsmasse zusätzlich **Dominanzkarten** erstellt. Dabei wird jede Ortschaft entsprechend desjenigen Erreichbarkeits- resp. Einflussbereichsmasses eingefärbt, welches für den betreffenden Ort am stärksten mit der syntaktischen Distanz korreliert. Das Resultat sind eingefärbte Voronoi-Polygon-Karten, welche bei allfällig erkennbaren räumlichen Mustern zusätzliche Erkenntnisse über den Einfluss der unscharfen Grenzen auf die syntaktischen Distanzen geben können.

### 5.3.3 Mittelwerttests

Bei den Analysen über den Einfluss der hier untersuchten scharfen Grenzen auf die syntaktischen Unterschiede der Ortschaften werden die drei Grenzarten separat untersucht. Dazu werden die Ortspaare für die Kantons-, Konfessionsgebiet- und Arbeitsmarktregionsgrenzen jeweils in zwei Gruppen eingeteilt. Die eine Gruppe beinhaltet diejenigen Ortspaare, zwischen denen eine scharfe Grenze liegt, die andere Gruppe beinhaltet alle Ortspaare, zwischen denen *keine* scharfe Grenze liegt. Da die Daten aller Gruppen nicht normalverteilt sind (vgl. Abb. 9.6 und 9.7 im Anhang), wird anschliessend mittels nicht-parametrischem *Mann-Whitney-U* Test überprüft, ob die zentralen Tendenzen der zwei Gruppen einen signifikanten Unterschied aufweisen.

Da die Methodik der Mittelwerttests und die aufgestellte  $H_0$ - resp. deren Alternativ-Hypothese  $H_1$  für alle untersuchten drei scharfen Grenzarten dieselbe ist, wird auf eine wiederholte Beschreibung verzichtet und stattdessen im nachfolgenden Abschnitt beispielhaft für alle drei scharfen Grenzen der Fall der Kantonsgrenzen beschrieben. Für alle acht Nb-Skalen werden die Tests einzeln durchgeführt.

#### 5.3.3.1 Getestete Hypothesen am Beispiel der Kantonsgrenzen

Die Arbeitshypothese lautet, dass die syntaktische Distanz von Ortspaaren, zwischen denen eine Kantonsgrenze liegt, grösser ist als die syntaktische Distanz von Ortspaaren, die im selben Kanton liegen. **Die Null-Hypothese** des Mann-Whitney-U Tests beschreibt, dass der Mittelwert der Gruppe *«im selben Kanton»* gleich ist wie als der Mittelwert der Gruppe *«nicht im selben Kanton»*.

$$H_0 : \mu_1 = \mu_2 \quad (5.7)$$

Dagegen lautet die **Alternativ-Hypothese**, dass der Mittelwert der Gruppe *«im selben Kanton»* kleiner ist als der Mittelwert der Gruppe *«nicht im selben Kanton»*:

$$H_1 : \mu_1 < \mu_2 \quad (5.8)$$

Getestet wird mit einem Signifikanzniveau  $\alpha = 0.01$  gegen die Null-Hypothese, womit die Wahrscheinlichkeit für einen Fehler 1. Art<sup>11</sup> nicht grösser als 1% ist. Um zu untersuchen, ob eine allfällige Signifikanz der Kantonsgrenzen auch bedeutend ist, wird zusätzlich noch Cohens  $d$  (Cohen, 1988) als Mass der Effektstärke berechnet. Dieses Effektstärkemass wird im nachfolgenden Abschnitt beschrieben.

#### 5.3.3.2 Das Mass der Effektstärke

Auch wenn statistische Tests signifikante Unterschiede nachweisen können, bedeutet dies nicht, dass die daraus gezogenen Schlüsse auch eine praktische Relevanz haben (vgl. z.B. Cohen, 1988; Durlak, 2009; Sullivan und Feinn, 2012). Statistische Signifikanz hängt unter anderem mit der

---

<sup>11</sup>Ein Fehler 1. Art liegt vor, wenn die  $H_0$ -Hypothese fälschlicherweise verworfen wird obwohl sie richtig wäre.

Stichprobengrösse zusammen. Bei kleinen Stichprobengrössen müssen deren Unterschiede gross sein, damit eine statistische Signifikanz nachgewiesen werden kann. Andersherum verhält es sich bei grossen Stichprobengrössen. Dort reichen bereits kleine Abweichungen der Unterschiede zwischen den Stichprobengruppen, um als statistisch relevant betrachtet zu werden. Aus diesem Grund ergibt es Sinn, dass man neben der statistischen Signifikanz auch deren praktische Relevanz überprüft. Es gibt verschiedene Möglichkeiten, die Stärke eines Effekts zu messen. Zwei bekannte Masse sind Cohens  $d$  (Cohen, 1988) und Pearsons  $r$ . Da sich die Werte von Pearsons  $r$  bei unterschiedlichen Gruppengrössen verzerren können, wird in dieser Arbeit Cohens  $d$  verwendet (Universität Zürich, 2016a). Cohens  $d$  zeigt auf, ob der Effekt der Grenzen auch bedeutend ist:

$$d = \frac{\mu_1 - \mu_2}{SD_{pooled}} \quad (5.9)$$

Dabei ist  $\mu_1$  der Mittelwert der Gruppe Ortspaare, die im selben scharfen Grenzgebiet (z.B. in denselben Kantonen) liegen und  $\mu_2$  der Mittelwert der Gruppe Ortspaare, die *nicht* im selben scharfen Grenzgebiet liegen.  $SD_{pooled}$  ist die gewichtete Standardabweichung der beiden Gruppen, bei der die individuellen Standardabweichungen gemittelt und anhand ihrer Gruppengrössen gewichtet werden (vgl. Volker, 2006):

$$SD_{pooled} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}} \quad (5.10)$$

$$SD_{1,2} = \frac{1}{n_{1,2} - 1} \sum_{i=1}^{n_{1,2}} (x_i - \mu_{1,2})^2 \quad (5.11)$$

$$\mu_{1,2} = \frac{1}{n_{1,2}} \sum_{i=1}^{n_{1,2}} x_i \quad (5.12)$$

Dabei ist  $n_1$  und  $SD_1$  die Anzahl Ortspaare resp. die Standardabweichung der Gruppe «*im selben scharfen Grenzgebiet*» und  $n_2$  sowie  $SD_2$  die Anzahl Ortspaare resp. die Standardabweichung der Gruppe «*nicht im selben scharfen Grenzgebiet*». <sup>12</sup>  $x_i$  ist die syntaktische Distanz eines Ortspaares <sub>$i$</sub> .

Für eine Beurteilung der Cohen  $d$ -Effektstärke benutzt Cohen (1988) folgende Richtlinien:

- $d = 0.2$  entspricht einem schwachen Effekt
- $d = 0.5$  entspricht einem mittleren Effekt
- $d = 0.8$  entspricht einem starken Effekt

Es gilt zu beachten, dass  $d$  die Differenz zwischen den Mittelwerten in Standardabweichungseinheiten darstellt. Ein Cohens  $d = 0.5$  bedeutet beispielsweise, dass die beiden Mittelwerte

<sup>12</sup>Cohens  $d$  mit einer gewichteten und gemittelten Standardabweichung wird manchmal auch als Hedgens  $g$  bezeichnet (Volker, 2006)

eine halbe Standardabweichung auseinander liegen (Cohen, 1992). Ist  $d > 1$ , so liegen die Mittelwerte weiter als eine Standardabweichung auseinander.

Es wird erwartet, dass die Mittelwerte der syntaktischen Distanzen der Gruppen «*Ortspaare, zwischen denen eine scharfe Grenze liegt*» signifikant grösser sind als die Mittelwerte der syntaktischen Distanzen der Gruppen «*Ortspaare, zwischen denen keine scharfe Grenze liegt*». Da Sprache sich über die Zeit verändert (Croft, 2000), wird vermutet, dass die Kantons- und die Konfessionsgrenzen als älteste der drei untersuchten scharfen Grenzen den stärksten Einfluss auf die syntaktischen Unterschiede haben. Dies würde auch mit der historischen Bedeutung der beiden Grenzarten übereinstimmen. Den geringsten Einfluss auf die syntaktischen Unterschiede sollten schliesslich die neuzeitlichen Arbeitsmarktregionen und ihre Pendler aufzeigen.

### 5.3.4 Lineare, multiple Regressionsmodelle

In diesem Abschnitt werden die Modellierungen der syntaktischen Distanzen vorgestellt. Es handelt sich um einen Vergleich mehrerer linearer, multipler Regressionsmodelle. Eine lineare Regression ist ein statistisches Verfahren zur Erklärung und Überprüfung eines allfälligen Zusammenhangs zwischen einer abhängigen von einer unabhängigen Variable (Backhaus et al., 2011). Sie basiert auf der Idee, dass man den Zusammenhang zwischen der abhängigen Variable  $y$  und der unabhängigen Variablen  $x$  mit einer linearen Gleichung  $f(x)$  mathematisch beschreiben kann (Universität Zürich, 2017a):

$$y = f(x) \tag{5.13}$$

Die Regressionsgerade ist durch folgendes Regressionsmodell definiert:

$$y_k = \beta_0 + \beta_1 \times x_k + \epsilon_k \tag{5.14}$$

$y_k$  ist der Wert der abhängigen Variable eines Datenmesspunkts  $k$ .  $x_k$  ist der Wert der unabhängigen Variable eines Datenmesspunkts  $k$ .  $\beta_0$  und  $\beta_1$  sind die Regressionskoeffizienten, wobei  $\beta_1$  in einer linearen Gleichung der Steigung  $m$  und  $\beta_0$  dem y-Achsenabschnitt entspricht (vgl. Abb. 5.17).  $\beta_1$  beschreibt die Veränderung der abhängigen Variable  $y$ , wenn die unabhängige Variable  $x$  um eine Einheit ansteigt. Da die Regressionsgerade in der Regel nicht direkt auf den Messpunkten liegt, beinhaltet das Regressionsmodell mit  $\epsilon_i$  einen zusätzlichen stochastischen Fehlerterm des Datenmesspunkts  $k$ . Dieser beschreibt alle zusätzlichen Einflüsse auf die abhängige Variable  $y_i$ , die nicht durch die unabhängige Variable  $x_k$  erklärt werden können (Universität Zürich, 2017a). Abbildung 5.17 zeigt als Beispiel das Streudiagramm von fünf fiktiven Datenpunkten  $y_1$  bis  $y_5$  (blau) und deren Regressionsgerade (rot), die mittels der Methode der kleinsten Quadrate (vgl. Formel 5.15) durch die Datenpunkte gelegt wurde (Universität Zürich, 2017a):

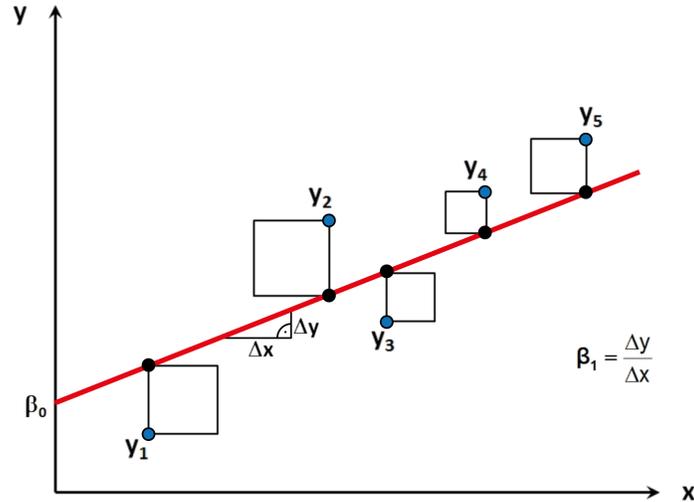


Abbildung 5.17: Die rote Regressionsgerade wird so durch das Streudiagramm der blauen Datenpunkte gelegt, dass die Residuenquadratsumme minimiert ist. Bildquelle: modifiziert nach Universität Zürich (2017a).

Vereinfacht gesagt wird bei einer linearen Regression eine Regressionsgerade so durch das Streudiagramm der Datenpunkte gelegt, dass die *Summe der quadrierten vertikalen Abstände* (auch Residuenquadratsumme genannt) von den Datenpunkten zur Regressionsgerade mathematisch minimiert ist (Backhaus et al., 2011; Universität Zürich, 2017a):

$$\sum_{k=1}^K e_k^2 = \sum_{k=1}^K [y_k - (\beta_0 + \beta_1 \times x_k)]^2 \rightarrow \min! \quad (5.15)$$

$K$  = Anzahl der Datenpunkte

$k = 1, 2, \dots, K$

$e_k$  = Abweichung des Datenpunkts  $k$  von der Regressionsgerade

Abbildung 5.18 zeigt am Beispiel der Ortschaft Küblis den Zusammenhang zwischen den euklidischen und syntaktischen Distanzen seiner 50 nächsten Nachbarorte. Nach der Regressionsmodellgleichung 5.14 ergibt sich folgendes Modell (Formel 5.16), mit der die lineare Abhängigkeit der syntaktischen von den euklidischen Distanzen zwischen Küblis und seinen 50 nächsten Nachbarn beschrieben wird:

$$\text{syntaktische Distanz} = \beta_0 + \beta_1 \times \text{euklidische Distanz} + \epsilon_k \quad (5.16)$$

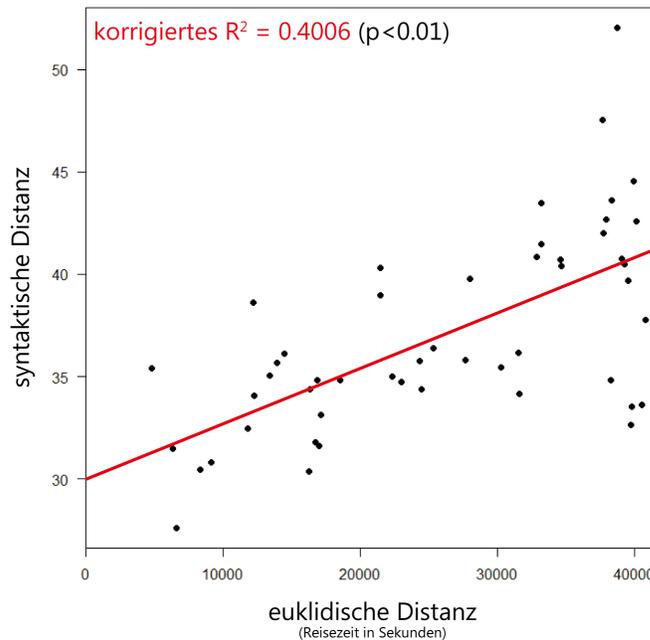


Abbildung 5.18: Ein Punkteplot der syntaktischen ( $y$ -Achse) und euklidischen Distanzen ( $x$ -Achse) der 50 nächsten Nachbarn von Küblis sowie die rote Regressionslinie des linearen Regressionsmodells.

Für die Regressionsgerade des Regressionsmodells von Formel 5.16 entspricht  $\beta_0 = 30,02$  und  $\beta_1 = 0,00027$ . Ein wenig komplizierter wird es, wenn man eine abhängige Variable durch *mehrere* unabhängige Variablen beschreiben möchte. Dazu eignen sich lineare, multiple Regressionsmodelle. Sie folgen dem gleichen Aufbau wie einfache, lineare Regressionsmodelle, besitzen aber mehrere Regressionskoeffizienten  $\beta_1$  bis  $\beta_n$ , wobei  $n$  der Anzahl der unabhängiger Variablen entspricht:

$$y = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \dots + \beta_n \times x_k + \epsilon_k \quad (5.17)$$

Angewandt auf die scharfen und unscharfen Grenzen ergibt sich für die syntaktische Distanz folgendes Regressionsmodell:

$$\text{synt. Distanz} = \beta_0 + \beta_1 \times \text{Distanzmass} + \beta_2 \times \text{Kant (0/1)} + \beta_3 \times \text{Konf (0/1)} + \beta_4 \times \text{AmRe (0/1)} + \epsilon \quad (5.18)$$

Das «*Distanzmass*» entspricht einer der unscharfen Grenzen ED, LCP, GI-ED oder GI-LCP. Die restlichen drei Variablen sind sogenannte *dichotom-kategorische* Regressionskoeffizienten, wobei «*Kant*» den Kantonsgrenzen, «*Konf*» den Konfessionsgebietsgrenzen und «*AmRe*» den Arbeitsmarktregionsgrenzen entspricht. Diese dichotom-kategorischen Regressionskoeffizienten stellen die acht möglichen Fälle der scharfen Grenzen für ein Ortspaar  $k$  dar und unterscheiden für jede scharfe Grenzart zwischen den beiden Fällen «*Es liegt eine Grenze zwischen dem Ortspaar*» und «*Es liegt keine Grenze zwischen dem Ortspaar*» (vgl. Tabelle 5.5). Es gilt zu beachten, dass in der vorliegenden Arbeit wegen Platzgründen auf eine explizite Untersuchung der Unterschiede zwischen den acht Fällen der Tabelle 5.5 verzichtet wird. Es wird somit nicht untersucht, ob der Einfluss von Grenzen auf Sprachunterschiede steigt, wenn mehrere Grenzarten zusammen vorkommen (vgl. Abschnitt 2.1). Dies würde eine mögliche zukünftige Erweiterung der hier vorgestellten Methode darstellen.

Tabelle 5.5: Mögliche acht Fälle der scharfen Grenzen, die zwischen einem Ortspaar liegen können.

Fall	Kantonsgrenze?	Konfessionsgrenze?	Arbeitsmarktregionsgrenze?
1.	Nein	Nein	Nein
2.	Ja	Nein	Nein
3.	Nein	Ja	Nein
4.	Nein	Nein	Ja
5.	Ja	Ja	Nein
6.	Ja	Nein	Ja
7.	Nein	Ja	Ja
8.	Ja	Ja	Ja

Abbildung 5.19 zeigt am Beispiel der Kantonsgrenzen und der euklidischen Distanzen die Streudiagramme, Regressionsmodelle und Regressionsgeraden der 50 nächsten Nachbarn von Küblis. 23 der 50 Nachbarorte liegen im selben Kanton wie Küblis. Zwischen den restlichen 27 Nachbarorte und Küblis liegt eine (oder auch mehrere) Kantonsgrenzen. Der kleine Plot in der oberen Hälfte von Abbildung 5.19 zeigt die 50 nächsten Nachbarn und ihre syntaktischen Distanzen zu Küblis, aufgeteilt in die zwei Gruppen «*Es liegt eine Kantonsgrenze zwischen Nachbarort<sub>i</sub> und Küblis*» und «*Es liegt keine Kantonsgrenze zwischen Nachbarort<sub>i</sub> und Küblis*». Die beiden grösseren Plots in der unteren Hälfte von Abbildung 5.19 stellen die einzelnen Streudiagramme der beiden Gruppen dar, wobei die y-Achsen die syntaktische Distanz und die x-Achsen die euklidische Distanz zwischen den Nachbarorten und Küblis beschreiben.

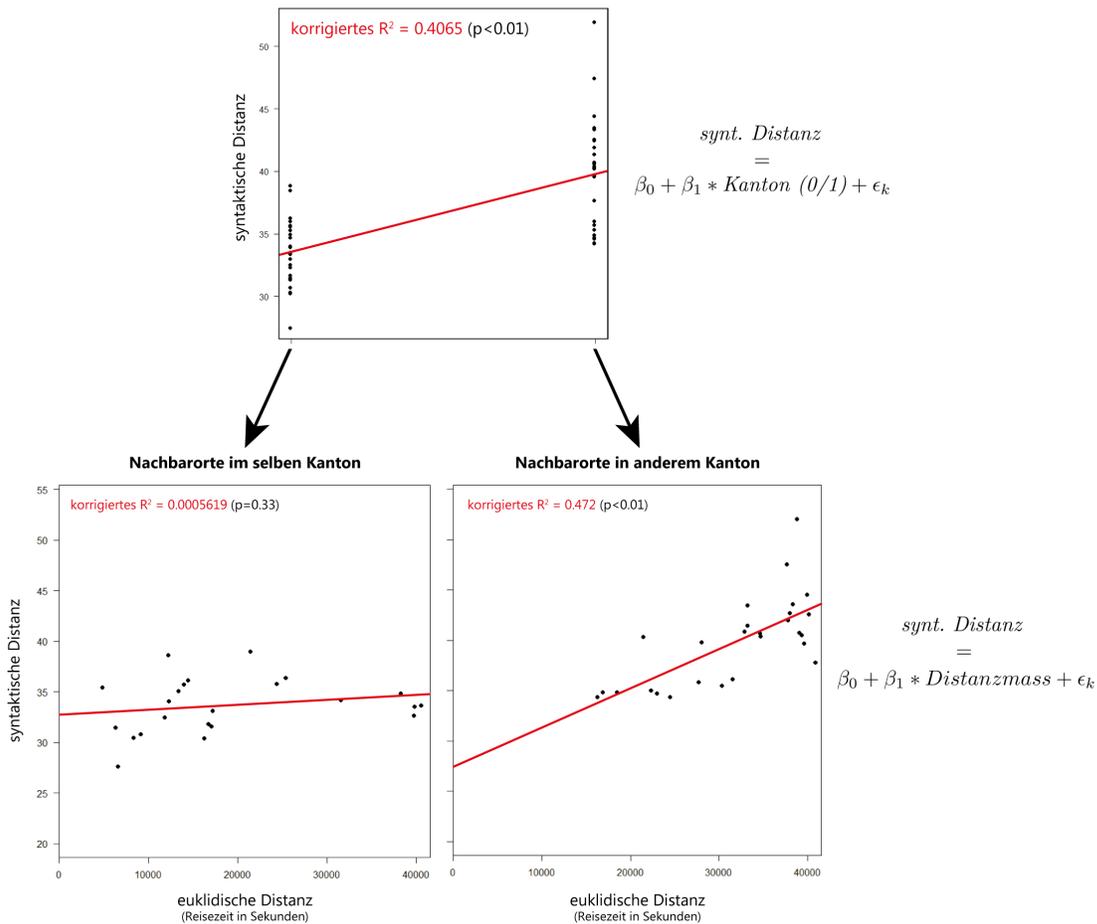


Abbildung 5.19: Streudiagramme, Regressionsmodelle und ihre Regressionsgeraden der 50 nächsten Nachbarn von Küblis, welche anhand des dichotom-kategorischen Regressionskoeffizienten *Kantonsgrenzen* aufgeteilt sind in zwei Gruppen

Eine Stärke von linearen, multiplen Regressionsmodellen ist, dass diese durch die Regressionskoeffizienten den Einfluss der verschiedenen erklärenden Variablen auf die abhängige Variable einzeln beschreiben. Regressionskoeffizienten zeigen den Effekt der Änderung einer unabhängigen Variable (die Grenzen) auf die abhängige Variable (syntaktische Distanz) an, wobei sie kein Massstab sind für die Wichtigkeit der Koeffizienten. Ein Vergleich der Regressionskoeffizienten ist nur bei einer Standardisierung<sup>13</sup> möglich (Backhaus et al., 2011). Wegen der hohen Anzahl der hier berechneten Modelle wird auf eine Einzeluntersuchung des Einflusses der einzelnen Grenzarten verzichtet. Es wird stattdessen nur das *korrigierte Bestimmtheitsmass*  $R_{\text{kor}}^2$  der Modelle betrachtet.

<sup>13</sup>Gemäss Backhaus et al. (2011) ist bei der Beurteilung der Wichtigkeit standardisierter Regressionskoeffizienten allerdings Vorsicht geboten, da ihre Aussagekraft durch allfällige Korrelationen (sog. Multikollinearitäten) zwischen den unabhängigen Variablen stark beeinträchtigt werden kann.

Das **Bestimmtheitsmass  $R^2$**  misst die Anpassung der Regressionsfunktion an die Datenpunkte und kann Werte zwischen 0 und 1 annehmen. Es zeigt bei linearen Regressionsmodellen an, wieviel Varianz der abhängigen Variable (im vorliegenden Fall die syntaktische Distanz) sich durch die unabhängigen Variablen (die untersuchten Grenzen) erklären lässt und berechnet sich wie folgt:

$$R^2 = 1 - \frac{\sum_{k=1}^K e_k^2}{\sum_{k=1}^K (y_k - \bar{y})^2} = 1 - \frac{\text{nicht erklärte Streuung}}{\text{Gesamtstreuung}} \quad (5.19)$$

$K$  = Anzahl der Datenpunkte

$k = 1, 2, \dots, K$

$e_k$  = Abweichung des Datenpunkts  $k$  von der Regressionsgerade

$y_k$  = Beobachtungswert des Datenpunkts  $k$

$\bar{y}$  = Durchschnitt der Beobachtungswerte

Dabei entspricht die *nicht erklärte Streuung* der Residuenquadratsumme (vgl. Formel 5.15) und die *Gesamtstreuung* der Summe der quadrierten Gesamtabweichungen aller Beobachtungen. Ein  $R^2$ -Wert von 0 bedeutet, dass kein linearer Zusammenhang zwischen der abhängigen und den unabhängigen Variablen besteht. Ein Wert von 1 bedeutet, dass das Regressionsmodell 100% der Varianz der abhängigen Variable erklärt. Das Bestimmtheitsmass ist durch die Anzahl der unabhängigen, erklärenden Variablen beeinflusst. Dies erschwert den Vergleich von Regressionsmodellen, welche eine unterschiedlich grosse Anzahl an erklärenden Variablen beinhalten. Das *korrigierte Bestimmtheitsmass  $R_{korr}^2$*  berücksichtigt diesen Sachverhalt (Backhaus et al., 2011):

$$R_{korr}^2 = R^2 - \frac{J \times (1 - R^2)}{K - J - 1} \quad (5.20)$$

$K$  = Anzahl der Datenpunkte

$J$  = Anzahl der unabhängigen Variablen

$K - J - 1$  = Anzahl der Freiheitsgrade

Es unterscheidet sich von  $R^2$  dadurch, dass es Regressionsmodelle bestraft, die viele unabhängige Variablen miteinbeziehen und folgt damit dem Sparsamkeitsprinzip von *Okhams Rasiermesser*<sup>14</sup> (vgl. auch Kapitel 8).

<sup>14</sup> Prinzip aus der Wissenschaft, nach dem bei mehreren möglichen Erklärungen diejenige ausgewählt werden soll, welche am einfachsten ist (Domingos, 1999).

Für jede der in dieser Arbeit untersuchten acht Nb-Skalen werden acht unterschiedliche Regressionsmodelle berechnet. Diese unterscheiden sich zum einen im verwendeten Distanzmass ED, LCP, GI-ED, GI-LCP (= unscharfe Grenze). Zum anderen dadurch, ob die scharfen Grenzen der Kantone, Konfessionsgebiete und Arbeitsmarktregionen miteinbezogen werden. Dies ermöglicht es zu sehen, ob die scharfen Grenzen einen Einfluss auf die erklärte Varianz der syntaktischen Distanz haben. Insgesamt werden so 64 verschiedene lineare, multiple Regressionsmodelle<sup>15</sup> berechnet. Erwartet wird, dass die Regressionsmodelle mit scharfen Grenzen die Varianz der syntaktischen Distanz besser erklären als die Modelle ohne scharfe Grenzen. Um zu überprüfen, ob sich die Residuenquadratsumme der unterschiedlichen Regressionsmodelle signifikant voneinander unterscheiden, werden für die zwei Regressionsmodelle mit jeweils denselben unscharfen Grenzen (= Distanzmass) ED, LCP, GI-ED oder GI-LCP innerhalb der selben Nb-Skala sogenannte ANOVA-Tests<sup>16</sup> durchgeführt. Im Falle dieser Arbeit überprüft der ANOVA-Test, ob sich die Varianz der Residuenquadratsumme des Regressionsmodells *mit* Einbezug der scharfen Grenzen signifikant von der Varianz der Residuenquadratsumme des Regressionsmodells *ohne* Einbezug der scharfen Grenzen unterscheidet. Es wird dabei dasjenige Regressionsmodell bevorzugt, dessen Residuenquadratsumme am kleinsten ist und somit *die Varianz der syntaktischen Distanz* ( $\approx$  die syntaktischen Unterschiede der Ortschaften) am besten erklärt.

Damit die Resultate der linearen, multiplen Regressionsmodelle eine gewisse Aussagekraft besitzen, müssen einige Voraussetzungen für die Anwendung von Regressionsmodellen erfüllt sein (vgl. z.B. Backhaus et al., 2011). Auf eine Abhandlung dieser statistischen Voraussetzungen wird wegen Platzgründen verzichtet. Es sei aber erwähnt, dass die meisten Voraussetzungen für lineare, multiple Regressionsmodelle erfüllt sind. Hier soll auch auf das Kapitel 8 verwiesen werden, welches die Limitierungen der vorliegenden Arbeit beinhaltet.

---

<sup>15</sup> 8 Nb-Skalen  $\times$  4 unscharfe Grenzen  $\times$  2 (einmal mit und einmal ohne Miteinbezug der scharfen Grenzen) = 64 Regressionsmodelle

<sup>16</sup> Für eine ausführliche Beschreibung der ANOVA-Tests sei auf die Fachliteratur verwiesen, z.B. Backhaus et al., 2011

## Kapitel 6

# Resultate und Interpretationen

Die nachfolgende Beschreibung der Resultate folgt einem ähnlichen Aufbau wie Kapitel 5. Erstens folgt eine kurze Betrachtung zweier Visualisierungen des Topographie-Einflusses auf die Reisezeiten der Orte. Als zweites werden für die unscharfen Grenzen die Korrelationsboxplots, die einzelnen Korrelationskarten der vier Distanzmasse ED, LCP, GI-ED und GI-LCP und die daraus folgenden, zusammenfassenden Korrelations-Dominanzkarten beschrieben. Drittens folgen die Resultate der scharfen Grenzen: Mittelwerttests und die berechneten Effektstärken von Cohens  $d$  decken auf, ob Kantons-, Konfessionsgebiets- und Arbeitsmarktregionsgrenzen einen Einfluss auf die syntaktischen Distanzen haben und falls ja, wie stark dieser ist. Viertens zeigt ein Vergleich der 64  $R_{korrr}^2$ -Werte der Regressionsmodelle sowie die ANOVA-Testresultate auf, ob die Regressionsmodelle *mit* Einbezug der scharfen Grenzen die Varianz der syntaktischen Distanz signifikant besser erklären als die Regressionsmodelle *ohne* scharfe Grenzen. Wie in Abschnitt 5.3.1 beschrieben, werden alle Analysen mit unterschiedlichen Massstabsgrößen durchgeführt.

### 6.1 Unscharfe Grenzen: Distanzmasse

#### 6.1.1 Topographie-Einfluss auf die syntaktische Distanz

Für eine erste Betrachtung über den Einfluss der Topographie auf die Reisezeiten und deren Zusammenhang zur syntaktischen Distanz der Ortschaften werden diejenigen syntaktischen Distanzen der Ortspaare miteinander verglichen, deren Reisezeiten am stärksten resp. am schwächsten von der Topographie beeinflusst sind. Dazu dient der in Abschnitt 5.1.1.4 vorgestellte Topographie-Einflussfaktor auf die Reisezeit. Für den Vergleich werden alle 1045 Voronoi-Nachbarschaftspaare nach ihrem Topographie-Einflussfaktor sortiert und die syntaktischen Distanzen der 25, 50, 100, 250 und 500 Ortspaare mit Boxplots zusammengefasst, deren Reisezeiten am stärksten resp. am schwächsten von der Topographie beeinflusst sind (vgl. Abbildung 9.1 auf im Anhang auf Seite 88 zum besseren Verständnis). Zusätzlich werden die LCP-Pfade der fünf Gruppen kartographisch dargestellt. Abbildung 6.1 auf der Seite 55 ist in

einen oberen und unteren Bereich gegliedert. Die *obere* Hälfte zeigt die visualisierten *Least-Cost*-Pfade an. Blau sind die Pfade zwischen Orten, deren Reisezeiten schwach durch die Topographie beeinflusst sind. Rot sind die Pfade, deren Reisezeiten stark von der Topographie beeinflusst sind. Die LCPs der Ortspaare, deren Reisezeiten schwach von der Topographie beeinflusst sind, liegen ausnahmslos im flachen Mittelland oder in den Tälern der Alpengebiete. Stark von der Topographie beeinflusste Reisezeiten der LCPs finden sich im Gegensatz dazu nur im bergigen Alpenraum sowie der hügeligen Jura-Region. Die *untere* Hälfte der Graphik zeigt mittels Boxplot-Darstellungen<sup>1</sup> die Verteilungen der syntaktischen Distanzen der zwei Gruppen sowie deren Veränderung mit grösser werdender Aggregierungsskala. Die Boxplots zeigen einen deutlichen Zusammenhang zwischen Topographie-Einfluss auf die Reisezeiten und den syntaktischen Distanzen der Nachbarorte. Ortspaare, deren Reisezeiten am schwächsten von der Topographie beeinflusst sind (blaue Boxplots), haben im Schnitt eine deutlich geringere syntaktische Distanz als Ortspaare, deren Reisezeiten am stärksten von der Topographie beeinflusst sind (rote Boxplots). Die Interquartilsbereiche der blauen Boxplots sind im Vergleich zu den roten Boxplots deutlich kleiner, was auf eine geringere Streuung der syntaktischen Distanzen der «blauen» Ortspaare hindeutet. Auch ist die Anzahl Ausreisser bei den roten Boxplots und ab der Skala mit  $\pm 250$  Ortspaaren deutlich grösser.

**Interpretation:** Die Ergebnisse zeigen, dass die Topographie die syntaktischen Unterschiede zwischen den Ortschaften deutlich beeinflusst. Je stärker der Topographie-Einfluss auf die Reisezeiten der Orte, desto grösser sind auch ihre syntaktischen Unterschiede. So sind die Mittelwerte der syntaktischen Distanzen der Ortspaare, deren Reisezeiten wenig von der Topographie beeinflusst werden, deutlich kleiner im Vergleich zu den Mittelwerten der Ortspaare, deren Reisezeiten stark von der Topographie beeinflusst werden. Zusätzlich lässt sich erkennen, dass die Streuung der syntaktischen Unterschiede der Ortspaare, deren Reisezeiten stark von der Topographie beeinflusst sind, viel grösser ist als die Streuung der syntaktischen Unterschiede der Ortspaare, deren Reisezeiten wenig von der Topographie beeinflusst sind. Die grössere Streuung kann ein Hinweis darauf sein, dass in den bergigen Regionen neben der Topographie noch weitere Faktoren die sprachlichen Unterschiede beeinflussen. Die Abnahme und Angleichung der Varianz resp. der Boxplots bei den  $\pm 500$  Ortspaargruppe folgt aus der hohen Anzahl von je 500 Ortspaaren (von Total 1045 Voronoi-Nachbarschaften), aufgrund derer die durchschnittlichen Unterschiede der blauen zur roten Gruppe viel geringer sind als bei den kleineren Skalen (wie z.B.  $\pm 25$  oder  $\pm 50$ ). Die sich nicht überlappenden Kerben der roten und blauen Boxplots deuten auf einen statistisch signifikanten Unterschied zwischen den zentralen Mittelwerten beider Gruppen hin (vgl. Boxplot-Erklärung 9.2 im Anhang). Die Resultate bestätigen die These, dass Kontakt wichtig für die Sprachveränderung ist. Je schwieriger der Kontakt resp. je grösser die vorhandenen Kontaktbarrieren in Form der Topographie, desto grösser sind auch die sprachlichen Unterschiede zwischen den Ortschaften.

---

<sup>1</sup>für eine Erklärung der Boxplot-Darstellung siehe Abb. 9.2 im Anhang

## Voronoi-Ortspaare mit **geringstem** resp. **grösstem** Einfluss der Topographie auf die Reisezeit und syntaktische Distanz zwischen diesen

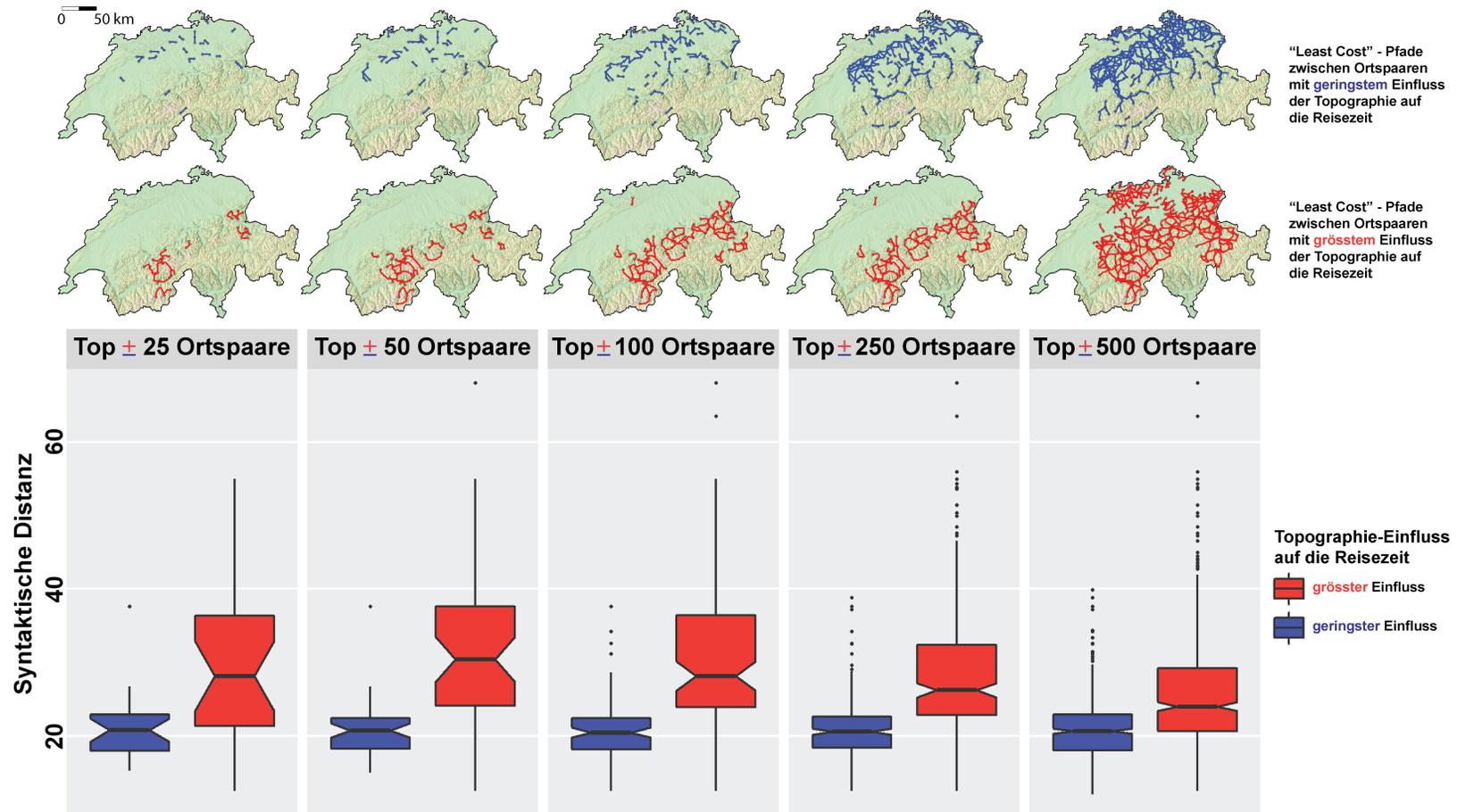


Abbildung 6.1: Die jeweils 25, 50, 100, 250 und 500 von insgesamt 1045 Voronoi-Ortspaaren, deren Reisezeiten am schwächsten resp. stärksten von der Topographie beeinflusst werden. *Oben:* Visualisierungen der LCPs. Blau sind die Pfade derjenigen 1045 Nachbarschaftsorte, deren Reisezeiten am wenigsten von der Topographie beeinflusst sind. Rot sind die Pfade der Orte, deren Reisezeiten am stärksten von der Topographie beeinflusst sind. *Unten:* Boxplots der syntaktischen Distanzen der Ortspaare, deren Reisezeiten am schwächsten (blau) resp. am stärksten (rot) von der Topographie beeinflusst werden.

### 6.1.2 Korrelationsanalysen: Boxplots

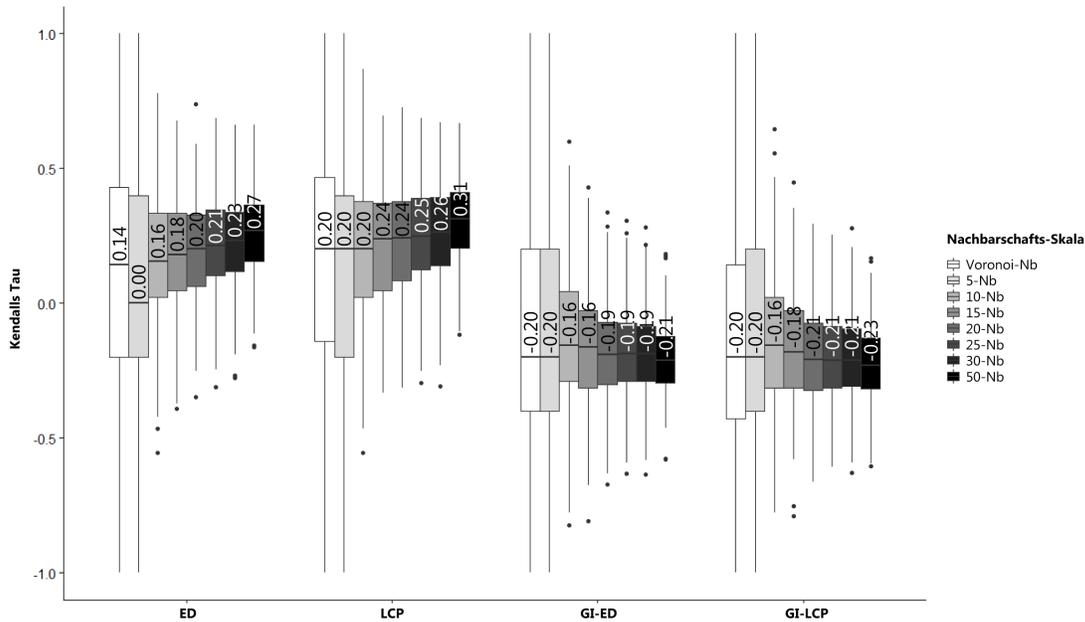


Abbildung 6.2: Boxplots der Korrelationswerte der syntaktischen Distanzen mit den vier unterschiedlichen Distanzmassen ED, LCP, GI-ED und GI-LCP sowie veränderten acht Nb-Skalen.

Die Korrelationswerte-Verteilungen der syntaktischen Distanzen zu den vier verschiedenen Distanzmassen unterscheiden sich für die unterschiedlichen Nb-Skalen nur gering. Wie erwartet sind die Mittelwerte der LCP-Distanzen aber für alle acht Nb-Skalen höher als bei den ED-Distanzen (vgl. Abb. 6.2). Ebenfalls wie erwartet liegen für die beiden Erreichbarkeitsmassen ED und LCP bei allen Nb-Skalen grösstenteils *positive* Korrelationen vor. Abgesehen von der 5-Nb-Skala der ED-Reisezeit liegen alle Median-Werte deutlich über Null und steigen bei grösser werdenden Nb-Skala stetig an. Erwartungsgemäss nimmt die Varianz der Korrelationswerte mit zunehmender Nb-Skala ab und gleicht sich langsam dem totalen Durchschnitt<sup>2</sup> an. Bei den Einflussbereichsmassen GI-ED und GI-LCP verhält es sich ähnlich wie bei den beiden Erreichbarkeitsmassen. Die Gravitations-Indizes korrelieren beide erwartungsgemäss grösstenteils *negativ* mit der syntaktischen Distanz. Alle Median-Werte liegen bei beiden Einflussbereichsmassen deutlich unter Null. Ab der 15-Nb-Skala und grösser sind die Mittelwerte der GI-LCP Distanz wie erwartet grösser als die der GI-ED Distanz. Im Vergleich zu den Erreichbarkeitsmassen steigen die Median-Werte der Korrelationen mit grösser werdender Nb-Skalen jedoch nicht an, sondern gleichen sich nach einem kurzen Anstieg bei der 10- und 15-Nb-Skala wieder dem Median der Voronoi- und 5-Nb-Skala an.

<sup>2</sup>Dies entspricht den Durchschnittskorrelationen von allen 383 SADS-Orten zusammen: 0.54 für ED, 0.58 für LCP, -0.42 für GI-ED und -0.43 für GI-LCP

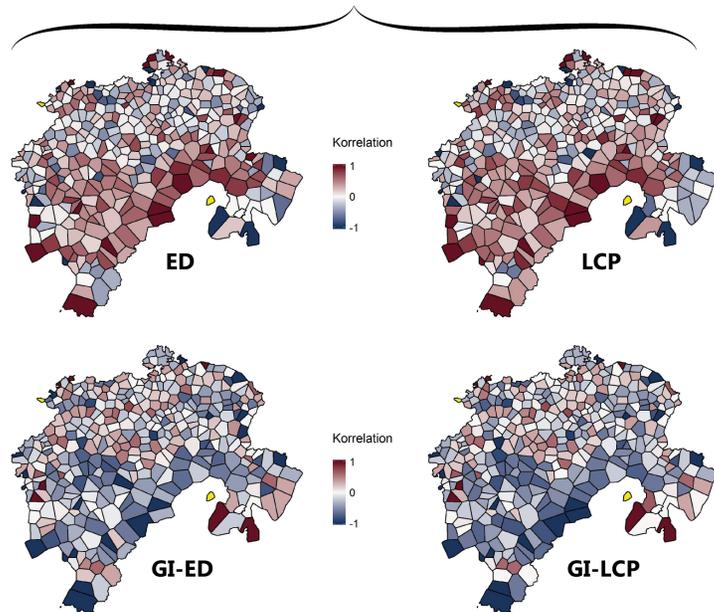
**Interpretation:** Obwohl die Korrelations-Mittelwerte der LCP gegenüber der ED über alle acht Nb-Skalen grösser sind, ist nicht klar ob diese Unterschiede eine statistische Signifikanz aufweisen. Ein Signifikanztest wurde nicht durchgeführt. Die Annahme, dass die LCP- stärker als die ED-Reisezeiten mit den syntaktischen Distanzen korrelieren, kann wegen den geringen Mittelwertunterschieden mit der Boxplotdarstellung nicht bestätigt werden. Beide Erreichbarkeitsmasse besitzen für die acht untersuchten Nb-Skalen jeweils ähnliche Korrelationswert-Verteilungen. Ebenso wie bei den Erreichbarkeitsmassen lässt sich die Hypothese, dass die GI-LCP Indizes durch Miteinbezugnahme der Topographie stärker mit den syntaktischen Distanzen korrelieren als die GI-ED Indizes, nicht bestätigen. Genauso wenig scheint der Einbezug der Bevölkerungsanzahl eine verstärkende Wirkung auf die Korrelation mit den syntaktischen Distanzen zu haben. Die absoluten Beträge der Einflussbereichsmass-Mediane ähneln denjenigen der Erreichbarkeitsmasse und zeigen keine stärkeren Korrelationen auf. Alle Distanzmasse korrelieren aber wie erwartet mit der syntaktischen Distanz.

### 6.1.3 Korrelationsanalysen: Korrelationskarten

Bei den flächenhaften Korrelationswertvisualisierung der vier Distanzmasse ED, LCP, GI-ED und GI-LCP zeigen sich vor allem bei grösser werdenden Nb-Skalen erkennbare Muster (siehe Abb. 6.3 sowie im Appendix Abb. 9.3, 9.4 und 9.5). Schon ab der Voronoi-Nb-Skala zeigt sich in den bergigen Alpenregionen eine stärkere generelle Tendenz zur jeweils erwarteten Korrelation als im flachen Mittelland (vgl. die linken vier Karten bei Abb. 6.3). Für die beiden Orte Obersaxen im Kanton Graubünden und Roggenburg im Kanton Basel-Landschaft können für die Voronoi-Nb-Skala keine Korrelationen berechnet werden, da beide mit Tamins resp. Kleinlützel jeweils nur einen direkten Voronoi-Nachbarort haben (vgl. die Visualisierung der Voronoi-Nachbarschaften bei Abb. 5.15). Auf den Korrelationskarten der Voronoi-Nb-Skala werden Obersaxen und Roggenburg deshalb gelb eingefärbt. Bei den kleinen Nb-Skalen der Erreichbarkeitsmasse ED und LCP korrelieren einige Ortschaften wider Erwarten negativ und für die Einflussbereichsmasse GI-ED und GI-LCP im Gegensatz dazu positiv. Ein anderes Bild zeigt sich bei den grösser werdenden Nb-Skalen. Die Korrelationswerte für die Erreichbarkeits- und Einflussbereichsmasse liegen ab dem Massstab der 15-Nb-Skala (und grösser) nur noch für wenige Ortschaften im jeweils konträr erwarteten Wertebereich. Ein Blick auf die Korrelationskarten der 50-Nb-Skala zeigt, dass die Korrelationen der beiden Erreichbarkeitsmasse ED und LCP mit den syntaktischen Distanzen für fast alle Ortschaften im positiven Bereich liegen, wobei im Vergleich die LCP-Karte noch ein wenig stärkere Korrelationen aufzeigt als die ED-Karte (Abb. 6.3). Die Alpengebiete weisen eine stärkere Korrelation zu den Erreichbarkeitsmassen auf als das flache Mittelland, wobei auch hier die LCP-Karte im Vergleich zur ED-Karte stärkere positive Korrelationen aufweist. Bei den Korrelationskarten der beiden Einflussbereichsmasse GI-ED und GI-LCP fallen bei der 50-Nb-Skala weisse und leicht rote, d.h. nicht resp. leicht positiv korrelierende, Gebiete um die Ortschaft Guttannen auf (vgl. die beiden unteren Korrelationskarten auf der rechten Seite von Abb. 6.3, Guttannen ist das dick umrandete Polygon auf der GI-ED Karte der 50-Nb-Skala). Dies ist entgegen den erwarteten negativen Korrelatio-

nen der Einflussbereichsmasse. Bei der kleinen Voronoi-Nb-Skala weisen die selben Ortschaften hingegen die erwartete negative Korrelation der Einflussbereichsmasse mit den syntaktischen Distanzen auf. Bei der 50-Nb-Skala zeigen sich neben den weissen Gebieten um die Ortschaft Guttannen auch bei anderen Regionen schwach korrelierende Gebiete (vgl. rechte vier Karten von Abb. 6.3).

## direkte Voronoi-Nachbarn



## 50 nächste Nachbarn

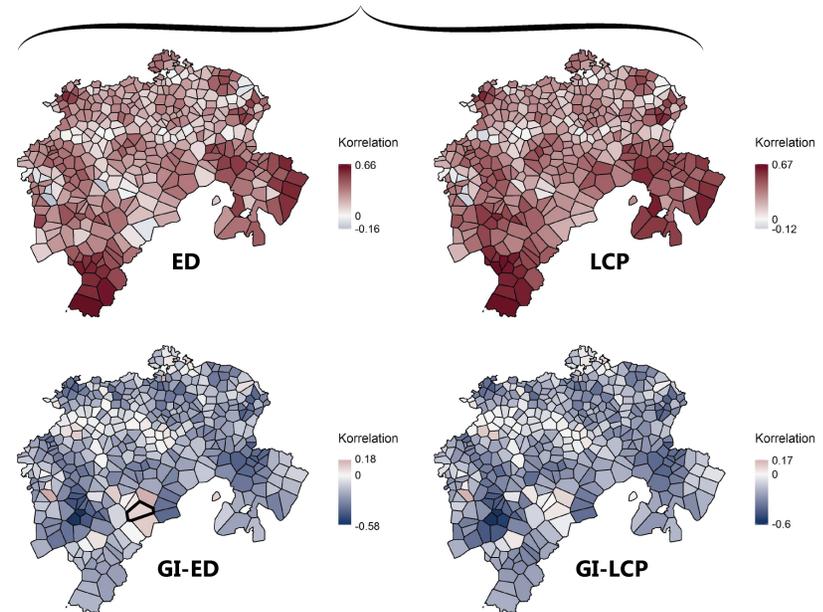


Abbildung 6.3: Korrelationskarten der vier Distanzmasse zur syntaktischen Distanz auf der Nb-Skala der jeweils angrenzenden Voronoi-Nachbarn (linke vier Karten) und der jeweils 50 nächsten Nachbarn (rechte vier Karten).

**Interpretation:** Eine mögliche Erklärung für die leicht positiv korrelierenden Gebiete bei den Einflussbereichsmassen der 50-Nb-Skala um die Ortschaft Guttannen sind die Berner Alpen (vgl. Abb. 6.4 linke Karte), welche die dortigen Ortschaften erreichbarkeitsmässig von der restlichen Deutschschweiz stark isolieren. Diese Erklärung stützt auch Abbildung 6.5 (linke Karte). Sie zeigt die Abweichungen der lokalen von der durchschnittlichen, syntaktischen Distanz deutschschweizer Orte und deckt so die Gebiete auf, deren lokale Syntax sich besonders stark vom Rest der Deutschschweiz unterscheidet<sup>3</sup> Die Orte rund um Guttannen (vgl. Abb. 6.5), deren syntaktischen Unterschieden zum Rest der Deutschschweiz am grössten sind, sind grösstenteils dieselben wie die leicht positiv korrelierenden Orte bei den GI-ED und GI-LCP Korrelationskarten der 50-Nb-Skala (vgl. Abb. 6.3). Die Diskrepanz zwischen den erwartungsgemäss korrekten Korrelationen bei der Voronoi-Nb-Skala und den unerwarteten Korrelationen beim Massstab der 50-Nb-Skala lässt sich durch Gravitations-Indizes der bevölkerungsschwachen Alpengebieten erklären (vgl. Abb. 6.4 rechte Karte), welche bei grossen Nb-Skalen mit den starken GI-Indizes im Flachland aggregiert werden und so einen schwächeren Zusammenhang mit den syntaktischen Distanzen aufzeigen. Dies im Gegensatz zur kleinräumigen Voronoi-Nb-Skala, bei der die lokalen GIs der Orte um Guttannen nicht mit denjenigen der bevölkerungsreichen Flachlandgebiete zusammengefasst werden und diese so die erwarteten negativen Korrelationen aufweisen (vgl. Abb. 6.5 rechte Karte). Die jeweils entgegen der Erwartung – mit falschem Vorzeichen – korrelierenden Gebiete bei den kleinen Nb-Skalen lassen sich zum einen über Randeffekte erklären, zum anderen durch die geringe Anzahl an Nachbarorten. Dadurch können unerwartete Korrelationen stochastisch häufiger auftreten<sup>4</sup>. Bei der 50-Nb-Skala zeigt die leicht deutlicheren Korrelationen der LCPs gegenüber der ED, dass der Miteinbezug der Topographie in die Distanzen die syntaktischen Unterschiede besser erklärt, als wenn nur die euklidischen Distanzen verwendet werden. Interessant an den vier Korrelationskarten der 50-Nb-Skala ist, dass die weissen Flächen, die eine keine Korrelation anzeigen, deckungsgleich sind mit den Dialektregionsgrenzen der Morphologie und Phonetik, die Sandra Kellerhals (2014) in ihrer Masterarbeit aufgedeckt hat (vgl. Abb. 6.6). Es scheint, dass die Korrelationskarten der syntaktischen Distanzen mit den vier unscharfen Grenzen ED, LCP, GI-ED und GI-LCP auf einer 50-Nb-Skala eine Approximation der Deutschschweizer Dialektregionen bilden.

<sup>3</sup>Dazu wurden für jeden Ort die durchschnittlichen syntaktischen Distanzen seiner 50 nächsten Nachbarn ermittelt und deren Abweichung von der durchschnittlichen syntaktischen Distanz der 50 nächsten Nachbarschaften aller 383 Ortschaften berechnet. Da sich die Karten der unterschiedlichen Nb-Skalen für die Abweichung der lokalen syntaktischen Distanzen (linke Karte) und für die GI-LCP Werte (rechte Karte) im Vergleich zum GI-ED Mass kaum voneinander unterscheiden, wurde aus Platzgründen auf ihre Darstellungen verzichtet und beispielhaft nur die Karte der 50-Nb-Skala in die Arbeit genommen.

<sup>4</sup>Das mathematische *Gesetz der grossen Zahlen* besagt, dass sich die relative Häufigkeit eines Zufallsergebnisses umso mehr der erwarteten theoretischen Wahrscheinlichkeit nähert, je öfters das Zufallsexperiment wiederholt wird () (vgl. z.B. Kolmogoroff, 1933). Da der Datensatz des SADS eine Stichprobe der Grundgesamtheit darstellt, trifft dieses Gesetz auch auf die SADS-Daten zu.

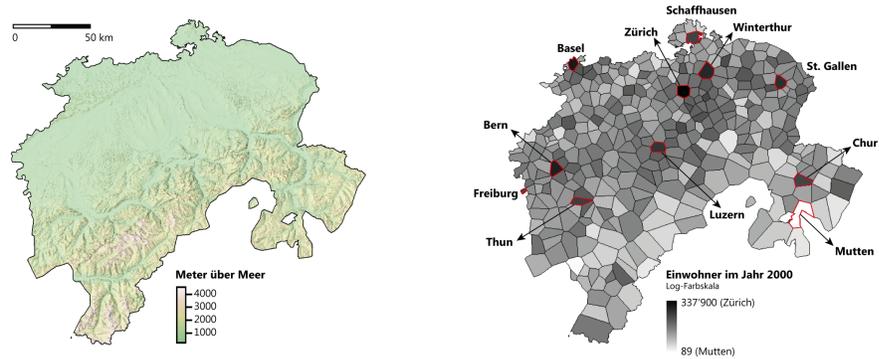


Abbildung 6.4: Topographie der Deutschschweiz (links) und Einwohnerzahl der SADS-Orte mit den elf bevölkerungsreichsten SADS-Ortschaften (rechts). Die Farbskala der rechten Karte ist zwecks besserer Sichtbarkeit logarithmisch.

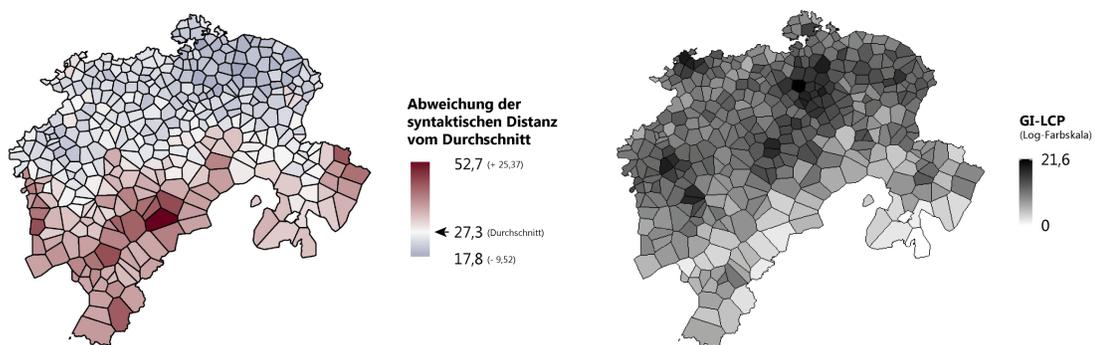


Abbildung 6.5: Die Abweichungen der syntaktische Distanz von durchschnittlichen syntaktischen Distanz der Deutschschweiz (links) und die durchschnittlichen GI-LCPs der jeweils 50 nächsten Nachbarorte (rechts). Zürich besitzt mit einem GI-LCP von 21,6 den grössten Einflussbereich der Deutschschweiz, die Ortschaft Avers im Kanton Graubünden mit einem GI-LCP von 0,0002 den kleinsten.

## 50 nächste Nachbarn

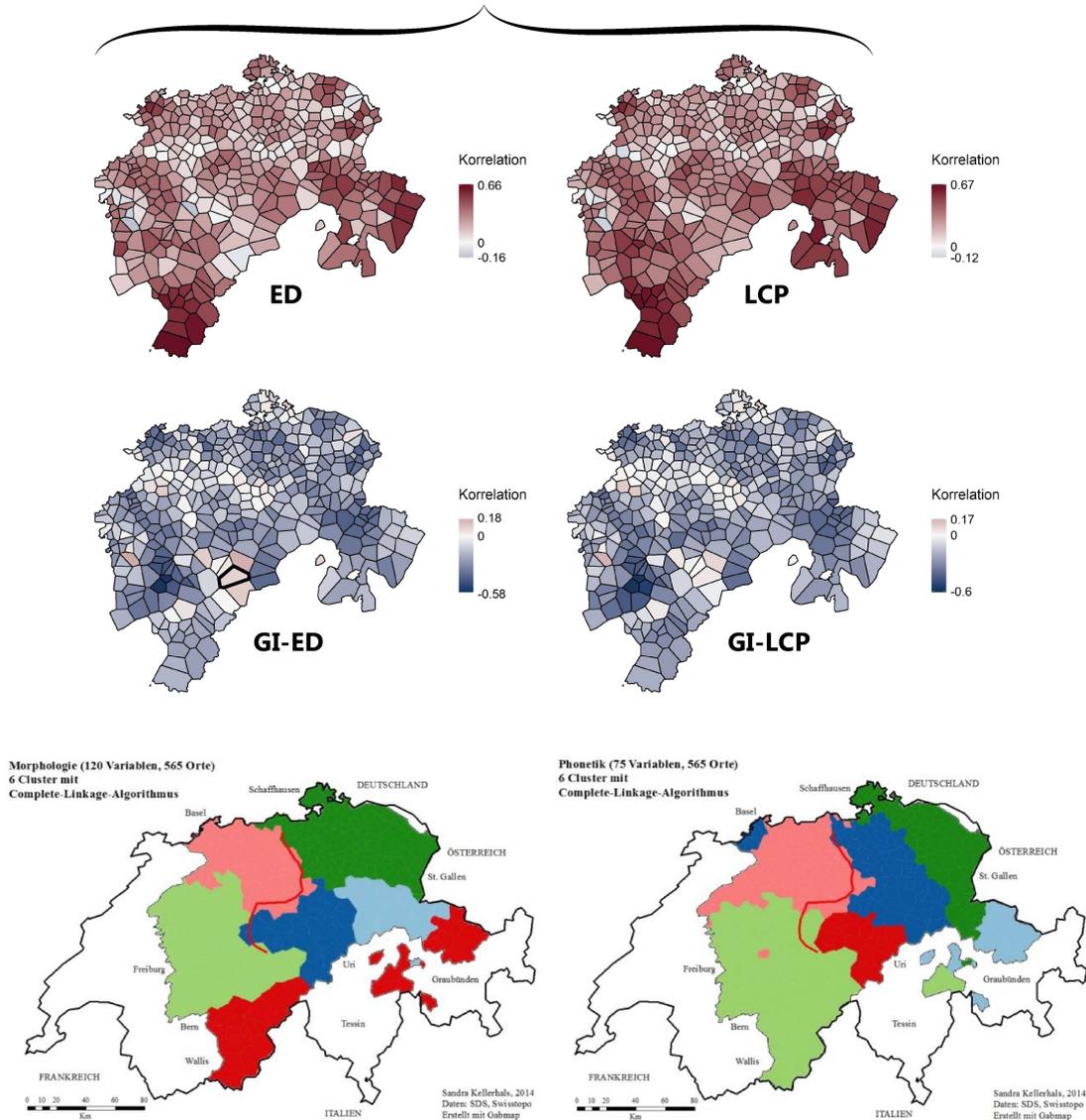


Abbildung 6.6: Der Vergleich der Korrelationswerte 50-Nb-Skala (obere vier Karten) mit den Dialektregionen der Deutschschweiz nach der Morphologie (links) und Phonetik (rechts), wie sie Kellerhals (2014) aufzeigte (untere zwei Karten). Die rote Linie stellt die Kulturgrenze der Brünig-Napf-Reuss-Linie dar. Bildquelle: Kellerhals (2014). Die Grenzlinien der Dialektregionen ähneln den weissen resp. schwach korrelierenden Ortschaften der 50-Nb-Skala (obere vier Karten).

### 6.1.4 Korrelationsanalysen: Dominanzkarten

Abbildung 6.7 zeigt für jeden Ort, welches der vier Distanzmasse ED, LCP, GI-ED und GI-LCP am stärksten mit der syntaktisch Distanz korreliert<sup>5</sup>. Jede Ortschaft ist entsprechend des lokal stärksten Distanzmasses eingefärbt (vgl. Legende der Abb. 6.7). Damit sich die beiden Einflussbereichsmasse nicht gegenseitig konkurrieren, treten diese nicht gegeneinander an. Die obere Hälfte der Abbildung zeigt jeweils die dominierenden Distanzmasse ED, LCP und GI-LCP. Die untere Hälfte der Abbildung verwendet anstelle des GI-LCP den GI-ED. Bei den kleinen Massstäben der Voronoi- und der 5-Nb-Skala gewinnt nach der Menge an Ortschaften jeweils die LCP Distanz vor den anderen Distanzmassen. Ab der Skalengrösse 10-Nb und grösser dominiert hingegen ortszahlmässig das Einflussbereichsmass GI-LCP resp. GI-ED. Im Vergleich dazu verliert die LCP-Distanz bei denselben Skalengrössen stetig an Ortschaften. Die Anzahl Ortschaften, wo die euklidische Distanz (ED) dominiert, bleibt über alle Skalen (mit einer leichten Zunahme) konstant. Bei den Karten zeigen sich erst bei den grösseren Skalen räumliche Muster. Die GI-dominierten Ortschaften weisen beim Vergleich der zwei Karten der 30- und 50-Nb-Skalen ein sehr ähnliches Muster auf. Dasselbe zeigt sich auch bei den euklidisch dominierten Ortschaften. Die Region Zürich, die Ortschaften nördlich des Brienzer- und Thunersees, die Gebiete süd- und nordwestlich von Thun, die Gegenden um Appenzell sowie die Bereiche um Bad Ragaz und Chur sind bei den 30- und 50-Nb-Skalen von den ED-Korrelationen dominiert.

**Interpretation:** Die anzahlmässig grosse Dominanz der LCP gegenüber den beiden Einflussbereichsmassen GI-ED und GI-LCP deutet darauf hin, dass die Topographie kleinräumig einen grösseren Einfluss auf syntaktische Unterschiede hat als die Bevölkerungsanzahl der Ortschaften. Bei den grösseren Nb-Skalen kehrt sich dies jedoch um. Die erst bei den Massstäben der 25-, 30- und 50-Nb-Skala erkennbaren Muster weisen darauf hin, dass der Einfluss der Bevölkerungsanzahl grossflächiger ist als jener der Topographie. Hier zeigt sich wie in Abschnitt 5.3.1 beschrieben der Effekt des MAUP. Zu beachten ist, dass die Dominanzen nicht auf ihre Signifikanz getestet wurden. Die jeweiligen Unterschiede der vier Distanzmass-Korrelationen können deshalb für die Ortschaften klein ausfallen (vgl. die einzelnen Korrelationskarten von Abb. 6.3 und im Anhang Abb. 9.3, 9.4 und 9.5)

---

<sup>5</sup>Um einen Vergleich der Korrelationswerte zu ermöglichen, werden für die Dominanzkarten die ursprünglich negativen Korrelationswerte der Einflussbereichsmasse GI-ED und GI-LCP mit dem Faktor  $-1$  multipliziert.

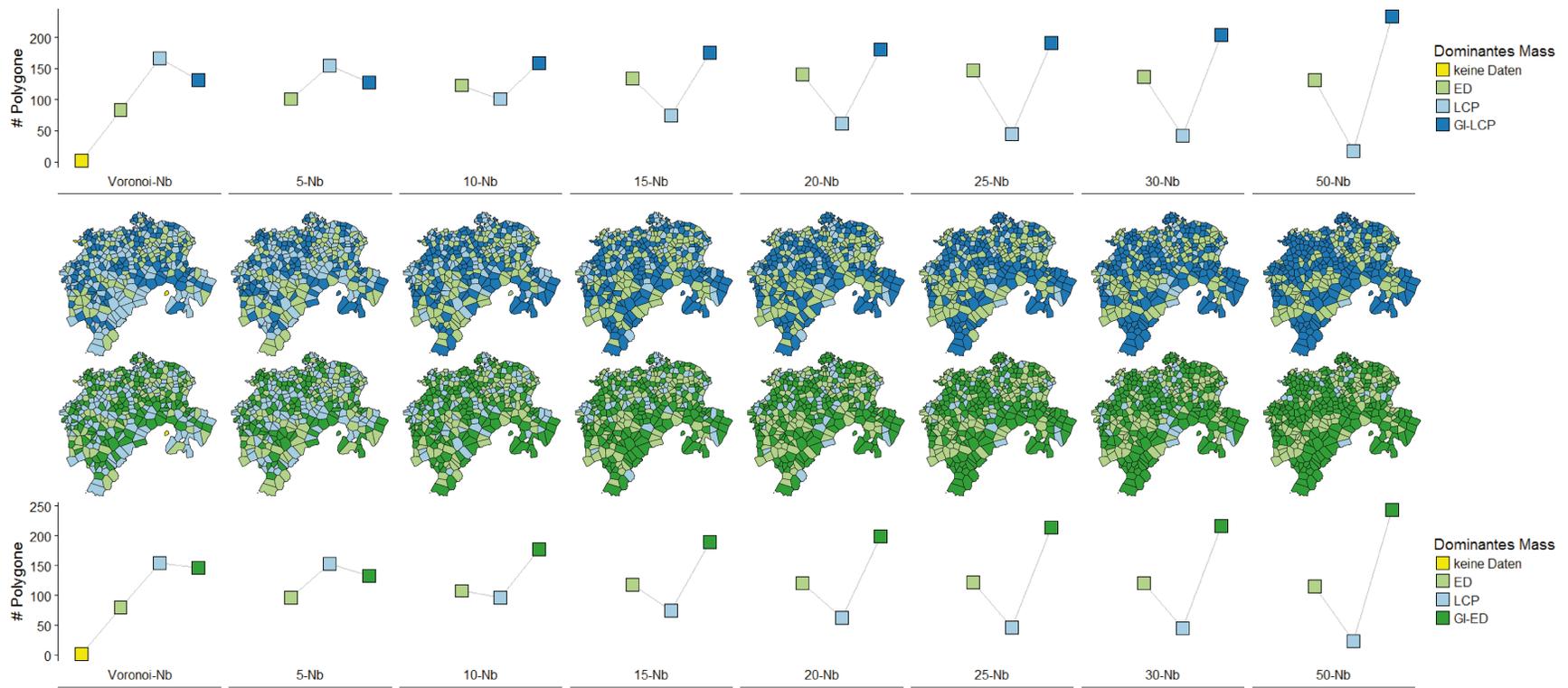


Abbildung 6.7: Dominanzkarten, welche der Distanzmasse die stärkste Korrelation aufweisen. *Obere Hälfte*: Stärkste Korrelation zwischen den syntaktischen Distanzen und den Distanzmasse ED, LCP und **GI-LCP**. *Untere Hälfte*: Wie in der oberen Hälfte der Abbildung werden auch in der unteren Hälfte die dominanten Korrelationen zwischen den syntaktischen Distanzen und den Distanzmassen ED und LCP visualisiert. Der Unterschied besteht aber darin, dass hier nicht das GI-LCP, sondern das **GI-ED** als Einflussbereichsmass verwendet wird.

## 6.2 Scharfe Grenzen: Kantone, Konfessionsgebiete und Arbeitsmarktregionen

Der Einfluss der scharfen Grenzen wurde mittels Mann-Whitney-U Test untersucht (vgl. Unterabschnitt 5.3.3). Die Nullhypothese gegen die getestet wurde beschreibt, dass der Mittelwert der Gruppe «*im selben Grenzgebiet*» gleich ist wie als der Mittelwert der Gruppe «*nicht im selben Grenzgebiet*».

$$H_0 : \mu_1 = \mu_2 \quad (6.1)$$

Dagegen lautet die **Alternativ-Hypothese**, dass der Mittelwert der Gruppe «*im selben Grenzgebiet*» kleiner ist als der Mittelwert der Gruppe «*nicht im selben Grenzgebiet*»:

$$H_1 : \mu_1 < \mu_2 \quad (6.2)$$

Getestet wird mit einem Signifikanzniveau  $\alpha = 0.01$  gegen die Null-Hypothese, womit die Wahrscheinlichkeit für einen Fehler 1. Art nicht grösser ist als 1%. Um zu untersuchen, ob eine allfällige Signifikanz der Unterschiede auch bedeutend ist, wird zusätzlich noch Cohens  $d$  als Mass der Effektstärke berechnet (vgl. Unterabschnitt 5.3.3)

### 6.2.1 Teststatistiken der Mittelwerttests

Die Teststatistiken der asymptotischen *Mann-Whitney-U*-Tests zeigen, dass die Kantonsgrenzen einen signifikanten Einfluss auf die Mittelwerte der syntaktischen Distanzen haben. Die zentrale Tendenz der syntaktischen Distanzen der Ortspaargruppe «*nicht im selben Kanton*» ist bei allen acht Nb-Skalen signifikant höher gegenüber dem Mittelwert der Ortspaargruppe «*im selben Kanton*» (vgl. Tabelle 6.1). Analoges gilt für die Konfessionsgrenzen und die Arbeitsmarktgrenzen. Auch dort ist der Mittelwert der syntaktischen Distanzen der Ortspaargruppe «*nicht im selben Konfessionsgebiet*» resp. «*nicht im selben Arbeitsmarktregionsgebiet*» bei allen acht Nb-Skalen signifikant höher gegenüber dem Mittelwert der Ortspaargruppe «*im selben Konfessionsgebiet*» resp. «*im selben Arbeitsmarktregionsgebiet*» (vgl. Tabellen 6.2 und 6.3). Somit kann die  $H_0$ -Hypothese für alle drei untersuchten scharfen Grenzen verworfen werden. Bei den Effektstärken von Cohens  $d$  zeigen sich für alle scharfen Grenzen und über alle Skalen mittlere ( $d = 0.5$ ) bis starke Effekte ( $d > 0.8$ ) (vgl. Abschnitt 5.3.3). Von den drei Grenzen weisen die Konfessionsgebietsgrenzen die stärksten Effekte auf die syntaktische Distanz auf, gefolgt von den Kantons- und den Arbeitsmarktregionsgrenzen. Abgesehen von den Kantonsgrenzen, bei denen die 50-Nb-Skala – knapp gefolgt von der Voronoi-Nb-Skala – den grössten Cohens  $d$ -Wert hat, weist bei den Konfessions- und Arbeitsmarktregionen die Voronoi-Nb-Skala die grössten Cohens  $d$ -Werte auf.

**Interpretation:** Die hohen Cohens  $d$ -Werte der Konfessionsgrenzen bestätigen auch linguistisch die Aussage von Schnyder (2013), wonach die Reformation zu den prägendsten Ereignissen der Schweizer Geschichte gehört. Dass die Kantonsgrenzen bei den grösseren Nb-Skalen höhere

Cohens  $d$ -Werte als die Arbeitsmarktregionsgrenzen aufweisen, könnten zwei Gründen erklären. Zum einen weisen die Kantone im Gegensatz zu den Arbeitsmarktregionen schroffere Grenzverläufe auf, zum anderen besitzen sie teilweise auch Enklavengebiete (vgl. Abb. 5.7 und 5.9). Beides kann dazu führen, dass Ortspaare bei grösseren Nb-Skalen der Gruppe «*im selben Kanton*» zugewiesen werden, die zwar im selben Kanton, jedoch nicht im selben, zusammengehörigen Kantonsgebiet liegen<sup>6</sup>. Solche Fälle können durch den Grenzverlauf der Arbeitsmarktregionen nicht auftreten, weshalb grössere Nb-Skalen Cohens  $d$  eher negativ beeinflussen. Solch unterschiedliche Resultate lassen sich durch das MAUP erklären (vgl. dazu auch die Ausführungen zum Interpretationsabschnitt von Abschnitt 6.3).

Tabelle 6.1: Kantonsgrenzen

Nb-Skala	Median <sup>†</sup>	Median <sup>††</sup>	$\mu^{\dagger}$	$\mu^{\dagger\dagger}$	$n^{\dagger}$	$n^{\dagger\dagger}$	$n_{total}$	$p$ -Wert <sup>*</sup>	$z$ -Wert	Cohens $d$
Voronoi	21.29	24.00	22.07	26.76	731	314	1045	0.00	8.08	0.72
5	21.36	23.93	22.10	25.90	690	270	960	0.00	7.21	0.66
10	21.71	24.51	22.65	26.57	1286	628	1914	0.00	10.52	0.60
15	21.94	25.12	23.02	27.20	1782	1069	2851	0.00	14.26	0.62
20	22.08	25.53	23.19	27.66	2202	1591	3793	0.00	18.04	0.64
25	22.18	25.74	23.32	27.93	2602	2140	4742	0.00	20.81	0.65
30	22.29	26.17	23.46	28.31	2971	2731	5702	0.00	23.78	0.67
50	22.66	27.40	23.93	29.59	4113	5451	9564	0.00	34.31	0.73

<sup>†</sup> Gruppe der Ortspaare, die jeweils im selben Kanton liegen

<sup>††</sup> Gruppe der Ortspaare, die jeweils *nicht* im selben Kanton liegen

<sup>\*</sup>  $\alpha = 0.01$  (asymptotischer Mann-Whitney  $U$  Test)

Tabelle 6.2: Konfessionsgebiete

Nb-Skala	Median <sup>†</sup>	Median <sup>††</sup>	$\mu^{\dagger}$	$\mu^{\dagger\dagger}$	$n^{\dagger}$	$n^{\dagger\dagger}$	$n_{total}$	$p$ -Wert <sup>*</sup>	$z$ -Wert	Cohens $d$
Voronoi	21.49	25.32	22.51	29.75	627	151	778	0.00	8.19	1.08
5	21.66	25.14	22.37	28.11	589	132	721	0.00	7.35	0.98
10	21.96	25.25	22.97	28.64	1088	322	1410	0.00	10.11	0.85
15	22.35	25.67	23.42	29.17	1542	564	2106	0.00	13.11	0.82
20	22.61	26.20	23.74	29.70	1948	858	2806	0.00	16.23	0.82
25	22.78	26.79	23.94	29.93	2319	1183	3502	0.00	18.76	0.82
30	22.98	27.28	24.23	30.37	2685	1525	4210	0.00	21.31	0.82
50	23.61	29.14	25.07	31.87	3894	3230	7124	0.00	31.37	0.84

<sup>†</sup> Gruppe der Ortspaare, die jeweils im selben Konfessionsgebiet liegen

<sup>††</sup> Gruppe der Ortspaare, die jeweils *nicht* im selben Konfessionsgebiet liegen

<sup>\*</sup>  $\alpha = 0.01$  (asymptotischer Mann-Whitney  $U$  Test)

<sup>6</sup>Selbiges Problem tritt auch bei den Konfessionsgrenzen auf, weshalb die Ortschaften dort in zusammenhängende Konfessionsgebiete eingeteilt wurden (vgl. Abschnitt 5.2.2)

Tabelle 6.3: Arbeitsmarktregionen

Nb-Skala	Median <sup>†</sup>	Median <sup>††</sup>	$\mu^{\dagger}$	$\mu^{\dagger\dagger}$	$n^{\dagger}$	$n^{\dagger\dagger}$	$n_{total}$	$p$ -Wert <sup>*</sup>	$z$ -Wert	Cohens $d$
Voronoi	21.64	22.99	22.57	27.38	847	198	1045	0.00	4.57	0.73
5	21.83	22.84	22.58	26.24	806	154	960	0.00	3.16	0.62
10	22.35	23.32	23.23	26.65	1520	394	1914	0.00	3.98	0.52
15	22.76	23.74	23.74	27.20	2154	697	2851	0.00	5.79	0.50
20	23.08	24.19	24.04	27.73	2737	1056	3793	0.00	7.77	0.52
25	23.27	24.54	24.24	27.99	3282	1460	4742	0.00	9.72	0.52
30	23.43	24.99	24.46	28.35	3763	1939	5702	0.00	11.77	0.52
50	24.05	26.77	25.04	29.70	5215	4349	9564	0.00	21.91	0.59

<sup>†</sup> Gruppe der Ortspaare, die jeweils in selber Arbeitsmarktregion liegen

<sup>††</sup> Gruppe der Ortspaare, die jeweils *nicht* in selber Arbeitsmarktregion liegen

\*  $\alpha = 0.01$  (asymptotischer Mann-Whitney  $U$  Test)

### 6.3 Regressionsmodelle mit scharfen und unscharfen Grenzen

Abbildung 6.8 und die Tabellen 6.4 und 6.5 zeigen die  $R_{korr}^2$ -Werte der Regressionsmodelle mit und ohne Einbezug der scharfen Grenzen für alle acht Nb-Skalen. Entgegen den Erwartungen erklären nicht die Regressionsmodelle mit den Einflussbereichsmassen GI-ED (max.  $R_{korr}^2 = 0.4035$  [30-Nb]) und GI-LCP (max.  $R_{korr}^2 = 0.4280$  [30-Nb]) die meiste Varianz der syntaktischen Distanzen, sondern die mit dem Erreichbarkeitsmass LCP (max.  $R_{korr}^2 = 0.4764$  [50-Nb]). Dies gilt für alle Nb-Skalen, wobei die  $R_{korr}^2$ -Werte beim Massstab der 5-Nb-Skala im Vergleich zur Voronoi-Nb-Skala zuerst kleiner werden, um danach bei den grösseren Skalen wieder stetig anzusteigen. Die Miteinbezugnahme der Bevölkerungsanzahl im GI-LCP führt im Vergleich mit dem LCP-Distanzmass bei allen Regressionsmodellen zu keiner grösseren erklärten Varianz. Bei den Regressionsmodellen mit Einbezug der scharfen Grenzen führt hingegen die Verwendung der GI-ED im Vergleich zum ED-Distanzmass bis zum Massstab der 50-Nb-Skala erwartungsgemäss zu einer grösseren erklärten Varianz (wenn auch nur minimal). Bei den Regressionsmodellen ohne Einbezug der scharfen Grenzen ist dies nur bei der Voronoi-Nb- und der 5-Nb-Skala der Fall. Bei den grösseren Skalen kehrt der Sachverhalt und der Miteinbezug der Bevölkerungsanzahl mittels GI-ED führt im Vergleich zum ED-Distanzmass zu einer kleineren erklärten Varianz. Von allen 64 Regressionsmodellen erklärt dasjenige auf der Skala der 50-Nb mit Einbezug der scharfen Grenzen *und* dem LCP als Distanzmass die meiste Varianz, dicht gefolgt von den Regressionsmodellen der 30-Nb-, Voronoi-Nb -und 25-Nb-Skalen (vgl. Tabelle 6.4). Bei den Regressionsmodellen ohne Einbezug der scharfen Grenzen ergibt sich fast dasselbe Bild. Das Regressionsmodell auf Skalengrösse der 50-Nb weist dicht gefolgt von den 30-Nb, 25-Nb, Voronoi-Nb und 20-Nb-Skalen den höchsten  $R_{korr}^2$ -Wert auf (vgl. Tabelle 6.5). Dass die Miteinbezugnahme der scharfen Grenzen die Residuenquadratsumme der Regressionsmodelle statistisch signifikant verkleinern, zeigen die paarweisen ANOVA-Tests der Regressionsmodelle mit demselben Distanzmass innerhalb derselben Nb-Skala. Für alle Regressionsmodelle mit Einbezug der scharfen Grenzen gilt  $p < 0.00$  bei einem Signifikanzniveau  $\alpha = 0.01$ .

**Interpretation:** Es ist nicht ganz klar, weshalb die  $R_{korr}^2$ -Werte von der Voronoi-Nb auf die 5-Nb-Skala kleiner werden und danach wieder ansteigen. Eine mögliche Erklärung könnte das MAUP sein (vgl. Abschnitt 5.3.1). Dies wäre nicht weiter verwunderlich, da die syntaktische Distanz inhärent durch den geographischen Abstand der Ortschaften ein räumliches Mass beinhaltet und somit auch durch das MAUP unterschiedliche Resultate hervorbringen kann. Gemäss Fotheringham und Wong (1991) ist das MAUP bei linearen, multiplen Regressionsmodellen bezüglich Effekten und Intensitäten unvorhersehbar. Wie in Abschnitt 5.3.1 bereits beschrieben, ist es deshalb auch eminent ein allfälliges MAUP mittels unterschiedlichen Analysenmassstäben zu berücksichtigen. Diesem Sachverhalt wird in dieser Arbeit mit den Analysen auf acht unterschiedlichen Nb-Skalen Rechnung getragen. Das räumliche Muster erst ab einem bestimmten Untersuchungsmassstab auftreten können, zeigen auch die Korrelationskarten, die in den Abschnitten 6.1.3 und 6.1.4 vorgestellt werden. Ebenso unklar ist, weshalb der Miteinbe-

zug der Bevölkerungsanzahl der Ortschaften die Varianz der syntaktischen nicht besser erklären als die Erreichbarkeitsmasse. Bei den Dominanzkarten (vgl. Abschnitt 6.1.4) zeigt sich, dass die GIs ab einer Nb-Skala von 10 und mehr Nachbarn anzahlmässig dominieren. Folglich wäre es zu erwarten, dass die Verwendung der GIs in den Regressionsmodellen die erklärte Varianz der syntaktischen Distanz erhöhen. Eine mögliche Erklärung für diesen nicht vorhandenen Umstand könnte sein, dass die Regressionsmodelle die Realität nicht korrekt abbilden und deshalb der Einfluss der GIs unterschätzt wird.

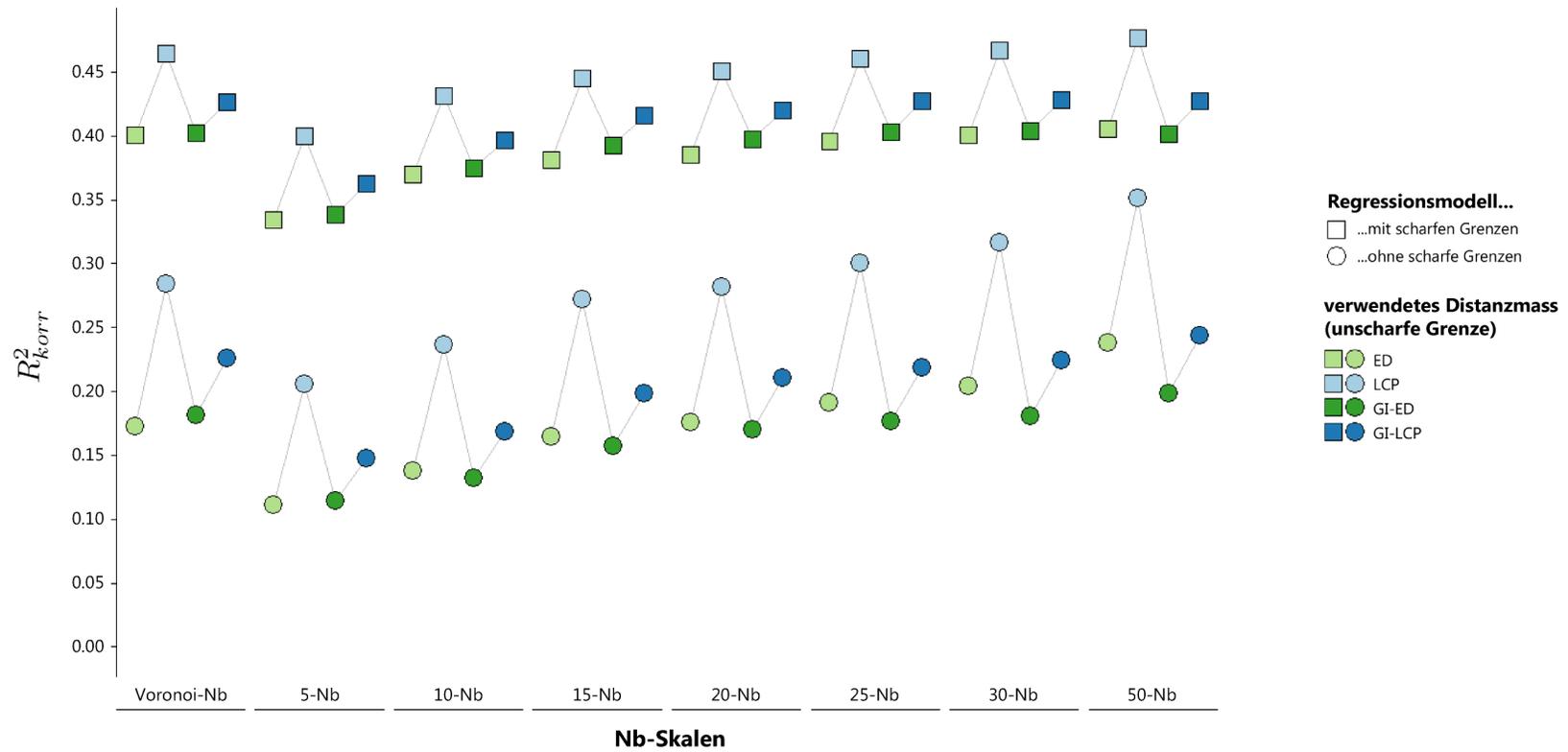


Abbildung 6.8:  $R^2_{korr}$ -Werte der Modelle mit und ohne Einbezug der scharfen Grenzen.

Tabelle 6.4: Regressionsmodelle *mit* Einbezug der scharfen Grenzen absteigend sortiert nach den grössten  $R_{korr}^2$ -Werten

Rang	$R_{korr}^2$	Nb-Skala	verwendetes Distanzmass
1.	0.4764	50-Nb	LCP
2.	0.4670	30-Nb	LCP
3.	0.4646	Voronoi-Nb	LCP
4.	0.4605	25-Nb	LCP
5.	0.4504	20-Nb	LCP
6.	0.4454	15-Nb	LCP
7.	0.4310	10-Nb	LCP
8.	0.4280	30-Nb	GI-LCP
9.	0.4271	50-Nb	GI-LCP
10.	0.4270	25-Nb	GI-LCP
11.	0.4267	Voronoi-Nb	GI-LCP
12.	0.4204	20-Nb	GI-LCP
13.	0.4159	15-Nb	GI-LCP
14.	0.4054	50-Nb	ED
15.	0.4035	30-Nb	GI-ED
16.	0.4033	25-Nb	GI-ED
17.	0.4022	Voronoi-Nb	GI-ED
18.	0.4010	50-Nb	GI-ED
19.	0.4010	30-Nb	ED
20.	0.4009	Voronoi-Nb	ED
21.	0.3999	5-Nb	LCP
22.	0.3970	20-Nb	GI-ED
23.	0.3968	10-Nb	GI-LCP
24.	0.3955	25-Nb	ED
25.	0.3922	15-Nb	GI-ED
26.	0.3855	20-Nb	ED
27.	0.3816	15-Nb	ED
28.	0.3747	10-Nb	GI-ED
29.	0.3702	10-Nb	ED
30.	0.3626	5-Nb	GI-LCP
31.	0.3380	5-Nb	GI-ED
32.	0.3343	5-Nb	ED

Tabelle 6.5: Regressionsmodelle *ohne* Einbezug der scharfen Grenzen absteigend sortiert nach den grössten  $R_{korr}^2$  -Werten

Rang	$R_{korr}^2$	Nb-Skala	verwendetes Distanzmass
1.	0.3510	50-Nb	LCP
2.	0.3166	30-Nb	LCP
3.	0.3007	25-Nb	LCP
4.	0.2845	Voronoi-Nb	LCP
5.	0.2819	20-Nb	LCP
6.	0.2718	15-Nb	LCP
7.	0.2437	50-Nb	GI-LCP
8.	0.2379	50-Nb	ED
9.	0.2361	10-Nb	LCP
10.	0.2258	Voronoi-Nb	GI-LCP
11.	0.2240	30-Nb	GI-LCP
12.	0.2187	25-Nb	GI-LCP
13.	0.2105	20-Nb	GI-LCP
14.	0.2053	5-Nb	LCP
15.	0.2042	30-Nb	ED
16.	0.1985	15-Nb	GI-LCP
17.	0.1985	50-Nb	GI-ED
18.	0.1914	25-Nb	ED
19.	0.1812	Voronoi-Nb	GI-ED
20.	0.1803	30-Nb	GI-ED
21.	0.1765	25-Nb	GI-ED
22.	0.1755	20-Nb	ED
23.	0.1724	Voronoi-Nb	ED
24.	0.1699	20-Nb	GI-ED
25.	0.1683	10-Nb	GI-LCP
26.	0.1647	15-Nb	ED
27.	0.1574	15-Nb	GI-ED
28.	0.1476	5-Nb	GI-LCP
29.	0.1378	10-Nb	ED
30.	0.1321	10-Nb	GI-ED
31.	0.1146	5-Nb	GI-ED
32.	0.1111	5-Nb	ED

# Kapitel 7

## Diskussion

Diese Arbeit untersuchte den Einfluss von verschiedenen scharfen und unscharfen Grenzen auf die syntaktischen Dialektunterschiede der deutschen Schweiz. Als scharfe Grenzen dienten Kantons-, Konfessions- und Arbeitsmarktregionsgrenzen; als unscharfe Grenzen wurden verschiedene Distanzmasse verwendet. Ein Ziel war, mehr über den strukturellen Zusammenhang zwischen vermuteten Kontaktbarrieren und Sprachunterschieden zu erfahren. Ein weiteres Ziel bestand darin, mit den (un-) scharfen Grenzen mehrere lineare, multiple Regressionsmodelle zu erstellen um damit die Dialektvariation innerhalb der Deutschschweiz möglichst gut erklären zu können. Die Regressionsmodelle sollen die Grundlage bilden für eine zukünftige Interpolationsmethode von Sprachdaten. Nachfolgend werden die zu Beginn der Arbeit vorgestellten Forschungsfragen beantwortet.

### 7.1 FF 1: Haben politische, religiöse und wirtschaftliche Grenzen einen Einfluss auf sprachliche Unterschiede?

In der linguistischen Forschung ist noch immer nicht klar, welche Faktoren und Mechanismen die Sprachen verändern (vgl. Abschnitt 2.1). Alle Resultate der vorliegenden Arbeit können jedoch die Theorie stützen, dass politische, religiöse und wirtschaftliche Grenzen als Kontakt- und Sprachbarrieren wirken (vgl. Haag, 1898; Kürschner und Gooskens, 2011; Pickl, 2013).

Die Zerteilung der Regressionsmodelle in solche *mit* und solche *ohne* Einbezug der scharfen Grenzen zeigte zusätzlich, dass die Hinzunahme der drei scharfen Grenzen die syntaktischen Unterschiede der Ortschaften signifikant besser erklärt.

Bei den Einzeluntersuchungen der drei scharfen Grenzarten (Kantone, Konfessionen und Arbeitsmarktregionen) zeigten die Mittelwerttests für alle drei Grenzarten statistisch signifikante Differenzen zwischen den syntaktischen Unterschieden der jeweiligen zwei Ortspaargruppen auf. Unterschieden wurden dabei die Ortspaargruppen «*ohne* scharfe Grenzen zwischen den Ortspaa- ren» von den Ortspaargruppen «*mit* scharfen Grenzen zwischen den Ortspaa- ren». Im Vergleich zu den Gruppen «*ohne* scharfe Grenzen zwischen den Ortspaa- ren» sind die Mittelwerte der syn-

taktischen Distanzen der Gruppen «mit scharfen Grenzen zwischen den Ortspaaren» signifikant höher.

Die berechneten Effektstärken des Cohens  $d$  zeigten zusätzlich auf, dass der Einfluss der scharfen Grenzen auch relevant ist. Es gilt aber zu beachten, dass die Einteilung von Cohen in die Effektstärken *schwach* ( $d = 0.2$ ), *mittel* ( $d = 0.5$ ) und *stark* ( $d = 0.8$ ) nicht empirisch begründet ist. Für eine abschliessende Beurteilung der Effektstärken von den scharfen Grenzen fehlen Vergleichswerte von Studien mit ähnlichen Daten. Es wäre deshalb wünschenswert, wenn zukünftige Studien die Effektstärken von verschiedenen Faktoren ebenfalls mit Cohens  $d$  messen würden. Je mehr Vergleichswerte vorhanden sind, desto eher lässt sich die wirkliche Effektstärke eines Faktors beurteilen.

Bei den Kantonsgrenzen stehen die Resultate im Einklang mit ihrer geschichtlichen Bedeutung in der föderalistisch aufgebauten Schweiz. Kantonsgrenzen beeinflussen bis heute das gesellschaftliche Leben und den interkantonalen Austausch innerhalb der Schweiz. Da die politische Tendenz in letzter Zeit eher zu einem schrittweisen Abbau der Kantonsautonomie führte (z.B. durch die Angleichung des kantonal reglementierten Schulwesens), wäre es interessant zu sehen, ob die Kantonsgrenzen in Zukunft einen schwächeren Einfluss auf die syntaktischen Unterschiede zeigen würden.

Bei den Konfessionsgrenzen decken sich die berechneten Effektstärken mit der Aussage der Historikerin Caroline Schnyder (2013), dass die Reformation zu den prägendsten Ereignissen innerhalb der Schweiz gehöre. Die aus der Reformation folgenden Spannungen und Kriege bedeuteten eine Zerreihsprobe und führten noch bis vor ein paar Jahrzehnten dazu, dass es für undenkbar gehalten wurde über Konfessionen hinweg zu heiraten. Zwar besteht in der Schweiz noch heute keine offizielle Trennung zwischen Kirche und Staat, doch sind die konfessionellen Unterschiede im alltäglichen Zusammenleben kaum mehr zu spüren. In Anbetracht der heute geringen gesellschaftlichen Bedeutung von Konfessionsunterschieden bestätigt sich damit auch die Aussage, dass die Sprache ein Abbild der Vergangenheit sei.

Die Resultate über den Einfluss von Arbeitsmarktregionsgrenzen auf die syntaktischen Unterschiede bekräftigen wiederum die Theorie von Auer et al. (2011) und Britain (2012), dass Pendlerbewegungen zu einem vermehrt interregionalen Dialektaustausch führen. Da die Arbeitsmarktregionen auf Pendlereinzugsgebieten basieren (vgl. Unterabschnitt 5.2.3 bestätigen die Resultate auch die Forschungsergebnisse von Lameli (2013) und Falck et al. (2012), nach denen der wirtschaftliche Austausch und die regionale Arbeitsmigration mit Dialektregionen erklärt werden kann.

Gemäss Thomason (2001) ist die Grammatik die stärkste und stabilste Ebene einer Sprache und verändert sich bei Kontakt mit anders Sprechenden am langsamsten. So ist es doch überraschend, dass die Effektstärken der Arbeitsmarktregionen für alle acht Nb-Skalen ähnlich hoch ausfallen wie bei den Kantonsgrenzen. Dies in Anbetracht dessen, dass das Phänomen des «Pendelns» zur Arbeit ein neuzeitliches ist und erst seit dem Aufkommen des tertiären Wirtschaftssektors einen anhaltenden Boom erlebt (vgl. Abschnitt 5.2.3). Wegen den jährlich steigenden Pendlerzahlen ist deshalb anzunehmen, dass der Einfluss von Pendlern auf die Dialektunterschiede auch in Zukunft zunehmen wird.

Es ist anzunehmen, dass die Grenzverläufe der drei scharfen Grenzarten jeweils einen ähnlichen Verlauf aufweisen (vgl. Abb. 5.7, 5.8 und 5.9). Dies hat historische Ursachen. Beispielsweise unterscheiden sich die Konfessionszugehörigkeiten der Kantone u.a. aufgrund der langen föderalistischen Geschichte der Schweiz. Ob und wie stark die drei untersuchten scharfen Grenzen deckungsgleich sind, wurde in dieser Arbeit aber nicht untersucht. Somit sind definitive Aussagen über die Effektstärken der *einzelnen* Grenzarten noch nicht möglich. Wenn beispielsweise Kantons- und Konfessionsgebietsgrenzen grösstenteils deckungsgleich sind, so ist es fast unmöglich, den Einfluss von *einer* der beiden Grenzen auf die sprachlichen Unterschiede zu bestimmen. Interessant wären dazu Vergleiche von einzelnen Grenzlinien-Abschnitten, die nur einer der drei Grenzarten entsprechen. Mit einem zusätzlichen Vergleich von Grenzlinien-Abschnitten, bei denen zwei oder gar drei Grenzarten deckungsgleich sind, könnte auch die These von Haag (1898) und Pickl (2013) überprüft werden. Nach dieser These wirken Grenzen vor allem dort als Sprachbarriere, wo mehrere Grenzarten zusammentreffen (vgl. Abschnitt 2.1).

## 7.2 FF 2: Haben geographische Distanzen einen Einfluss auf sprachliche Unterschiede?

In der bisherigen Forschung gab es diskrepante Ergebnisse bezüglich dem Einfluss von geographischen Distanzen auf sprachliche Unterschiede. So stellte Szmrecsanyi (2012) das von Nerbonne und Kleiweg (2007) für die Linguistik adaptierte erste Gesetz der Geographie in Frage (vgl. Abschnitt 2.1). Seine Resultate für britische Dialekte zeigten auf, dass die euklidische Distanz nur 4%, die LCPs mittels Auto nur 8% und Trudgills Gravitations-Index ca. 24% der Dialektunterschiede erklärt (Szmrecsanyi, 2012). Zumindest für die syntaktischen Unterschiede der Schweiz kann seine Behauptung, dass geographische Nähe ein schlechter Prädiktor für sprachliche Ähnlichkeit sei, widerlegt werden. Für alle vier in dieser Arbeit untersuchten Distanzen (= unscharfe Grenzen) konnte aufgezeigt werden, dass ein klarer Zusammenhang zwischen geographischer Distanz und sprachlichen Unterschieden besteht. Dies bestätigt auch die Resultate von Jeszenszky und Weibel (2014), die ihre Studie mit denselben syntaktischen Daten, welche auch in der vorliegenden Arbeit verwendet wurden, durchgeführt haben. Zwar sind Jeszenszkys und Weibels (2014) Resultate weniger klar zu deuten, da sie einen lokalen und einen globalen Massstab untersucht haben, doch zeigen auch ihre Ergebnisse zwischen den euklidischen Distanzen sowie den historischen resp. neuzeitlichen Autoreisezeiten und den syntaktischen Distanzen der Ortschaften starke, positive Korrelationen. Da der Autor dieser Arbeit keinen Zugriff auf die von Jeszenszky und Weibel (2014) berechneten Autoreisezeiten hatte, konnten deren Korrelationen nicht mit den in dieser Arbeit verwendeten Wanderzeit-Korrelationen verglichen werden. Somit kann die Frage, welche der beiden Fortbewegungsmittel (Auto oder zu Fuss) in einem stärkeren Zusammenhang zu den syntaktischen Unterschieden der Deutschschweiz steht, nicht beantwortet werden.

Im Vergleich mit Szmrecsanyis (2012) Resultaten erklärt von allen untersuchten geographischen Distanzen auch in dieser Arbeit die euklidische Distanz am wenigsten Dialektvariation,

doch liegen die erklärten Varianzen der deutschschweizer Dialekte mit maximal knapp 24% (50-Nb-Skala) um das 6-fache resp. mit minimal 11% (5-Nb-Skala) fast um das 3-fache höher als die 4% erklärter Varianz britischer Dialekte bei Szmrecsanys Arbeit (vgl. 8. Rang und 32. Rang bei Tabelle 6.5).

Genau gegenteilig verhalten sich die Resultate im Bezug auf die Gravitations-Indizes. Im Gegensatz zu Szmrecsanys Resultaten, bei denen der GI viel mehr Dialektvarianz als die euklidische Distanz erklärte, führten in dieser Arbeit die Gravitations-Indizes mit euklidischer Distanz (GI-ED) bei den sechs grössten untersuchten Nb-Skalen sogar zu einer geringeren erklärten Varianz als mit den euklidischen Distanzmassen (ED) (vgl. die  $R_{kor}^2$ -Werte der Regressionsmodelle *ohne* Miteinbezug der scharfen Grenzen bei Abb. 6.8 und die Tabelle 6.5). Dass die GIs ein schlechter Prädiktor für die sprachlichen Unterschiede sind, zeigt sich noch stärker beim Vergleich der tieferen  $R_{kor}^2$ -Werte der Regressionsmodelle mit GI-LCP als Distanzmass gegenüber den höheren Werten der Modelle mit LCP als Distanzmass. Die Hinzunahme eines demographischen Faktors, welcher mit der Bevölkerungsanzahl in den Gravitations-Indizes eingebunden ist, führt somit nicht wie erwartet zu einer Zunahme, sondern zu einer Abnahme der erklärten Varianz der syntaktischen Unterschiede.

Eine mögliche Erklärung liegt in der Natur der verwendeten Daten. Die 383 Erhebungsorte sind eine Stichprobe und stellen damit nur knapp einen Viertel aller Deutschschweizer Ortschaften dar. Dies kann dazu führen, dass der Einfluss von bevölkerungsreichen, im Datensatz nicht vorhandenen Dritt-Ortschaften, nicht berücksichtigt wird. Wenn solche im SADS nicht aufgenommene Dritt-Ortschaften zwischen oder in der Nähe der betrachteten SADS-Ortspaaren liegen, führt dies zu einer Verfälschung der GI-Ergebnisse. Neben im Datensatz *nicht vorhandenen* Dritt-Orten kann derselbe Effekt aber auch mit im Datensatz vorhandenen Dritt-Orten auftreten. Dies deshalb, weil die Beziehungen der Orte in den verwendeten Korrelationsanalysen sowie in den Regressionsmodellen immer nur paarweise betrachtet werden. Dritt-Orte, welche einen grossen Einfluss auf die betrachteten paarweisen Ortschaften haben, werden dabei jedoch ignoriert. Somit ist dieser Effekt nicht nur ein Datenproblem, sondern auch ein methodisches.

Eine andere mögliche Erklärung ist, dass die Gravitations-Indizes und die Bevölkerungsanzahl der Orte für die Schweiz schlicht einen weniger grossen Einfluss auf die syntaktischen Unterschiede haben als angenommen. Vor allem die grossen Städte der Schweiz haben im Vergleich zu den von Szmrecsanys verwendeten britischen Städten eine viel geringere Einwohnerzahl. Die kleineren Unterschiede der Bevölkerungszahlen der Schweizer Städte, die gleichmässige Besiedelung der Schweiz und die ebenfalls geringeren geographischen Distanzen im Vergleich mit Grossbritannien könnten dazu führen, dass die Gravitations-Indizes einen schwächeren Einfluss auf die syntaktischen Unterschiede haben.

Eine Erklärung für die in dieser Arbeit viel grössere erklärte Dialektvarianz (max. 35% [50-Nb-Skala] bis min. 20% [5-Nb-Skala]) der LCPs im Vergleich zu den nur 8% erklärter Varianz britischer Dialekte kann in den topographischen Unterschieden der beiden Länder liegen. So ist die Schweiz mit ihrer hügeligen Topographie stärker von Bergen geprägt als das flachere Grossbritannien. Die Folge ist, dass die Topographie als Kontakt- und somit Sprachbarriere in der Schweiz einen stärkeren Einfluss hat als in Grossbritannien.

Eine andere Erklärung für die Unterschiede der durch die LCPs erklärten Dialektvarianzen könnte in der Methode liegen, mit der Szmrecsanyi seine LCP-Distanzen berechnete. Szmrecsanyi verwendete als LCPs die schnellsten Reisezeiten per Auto, die er mit Google Maps berechnete. Im Gegensatz dazu wurden in der vorliegenden Arbeit die LCPs mit Wanderzeiten berechnet. Da die heutigen Dialekte ein Abbild der Vergangenheit sind, kann die berechnete Frage gestellt werden, wie viel Sinn ein Vergleich der modernen Autoreisezeiten mit den Dialektunterschieden ergibt. Zwar zeigte sich auch in dieser Arbeit mit dem Testen über den Einfluss von Arbeitsmarktregionen bereits ein starker Effekt der modernen Reisemöglichkeiten per Auto und Zug, doch basieren die in dieser Arbeit verwendeten Arbeitsmarktregionen auf Gebieten, innerhalb derer sich ein Pendleraustausch konzentriert (vgl. Abschnitt 5.2.3). Zudem ist es noch ungeklärt, wie der starke Effekt der Arbeitsmarktregionsgrenzen und somit der modernen Reisemitteln im Zusammenhang mit den Kantons- und Konfessionsgrenzen steht.

Die wohl wahrscheinlichsten zwei Erklärungen für Szmrecsanyis geringe erklärte Varianz mit geographischen Distanzen sind folgende: Erstens verwendet er auf County-Grösse aggregierte Daten. Dies steht in grossem Gegensatz zu den hier verwendeten, räumlich hoch aufgelösten Daten. Dass stark aggregierte Daten in der räumlichen Datenanalyse ein Problem darstellen können, wurde im Abschnitt 5.3.1 über den Einfluss des MAUPs beschrieben. Die Auswirkungen des MAUPs zeigen sich auch deutlich in der vorliegenden Arbeit. Bei allen durchgeführten Analysen unterscheiden sich die Resultate je nach betrachteter Nb-Skala. Auch räumliche Muster lassen sich erst ab einem gewissen Massstab auf den Karten erkennen (vgl. Abb. 6.3). Wie wichtig unterschiedliche Untersuchungsmassstäbe sind, zeigt auch die in Abschnitt 2.1 vorgestellte Arbeit von Jeszenszky und Weibel (2014). Jeszenszky und Weibel konnten aufzeigen, dass die Korrelationen zwischen syntaktischen Unterschieden und geographischer Distanz stark von einer lokalen oder einer globalen Betrachtungs-Skala abhängig sind.

Zweitens, und wie von Szmrecsanyi (2012) ebenfalls als Grund für seine abweichenden Resultate bereits erwähnt, sind seine verwendeten Korpusdatensätze. Dies steht im Gegensatz zu den allermeisten geolinguistischen Studien, welche üblicherweise Sprachatlanten verwenden (der in dieser Arbeit verwendete SADS ist ebenfalls ein Sprachatlas). Sprachatlanten unterscheiden sich von Korpusdatensätzen darin, dass erstere üblicherweise nur sprachliche Phänomene enthalten, die eine räumliche Variation aufweisen. Bei Korpusdaten ist dieser Schwerpunkt nicht gegeben. Sie sind reine Sammlungen von schriftlichen Texten einer Sprache. Ob man Sprachatlanten oder Korpusdatensätze verwendet, ist dabei abhängig von der wissenschaftlichen Fragestellung. Es ist jedoch fraglich, ob eine räumliche Analyse von Daten, die nachweislich keine spezifischen, räumlichen Zusammenhänge besitzen, sinnvoll ist.

Bei den  $R_{korr}^2$ -Werten der Regressionsmodelle zeigen sich je nach Nb-Skala unterschiedliche Resultate. Besonders interessant ist dabei die Tatsache, dass die  $R_{korr}^2$ -Werte der Voronoi-Nb-Skala denen der 50-Nb-Skala ähneln. So sinken die  $R_{korr}^2$ -Messgrössen vom Voronoi- auf die 5-Nb-Skala jeweils stark ab, um danach wieder mit grösser werdender Nb-Skala anzusteigen und schliesslich bei der 50-Nb-Skala ähnlich hohe Werte zu erreichen wie bei der Voronoi-Nb-Skala. Dieselbe Tendenz zeigt sich auch bei den berechneten Effektstärken der scharfen Grenzen (vgl. die Cohens  $d$  Werte der unterschiedlichen Nb-Skalen in den Tabellen 6.1, 6.2 und 6.3) und

bei den Medianen der GI-Indizes der Korrelationsboxplots (vgl. Abb. 6.2). Neben dem MAUP als mögliche Erklärung für die hohen Messgrößen auf der Voronoi-Nb-Skala sei nochmals auf die Wellentheorie als mögliche Erklärung für die Sprachveränderung verwiesen (vgl. Abschnitt 5.3.1.1). Gemäss diesem linguistischen Modell ist die relevanteste Form von Kontakt diejenige zwischen zwei benachbarten Orten. Die Voronoi-Nb-Skala ist ein Abbild der direkten Nachbarschaften und so sollten diese gemäss der Wellentheorie auch die stärksten Effekte aufweisen. Die Wellentheorie wird auch von der Tatsache gestützt, dass die  $R_{korr}^2$ -Werte der Regressionsmodelle von der Voronoi-Nb-Skala hin zur 5-Nb- und 10-Nb-Skala jeweils stark abnehmen (vgl. Abb. 6.8).

### 7.3 FF 3: Können syntaktische Unterschiede in einem Modell mit mehreren geographischen und demographischen Faktoren erklärt werden?

Die vorliegende Arbeit zeigt auf, dass mit der Methode der linearen, multiplen Regressionsanalyse eine Möglichkeit besteht, mehrere unabhängige Faktoren zusammenzufassen und dass damit die syntaktischen Unterschiede gut erklärt werden können. Lineare, multiple Regressionsmodelle haben mehrere Stärken. So können beispielsweise ANOVA-Tests aufzeigen, ob der Miteinbezug von zusätzlichen erklärenden Variablen die erklärte Dialektvarianz signifikant erhöht oder nicht. Die hier nicht benutzten Entscheidungskriterien wie das AIC oder BIC<sup>1</sup> können dazu dienen, die besten Modelle bezüglich ihrer Einfachheit auszusuchen. Die Güte des Modells lässt sich mit dem in dieser Arbeit verwendeten Bestimmtheitsmass  $R_{korr}^2$  bestimmen. Auch der Einfluss der einzelnen erklärenden Variablen auf die syntaktische Distanz könnte bei linearen, multiplen Regressionsmodellen untersucht werden (vgl. Abschnitt 5.3.4). Obwohl in dieser Arbeit nicht verwendet, wäre es möglich (ähnlich wie bei den Mittelwerttests) die Bestimmtheitsmasse  $R_{korr}^2$  der Regressionsmodelle auf ihre Effektstärke mit Cohens  $d$  überprüfen. Dies würde zeigen, ob die erklärte Dialektvarianz auch genügend hoch ist, um als bedeutend eingestuft zu werden (Universität Zürich, 2016a).

---

<sup>1</sup> Das Akaike-Informationskriterium (AIC) von Hirotugu Akaike (1973) oder auch das Bayessche Informationskriterium (BIC) von Gideon Schwarz (1978) sind beides Kriteriumsmaße für die Auswahl von Regressionsmodellen. Sie folgen dem Sparsamkeitsprinzip von *Okhams Rasiermesser*, nach dem bei mehreren möglichen Erklärungen dasjenige ausgewählt werden soll, welches am einfachsten ist (Domingos, 1999).

## 7.4 Die abschliessende Relevanz der Resultate

Die vorliegende Arbeit zeigte mit der Methode der linearen, multiplen Regressionsmodelle eine Möglichkeit auf, wie strukturelle Raumkomponenten in Form von unscharfen und scharfen Grenzen in ein einziges Modell für die Erklärung der Dialektunterschiede zusammengefasst werden können. Die Relevanz der in dieser Arbeit vorgestellten Resultate zeigt sich in folgenden zwei Punkten.

Erstens bieten sich die hier vorgestellten Modelle für die Wirtschaftswissenschaften an. Gemäss Falck et al. (2012) und Lameli (2013) besteht zwischen kulturell ähnlichen Regionen mehr wirtschaftlicher Austausch als zwischen Regionen, die sich kulturell fremder sind. Falck et al. (2016) konnten zudem aufzeigen, dass regionale kulturelle Unterschiede ein Hemmnis für die interregionale Arbeitsmigration darstellen und dass Arbeitsmigranten erst ab einer, von den kulturellen Unterschieden abhängigen, Lohnerhöhung dazu bereit sind, ihren Arbeitsort zu wechseln. Dies gilt gemäss Falck et al. vor allem für besser verdienende, männliche Arbeitnehmer und über geringe Distanzunterschiede zum neuen Arbeitsort. Da Falck et al. (2012), Lameli (2013) und Falck et al. 2016 kulturell ähnliche Gebiete mit historischen Dialektdaten und den daraus berechneten Distanzen approximieren, ist es augenscheinlich, dass das in dieser Arbeit erstellte multiple Regressionsmodell im Kontext quantitativer Wirtschaftsgeographie angewendet werden könnte.

Zweitens stellen die in dieser Arbeit erstellten Regressionsmodelle eine erstmalig durchgeführte, empirische-quantitative Kombination von verschiedenen, einzelnen linguistischen Theorien dar, welche in den Sprachwissenschaften diskutiert werden. Zwar beschrieb Gooskens bereits 2004 den Nutzen einer hier durchgeführten, quantitativen Kombination von mehreren geographischen Faktoren in einem Modell zur Erklärung der Dialektunterschiede (vgl. Abschnitt 2.1), doch ist ein solch zusammenfassendes Modell nach dem Wissen des Autors der vorliegenden Arbeit bis jetzt noch nie erstellt worden. Da die Sprache von diversen Faktoren beeinflusst wird, kann gemäss Gooskens ein Modell, das mehrere erklärende Variablen gleichzeitig miteinbezieht, zu einem verbessertem Verständnis der Sprachunterschiede führen. Des Weiteren können die hier erstellten Regressionsmodelle für Interpolationen verwendet werden um nicht vorhandene Datenpunkte zu schätzen. Die damit erstellten Karten können zum einen helfen, weitere räumliche, sprachbeeinflussende Strukturen zu entdecken, zum anderen können mit den analysierten Schätzfehlern der Karten zukünftige Sprachstudien besser geplant und die Erhebungsorte der Daten adäquater ausgewählt werden.

Ob die hier gewählten linearen, multiplen Regressionsmodelle am geeignetsten sind um mehrere Faktoren für die Modellierung der syntaktischen Distanz zu berücksichtigen oder ob sich andere Modelle, wie z.B. nicht-lineare Regressionsmodelle, besser eignen, sollte in weiterführender Forschung untersucht werden. Über die methodischen Schwierigkeiten, mehrere erklärende Faktoren in ein einziges Modell «hineinzupacken», geben die im nachfolgenden Kapitel beschriebenen Limitierungen Aufschluss.



# Kapitel 8

## Limitierungen

In diesem Kapitel werden die Limitierungen dieser Arbeit aufgeführt. Sie sollen aufzeigen, dass die hier vorgestellten Methoden und Resultate gewissen Limitierungen unterworfen und die Resultate im Kontext der nachfolgenden Punkte zu betrachten sind. Dies im Hinblick auf weiterführende Studien und wissenschaftliche Arbeiten, welche die hier vorgestellte Methodik des linearen, multiplen Regressionsmodells für die Modellierung von sprachlichen Unterschieden weiterverfolgen.

### 8.1 Limitierungen in Bezug auf die Modellannahmen

Die Aggregation von Daten stellt ein Problem dar, das in dieser Arbeit mehrfach beschrieben wurde. So beeinflusst erstens das MAUP die hier vorgestellten Resultate, da zum einen die verwendeten Sprachdaten bereits auf Ortsebene aggregiert sind, zum anderen für die Untersuchungen der Nb-Skalen stets mehrere Orte zusammengefasst wurden. Zwar wurde mittels unterschiedlicher Nb-Skalen versucht das MAUP zu berücksichtigen, doch bleibt diese Methode wie bereits in Abschnitt 5.3.1 beschrieben, die einzige Möglichkeit dem MAUP Rechnung zu tragen. Des Weiteren hat die Aggregation der Orte in Nb-Skalen zur Folge, dass dieselben Ortspaare mehrmals in die Berechnungen miteinbezogen werden. Damit wird die Voraussetzung der Unabhängigkeit der Daten verletzt und die Resultate werden verzerrt.

Eher problematisch ist die generelle Annahme der Linearität zwischen geographischen Faktoren und syntaktischer Distanz. Sprachen und ihre Unterschiede haben sich über tausende von Jahren entwickelt und wurden von unzähligen Faktoren beeinflusst. Aus diesem Grund ergibt sich zwischen den sprachlichen Unterschieden und den erklärenden Variablen auch kein perfekt-linearer Zusammenhang. Die in dieser Arbeit verwendeten linearen Regressionsmodelle setzen einen linearen Zusammenhang der untersuchten Variablen voraus. Zusätzlich soll die Komplexität eines Regressionsmodells auch möglichst gering sein (vgl. Abschnitt 5.3.4). Gemäss Backhaus et al. (2011) ist es deshalb oft ausreichend, nicht-lineare Zusammenhänge durch lineare Modelle zu approximieren. Zudem lassen komplexere, nicht-lineare Regressionsmodelle keine Berechnung des Bestimmtheitsmass ( $R^2$ ) zu, welches das Vergleichen verschiedener Modelle verunmöglicht.

Aus diesen Gründen werden in dieser Arbeit lineare (und keine komplexeren, nicht-lineare) Regressionsmodelle verwendet.

Bei linearen, multiplen Regressionsmodellen stellt sich bei der Verwendung von mehreren erklärenden Variablen auch immer die Frage nach der Multikollinearität. Die scharfen Grenzen der Kantone, Konfessionsgebiete und der Arbeitsmarktregionen haben naturgegebene Korrelationen untereinander. Dies führt zu hohen Varianzinflationsfaktoren (VIF<sup>1</sup>). Multikollinearität ist vor allem dann ein Problem, wenn man den Einfluss von einzelnen erklärenden Faktoren untersuchen möchte. Auf eine solche Untersuchung wurde in dieser Arbeit jedoch verzichtet (vgl. Abschnitt 5.3.4). Wenn beispielsweise Kantons- und Konfessionsgebietsgrenzen grösstenteils deckungsgleich sind, so ist es fast unmöglich, den Einfluss von einer der beiden Grenzen auf die sprachlichen Unterschiede zu bestimmen. Interessant wären dazu Vergleiche von Grenzlinien-Abschnitten, die beispielsweise einer Kantons- *und* einer Konfessionsgrenze entsprechen, mit solchen Grenzabschnitten, deren Verläufe nur zu einer der beiden Grenzarten gehören. Damit könnte auch die These von Haag (1898) und Pickl (2013) überprüft werden, nach der Grenzen vor allem dort eine Sprachbarriere darstellen, wo mehrere Grenzarten zusammentreffen (vgl. Abschnitt 2.1).

Eine weitere Limitierung betrifft die Frage der Kausalität der Korrelationsresultate. Da sich die Sprachen und Dialekte über tausende von Jahren verändert haben und von unzähligen Faktoren beeinflusst wurden, kann es jedoch nicht das Ziel sein, kausale Zusammenhänge zwischen erklärenden Faktoren und sprachlichen Unterschieden zu entdecken.

## 8.2 Weitere Limitierungen

Eine ungünstige Tatsache liegt in der Grösse und dem Grenzverlauf der Deutschschweiz. Viele SADS-Orte grenzen ans Ausland oder an anderssprachige Landesteile. Dies führt bei den Analysen zu vielen Randeffekten. Randeffekte entstehen, wenn die Resultate von Randgebieten stark von ihrer peripheren Lage beeinflusst sind. So hat beispielsweise bei grösseren Nb-Skalen eine Ortschaft in der Mitte der Deutschschweiz von allen Seiten eine ähnliche Anzahl Nachbarschaftsorte. Dies im Gegensatz zu einem peripher liegendem Untersuchungsort. Seine Lage hat zur Folge, dass für ihn nur von einer Himmelsrichtung Nachbarschaftsorte für die Berechnung hinzugenommen werden können. Dies kann zu einer Verzerrung der Resultate führen.

Eine zusätzliche Limitierung ist die nicht vollständig erfasste Beziehung zwischen den untersuchten Ortspaaren. Wenn nur die Masse zwischen Ortspaaren verglichen werden, ohne den lokalen Kontext zu berücksichtigen – z.B. den Einfluss von nahen, im Datensatz nicht vorhandenen Dritt-Orten –, kann dies zu widersprüchlichen Resultaten führen. Dies ist auch eine mögliche Erklärung für den schwachen Einfluss der Gravitations-Indizes, welche bereits in Abschnitt 7 diskutiert wurde.

---

<sup>1</sup>Der  $VIF_i = \frac{1}{1 - R_i^2}$  wird für jede der erklärenden Variablen der Regressionsmodelle berechnet. Er beginnt bei 1 und hat keine Obergrenze. Je höher der VIF, desto stärker ist der Hinweis auf Multikollinearität der erklärenden Variablen (Backhaus et al., 2011). Es gibt keinen definitiven Wert, ab wann Multikollinearität vorherrscht, aber mehrere unterschiedliche Faustregeln (VIF höher als 2, 5 oder 10).

Darüber hinaus ist das in dieser Arbeit verwendete Mass der syntaktischen Distanz, nur eines von vielen möglichen Massen um sprachliche Unterschiede zu messen. Es ist anzunehmen, dass die Verwendung eines anderen Masses auch zu anderen Resultaten führen würde. Zudem ist die hier verwendete syntaktische Distanz nicht statistisch berechnet, sondern nur die Summe der Unterschiede zwischen den Ortschaften. Aus diesem Grund tragen syntaktische Variablen mit vielen Varianten mehr zum sprachlichen Unterschied bei als Variablen mit wenigen Varianten (vgl. Abschnitt 4 über die Berechnung der syntaktischen Distanz). Des weiteren ist es wahrscheinlich, dass einige der syntaktischen Variablen voneinander abhängig sind. Dies bedeutet, dass die betroffenen Variablen dasselbe syntaktische Phänomen untersuchen. Die damit entstehenden Kollinearitäten führen zu einer Übervertretung einiger Variablen. Hier würde es Sinn ergeben ein Mass zu entwickeln, das die syntaktischen Unterschiede ausgeglichener repräsentiert und das statistisch berechnet wurde.

Überdies repräsentieren die hier untersuchten scharfen Grenzlinien nur Momentaufnahmen und haben sich in ihrem Verlauf – abgesehen von den neuen Arbeitsmarktregionen – mehrfach verändert. Die Dialektähnlichkeit ist hingegen das Produkt eines Prozesses, der über viel Zeit zustande gekommen ist. Eine perfekte Untersuchung über den Einfluss des Raumes auf die Sprache sollte beides, die Dynamik der Grenzen und die Dynamik der Sprache, als parallele Entwicklung untersuchen. Dafür gibt es aber keine Daten. Eine Möglichkeit, die sprachliche Entwicklung über die Zeit zu verfolgen, wäre eine Zusammenfassung mehrerer grosser Spracherhebungen wie z.B. dem *Sprachatlas der deutschen Schweiz* (SDS) (Hotzenköcherle, 1962), dem SADS oder auch dem *Deutschen Sprachatlas* (DSA) (Schmidt und Herrgen, 2001). Das Problem dabei ist, dass ihre unterschiedlichen Methoden zur Datenerhebung, ihre verschiedenen erhobenen Daten (z.B. syntaktische und phonologische Daten) sowie ihre voneinander abweichenden Erhebungsorte es faktisch unmöglich machen, ein kohärentes Mass der Dialektähnlichkeit zu erstellen.

Bei der Frage nach der Definition, was eine *scharfe* Grenze von einer *unscharfen* Grenze unterscheidet, zeigen sich ebenfalls Probleme. So sind die untersuchten Arbeitsmarktregionen nicht zwingend «scharfe» Grenzen, wobei ähnliches auch für die Konfessions- und Kantonsregionen gilt. Die Grenzarten sollten aber auch nicht als scharfe, undurchlässige Trennlinie, sondern als durchlässige Kontaktbarrieren verstanden werden. Kontaktbarrieren können selbstverständlich umgangen und überbrückt werden, doch stellen sie Hindernisse für den Menschen dar. Die Vereinfachung in «scharfe» Grenzen erfolgte zwecks klarerer Einteilung der Ortschaften in eine der beiden Gruppen.



## Kapitel 9

# Konklusion

Es besteht die linguistische Theorie, dass scharfe und unscharfe Grenzen als Kontakt- und somit als Sprachbarrieren wirken. Ihre effektive Bedeutung ist in der linguistischen Forschung jedoch noch umstritten. Im Fokus dieser Arbeit stand deshalb der quantitative Einfluss verschiedener scharfer und unscharfer Grenzen auf die syntaktischen Dialektunterschiede von Orten in der deutschen Schweiz. Als *unscharfe* Grenzen wurden vier verschiedene Distanzmasse definiert. Die «euklidischen Distanzen» und «Least-Cost-Path-Distanzen» dienten als rein geographische Distanzmasse zwischen den untersuchten Ortschaften. Zwei Gravitations-Indizes beinhalteten neben ebendiesen geographischen Distanzmassen noch zusätzlich die Bevölkerungszahl der Orte. Damit sollte aufgezeigt werden, ob die Hinzunahme eines demographischen Faktors die sprachlichen Unterschiede der deutschschweizer Ortschaften besser erklären kann. Als *scharfe* Grenzen dienten die Kantons-, Konfessionsgebiets- und Arbeitsmarktregionsgrenzen. Diese drei scharfen Grenzarten beeinflussen das gesellschaftliche Zusammenleben der Schweiz. Mit Korrelationsuntersuchungen und Mittelwertstests wurde zuerst der Einfluss der einzelnen scharfen und unscharfen Grenzarten untersucht. In einem zweiten Schritt dienten lineare, multiple Regressionsmodelle dazu, alle scharfen und unscharfen Grenzen zusammen in ein Modell einzubinden. Dies mit dem Ziel, die Dialektunterschiede der Deutschschweiz möglichst gut erklären zu können.

Sowohl für die Kantons- als auch die Arbeitsmarktregionsgrenzen zeigten sich ähnlich starke Effekte auf die syntaktischen Unterschiede der Orte. Die Konfessionsgrenzen wiesen von den drei scharfen Grenzarten über alle acht Nb-Skalen die grössten Effektstärken auf. Bei den Regressionsmodellen erklärten über sämtliche acht untersuchten Nb-Skalen diejenigen die meiste Varianz, welche als Distanzmass die LCP-Distanzen verwendeten. Mit gut 47% konnte das Regressionsmodell der 50-Nb-Skala mit dem LCP-Distanzmass sowie den drei scharfen Grenzen am meisten syntaktische Varianz erklären.

**Die zwei ersten Ziele** dieser Arbeit waren, sowohl den Einfluss von diskreten Grenzen als auch den Einfluss des geographischen Raumes auf die sprachlichen Unterschiede zu untersuchen. Beide Zielsetzungen wurden erreicht. Die Resultate zeigten deutlich auf, dass die untersuchten Grenzarten in Zusammenhang mit den Dialektunterschieden der Deutschschweiz stehen. Die Theorie, dass politische, religiöse, wirtschaftliche und geographische Grenzen als Sprachbarrie-

ren wirken, konnte bestätigt werden. Dies zeigte sich sowohl in den Einzeluntersuchungen der Korrelationsanalysen und der Mittelwertstests als auch bei den Regressionsmodellen. Die Erwartung, dass der Miteinbezug der Bevölkerungszahl der Ortschaften deren Dialektunterschiede noch besser erklärt, konnte jedoch nicht bestätigt werden. Bei den Regressionsmodellen führte die Hinzunahme der Bevölkerung zu einer Abnahme der erklärten Dialektunterschiede. Die Resultate der Regressionsanalysen zeigten auf, dass die Hinzunahme der drei scharfen Grenzarten die Dialektunterschiede zwischen den Ortschaften signifikant besser erklärt.

**Das dritte Ziel** dieser Arbeit war, ein Modell zu erstellen, das mit dem gleichzeitigen Einbezug von mehreren erklärenden Faktoren die syntaktischen Unterschiede zwischen den Ortschaften möglichst gut erklärt. Mit den Resultaten der linearen, multiplen Regressionmodelle wurde auch dieses Ziel erreicht.

Der Nutzen eines zusammenfassenden, quantitativen Modells, das gleichzeitig mehrere scharfe und unscharfe Grenzen miteinbezieht und damit die Dialektunterschiede zwischen Ortschaften möglichst genau erklärt, ist vielfältig. So hilft es, Sprachveränderungen besser zu verstehen, indem es räumliche Prozesse und Strukturen innerhalb der Dialektverteilung aufdeckt. Dies ist zum einen für die Sprachwissenschaften von Interesse, zum anderen auch für alle anderen Wissenschaften, welche sich mit dem Einfluss von historisch-kulturellen oder topographischen Grenzen auf eine Gesellschaft auseinandersetzen. Als Folgearbeit ist es deshalb denkbar, dass ein empirisch-quantitatives Modell zur Interpolation von Dialektdaten oder auch in der Wirtschaftsgeographie eingesetzt wird.

# Anhang

	Ort	Nachbarort	SyntDist60	Faktor					
1.	Bleienbach	Langenthal	22.88	0.99559	...	...	...	...	...
2.	Walenstadt	Flums	17.65	0.99730	984.	Visperterminen	Simplon-Dorf	19.48	1.98316
3.	Sitterdorf	Amriswil	17.85	0.99829	985.	Fiesch	Grindelwald	46.52	1.98561
4.	Fällanden	Uster	20.85	0.99838	986.	Vals	Safien	24.38	1.98658
5.	Solothurn	Bettlach	22.39	0.99872	987.	Gadmen	Giswil	49.96	1.99386
6.	Wädenswil	Rapperswil (SG)	16.56	1.00143	988.	Engelberg	Melchtal	25.88	2.00107
7.	Lenzburg	Aarau	17.92	1.00302	989.	Iseltwald	Schangnau	33.85	2.00488
8.	Rümlang	Bilach	18.67	1.00317	990.	Churwalden	Arosa	27.96	2.02919
9.	Illnau	Pfäffikon	15.92	1.00678	991.	Inden	Adelboden	43.21	2.03130
10.	Stein (AG)	Kaisten	22.11	1.01003	992.	Muotathal	Unterschächen	36.28	2.03399
11.	Flums	Mels	20.44	1.01014	993.	St. Niklaus	Agarn	22.76	2.05020
12.	Basadingen	Unterstammheim	19.56	1.01018	994.	Brienz (BE)	Marbach (LU)	48.47	2.06090
13.	Hünenberg	Rotkreuz	20.64	1.01019	995.	Weisstannen	Valens	23.99	2.06512
14.	Iseltwald	Brienz (BE)	26.17	1.01105	996.	Engelberg	Gurtellen	32.05	2.07973
15.	Reckingen	Oberwald	25.23	1.01430	997.	St. Niklaus	Saas Grund	28.05	2.08611
16.	Rümlang	Zürich	21.11	1.01494	998.	Kandersteg	Mürren	41.67	2.08960
17.	Hünenberg	Zug	15.23	1.01703	999.	Reckingen	Grindelwald	43.92	2.09168
18.	Tuggen	Benken (SG)	23.04	1.01717	1000.	Kiental	Mürren	42.94	2.09397
19.	Aarburg	Zofingen	19.30	1.01770	1001.	Rheinwald	Vals	25.65	2.10024
20.	Boswil	Merenschwand	17.74	1.01795	1002.	Frutigen	Diemtigen	22.66	2.10331
21.	Niederrohrdorf	Birmenstorf (AG)	20.77	1.01841	1003.	Lenk	Inden	47.23	2.11229
22.	Murten	Gurmels	37.60	1.01912	1004.	Visperterminen	Bürchen	22.18	2.11287
23.	Riken	Niederbipp	21.05	1.02141	1005.	Oberberg	Linthal	35.63	2.12323
24.	Solothurn	Bibern (SO)	26.50	1.02206	1006.	Davos-Monstein	Arosa	37.59	2.12402
25.	Bleienbach	Aeschi (SO)	26.74	1.02450	1007.	Inden	Ferden	36.60	2.13230
26.	Ellikon an der Thur	Frauenfeld	18.05	1.02466	1008.	Habkern	Iseltwald	25.43	2.13623
27.	Oberriet (SG)	Diepoldsau	18.24	1.02509	1009.	Brig	Blatten	37.59	2.13837
28.	Muttenz	Pratteln	19.26	1.02600	1010.	Walenstadt	Grabs	24.68	2.14598
29.	Ins	Murten	17.11	1.02618	1011.	Guttannen	Göschenen	68.03	2.15582
30.	Näfels	Mollis	15.00	1.02807	1012.	Betten	Blatten	36.83	2.16107
31.	Stüsslingen	Aarau	20.47	1.02818	1013.	Grindelwald	Brienz (BE)	26.21	2.16144
32.	Ins	Laupen	20.37	1.02876	1014.	Wiesen (GR)	Arosa	32.77	2.16649
33.	Brunnen	Schwyz	19.07	1.02888	1015.	Oberberg	Luchsingen	33.04	2.19416
34.	Uster	Nänikon	20.69	1.02897	1016.	Ausserberg	Blatten	31.28	2.19450
35.	Ins	Täuffelen	15.14	1.02955	1017.	Guttet-Feschel	Agarn	19.69	2.23686
36.	Lenzburg	Häggingen	22.30	1.02981	1018.	Zermatt	Saas Grund	26.82	2.24607
37.	Giswil	Melchtal	22.41	1.02983	1019.	Obstallden	Amden	22.35	2.25029
38.	Aesch (BL)	Basel Stadt	24.94	1.03118	1020.	Boltigen	Schwarzsee	40.20	2.26624
39.	Näfels	Glarus	18.19	1.03159	1021.	Kiental	Lauterbrunnen	36.36	2.26678
40.	Luchsingen	Schwanden (GL)	21.12	1.03209	1022.	Flums	Grabs	21.35	2.28358
41.	Lenzburg	Menziken	21.70	1.03238	1023.	Blatten	Kandersteg	54.94	2.30968
42.	Lachen	Eschenbach (SG)	19.25	1.03245	1024.	Ferden	Kandersteg	50.39	2.31288
43.	Lupfig	Brugg	16.61	1.03493	1025.	Grindelwald	Iseltwald	30.85	2.31549
44.	Grosswangen	Ruswil	21.08	1.03510	1026.	Adelboden	Kandersteg	29.96	2.31624
45.	Basel Stadt	Muttenz	22.64	1.03569	1027.	Mörel	Blatten	32.96	2.31746
46.	Langenthal	Niederbipp	20.95	1.03579	1028.	Walenstadt	Wildhaas	24.02	2.32673
47.	Tüscherz	Täuffelen	21.53	1.03609	1029.	Lauterbrunnen	Mürren	29.19	2.34677
48.	Aeschi (SO)	Utzenstorf	23.66	1.03697	1030.	Blatten	Mürren	47.63	2.36775
49.	Winterthur	Ellikon an der Thur	22.88	1.03707	1031.	Vättis	Chur	18.29	2.36828
50.	Merenschwand	Hedingen	23.77	1.03730	1032.	Bürchen	Ausserberg	18.74	2.37267
51.	Aarau	Muhen	31.21	1.03760	1033.	Fiesch	Mürren	42.71	2.37681
52.	Bassersdorf	Fällanden	18.55	1.03835	1034.	Stein (SG)	Quarten	24.31	2.40430
53.	Illnau	Nänikon	17.80	1.03862	1035.	Stein (SG)	Walenstadt	20.10	2.40667
54.	Diepoldsau	Altstätten	18.84	1.03903	1036.	Weisstannen	Vättis	22.15	2.43710
55.	Neftenbach	Andelfingen	21.23	1.03916	1037.	Vättis	Trimmis	18.74	2.45012
56.	Riken	Aarburg	22.80	1.03936	1038.	Simplon-Dorf	Saas Grund	28.10	2.46053
57.	Homburg	Weinfelden	18.67	1.04019	1039.	Betten	Mürren	43.60	2.46835
58.	Solothurn	Utzenstorf	34.24	1.04030	1040.	Valens	Untervaz	20.41	2.49708
59.	Amriswil	Roggwil (TG)	20.58	1.04065	1041.	Vättis	Untervaz	12.48	2.64565
60.	Urdorf	Rümlang	16.32	1.04084	1042.	Mürren	Wengen	25.85	2.66068
61.	Frutigen	Reichenbach	21.83	1.04092	1043.	Randa	Saas Grund	33.67	2.83350
...	...	...	...	...	1044.	Gadmen	Engelberg	47.58	2.87848
...	...	...	...	...	1045.	Lauterbrunnen	Wengen	27.81	3.39087

Abbildung 9.1: Ausschnitte aus der Tabelle aller 1045 Voronoi-Nachbarschaftspaaren, welche nach dem Topographie-Einflussfaktor (Spalte "Faktor") sortiert sind. Der linke Teil zeigt die 61 Ortspaare, deren Reisezeiten am schwächsten von der Topographie beeinflusst sind. Die rechte Seite zeigt die 61 Ortspaare (1045. bis 984.), deren Reisezeiten am stärksten von der Topographie beeinflusst sind. Die jeweils 25, 50, 100, 250 und 500 Voronoi-Ortspaare, deren Reisezeiten am schwächsten resp. stärksten von der Topographie beeinflusst sind, werden in Gruppen zusammengefasst und deren syntaktische Distanzen in Abbildung 6.1 auf Seite 55 mit Boxplots miteinander verglichen.

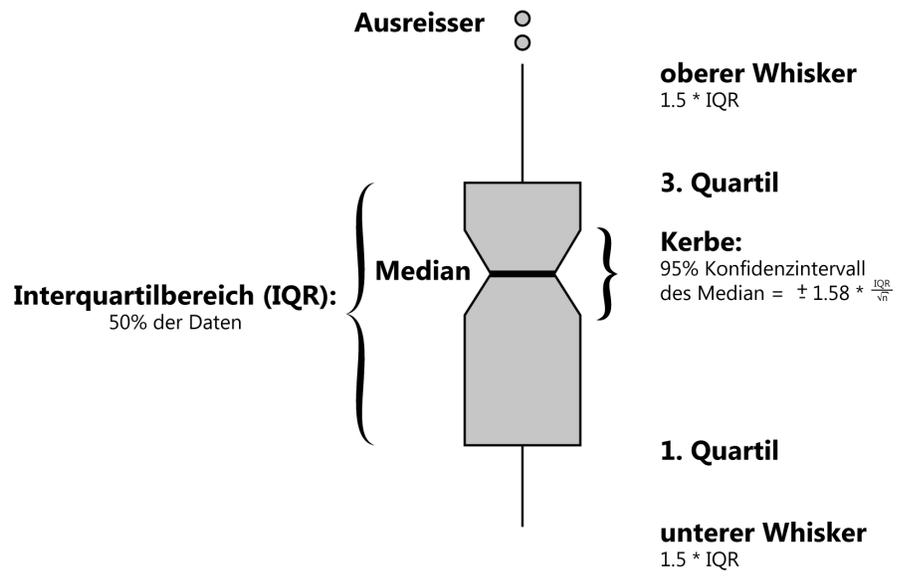
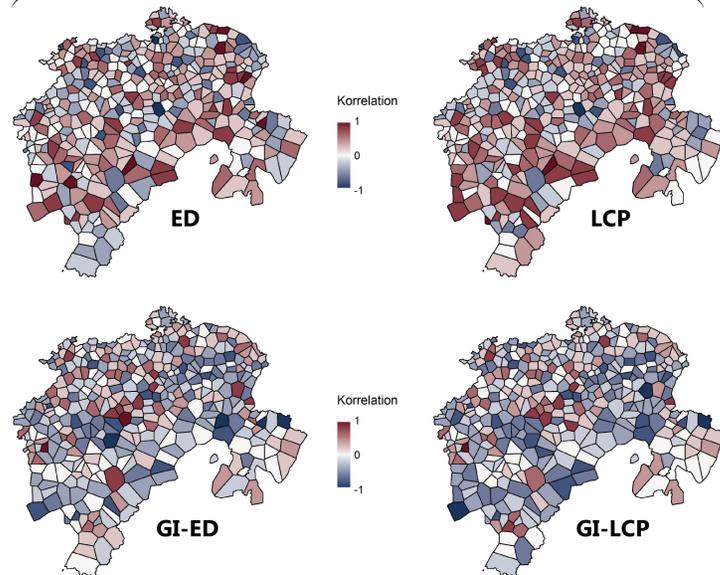


Abbildung 9.2: Erklärung zur Bedeutung der Boxplot-Darstellung. Modifiziert nach Doyle (2016)

## 5 nächste Nachbarn



## 10 nächste Nachbarn

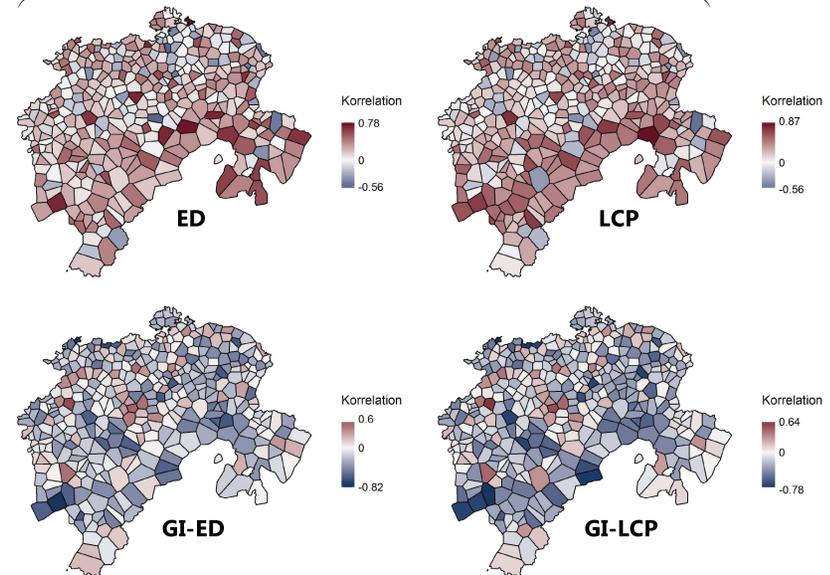


Abbildung 9.3: Korrelationskarten der syntaktischen Distanz mit unterschiedlichen Distanzmassen und veränderten Nb-Skalen.

## 15 nächste Nachbarn

## 20 nächste Nachbarn

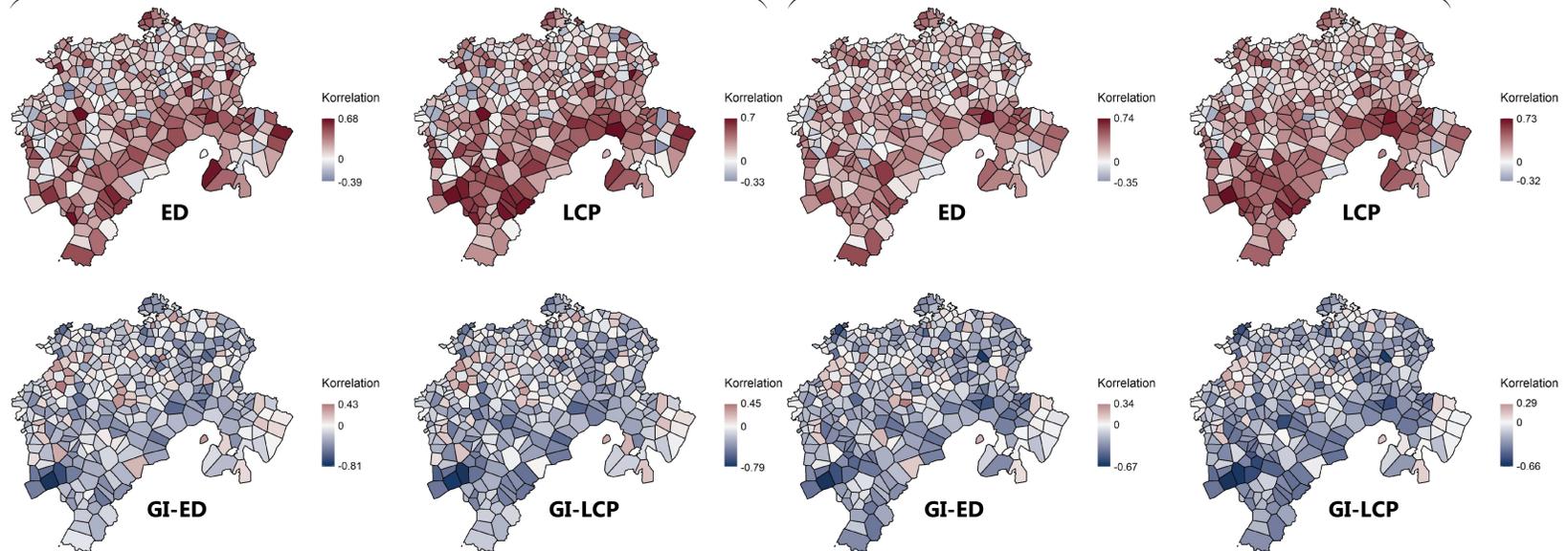


Abbildung 9.4: Korrelationskarten der syntaktischen Distanz mit unterschiedlichen Distanzmassen und veränderten Nb-Skalen.

## 25 nächste Nachbarn

## 30 nächste Nachbarn

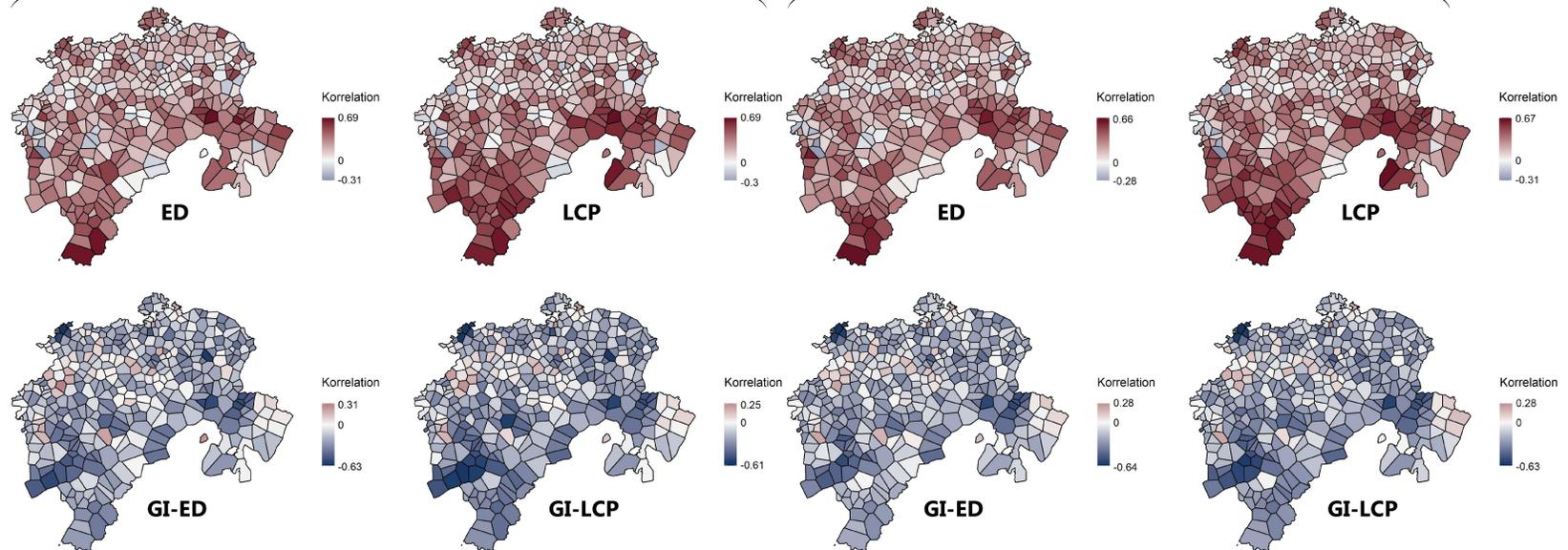


Abbildung 9.5: Korrelationskarten der syntaktischen Distanz mit unterschiedlichen Distanzmassen und veränderten Nb-Skalen.

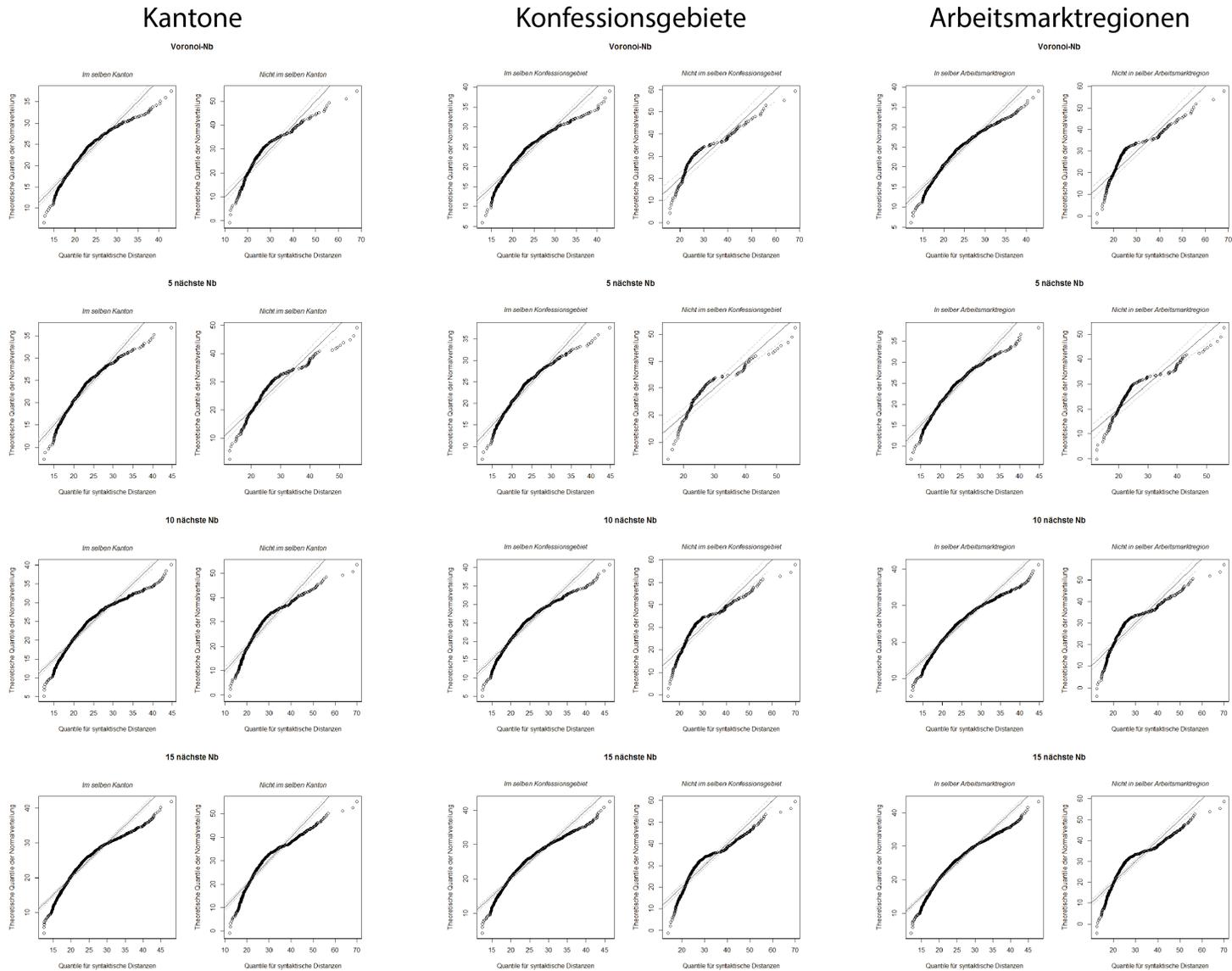
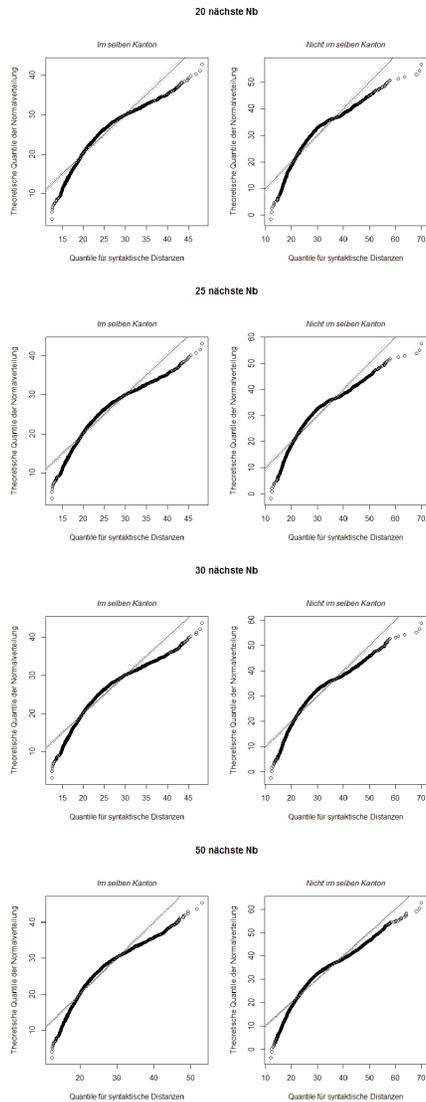
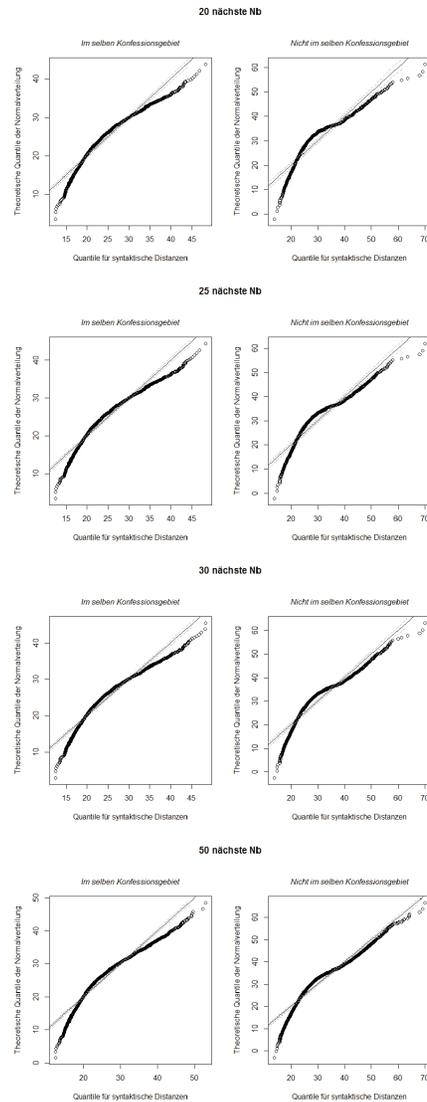


Abbildung 9.6: Die Q-Q Plots der syntaktischen Distanzen von Ortsparen unterschiedlicher Nb-Skalen, die (nicht) im selben Kanton (linke Spalte), (nicht) im selben Konfessionsgebiet (mittlere Spalte) resp. (nicht) in derselben Arbeitsmarktregionen liegen (rechte Spalte). Die gestrichelten Linien und die darin mittigen schwarze Linien zeigen die theoretischen Normalverteilungen resp. deren 99%-Konfidenzintervalle an.

## Kantone



## Konfessionsgebiete



## Arbeitsmarktregionen

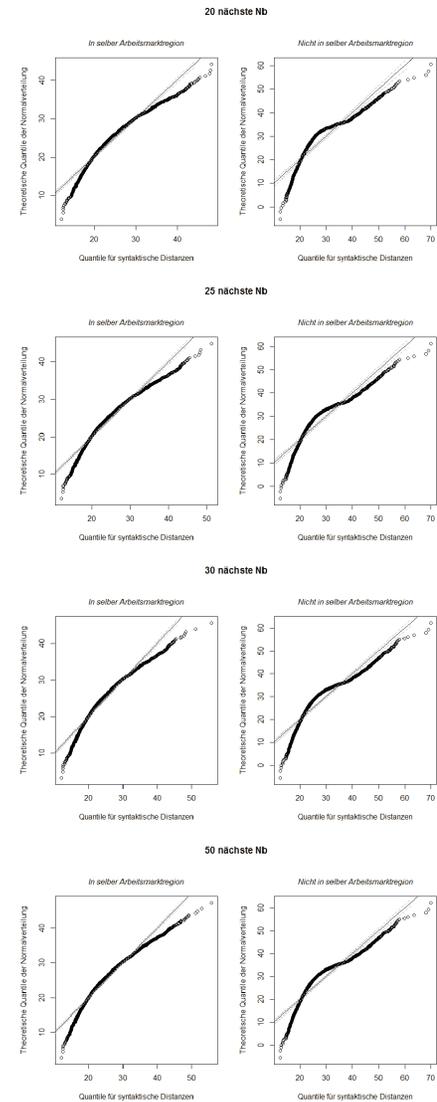


Abbildung 9.7: Die Q-Q Plots der syntaktischen Distanzen von Ortspaaren unterschiedlicher Nachbarschaftsgrößen, die (nicht) im selben Kanton (linke Spalte), (nicht) im selben Konfessionsgebiet (mittlere Spalte) resp. (nicht) in derselben Arbeitsmarktregionen liegen (rechte Spalte). Die gestrichelten Linien und die darin mittigen schwarze Linien zeigen die theoretischen Normalverteilungen resp. deren 99%-Konfidenzintervalle an.

# Literatur

- Akaike, Hirotugu (1973): „Information theory and an extension of the maximum likelihood principle.“ In: *B. Petrov und B. Csaki (Eds). Second International Symposium on Information Theory*, S. 267–281.
- Andrews et al. (2012): *2012 Presidential Election Results*. Zugriff: 06.06.2017. URL: <http://www.washingtonpost.com/wp-srv/special/politics/election-map-2012/president/>.
- Atkinson, Peter und Tate, Nicholas (2000): „Spatial scale problems and geostatistical solutions: a review“. In: *The Professional Geographer* 52.4, S. 607–623.
- Auer, Peter (2013): „The Geography of Language: Steps toward a new approach“. In: *Freiburger Arbeitspapiere zur Germanistischen Linguistik (FRAGL)*. URL: <http://portal.unifreiburg.de/sdd/fragl/2013.16>.
- Auer et al. (2011): „Vertical vs. horizontal change in the traditional dialects of southwest Germany: a quantitative approach“. In: *taal en tongval* 63.1, S. 13–41.
- Backhaus, Klaus et al. (2011): *Multivariate Analysemethoden: eine anwendungsorientierte einföhrung*. Springer-Verlag.
- Bartholy, Heike (1992): *Sprache, kulturelle Identität und Unabhängigkeit, dargestellt am Beispiel Maltas*. Schuch.
- Berry, Joseph (2007): *Map analysis: understanding spatial patterns and relationships*. GeoTec Media San Francisco, CA.
- Blommaert, Jan (2010): *The sociolinguistics of globalization*. Cambridge University Press.
- Bowern, Claire (2013): „Relatedness as a factor in language contact“. In: *Journal of Language Contact* 6.2, S. 411–432.
- Bowern, Claire und Evans, Bethwyn (2015): *The Routledge handbook of historical linguistics*. Routledge.
- Britain, David (2012): „Countering the urbanist agenda in variationist sociolinguistics: dialect contact, demographic change and the rural-urban dichotomy“. In: *Hansen, Sandra; Schwarz, Christian; Stoeckle, Philipp; Streck, Tobias (Eds.): Dialectological and folk dialectological concepts of space*. S. 12–30.
- Bucheli, Claudia und Glaser, Elvira (2002): „The syntactic atlas of Swiss German dialects: empirical and methodological problems“. In: *Syntactic microvariation* 2, S. 41–73.
- Bundesamt für Statistik (2016): *Pendlermobilität (PEND)*. Zugriff: 12.06.2017. URL: <https://www.bfs.admin.ch/bfs/de/home/statistiken/mobilitaet-verkehr/personenverkehr/pendlermobilitaet.html>.

- Champion, Tony, Coombes, Mike und Brown, David (2009): „Migration and longer-distance commuting in rural England“. In: *Regional Studies* 43.10, S. 1245–1259.
- Christen, Helen et al. (2010): *Kleiner Sprachatlas der deutschen Schweiz*. Huber.
- Cohen, Jacob (1988): „Statistical power analysis for the behavioral sciences (2nd edition)“. In: — (1992): „Statistical power analysis“. In: *Current directions in psychological science* 1.3, S. 98–101.
- Croft, William (2000): *Explaining language change: An evolutionary approach*. Pearson Education.
- Dijkstra, Edsger W (1959): „A note on two problems in connexion with graphs“. In: *Numerische mathematik* 1.1, S. 269–271.
- Domingos, Pedro (1999): „The role of Occam’s razor in knowledge discovery“. In: *Data mining and knowledge discovery* 3.4, S. 409–425.
- Doyle, David (2016): *Notched Box Plots*. Zugriff: 15.06.2017. URL: <https://sites.google.com/site/davidstatistics/home/notched-box-plots>.
- Durlak, Joseph (2009): „How to select, calculate, and interpret effect sizes“. In: *Journal of pediatric psychology*.
- ESRI (2017): *ArcGIS 10.4.1 for Desktop*.
- Esser, Paul (1983): *Dialekt und Identität: Diglottale Sozialisation und Identitätsbildung*. Lang.
- Falck, Oliver, Lameli, Alfred und Ruhose, Jens (2016): „Cultural biases in migration: Estimating non-monetary migration costs“. In: *Papers in Regional Science*.
- Falck, Oliver et al. (2012): „Dialects, cultural identity, and economic exchange“. In: *Journal of urban economics* 72.2, S. 225–239.
- Fotheringham, Stewart, Brunson, Chris und Charlton, Martin (2002): *Geographically weighted regression*. Hoboken.
- Fotheringham, Stewart und Wong, David (1991): „The modifiable areal unit problem in multivariate statistical analysis“. In: *Environment and planning A* 23.7, S. 1025–1044.
- GITTA (2015): *Thiessen-Polygone*. URL: [http://www.gitta.info/Accessibilit/de/html/UncProxAnaly\\_learningObject4.html](http://www.gitta.info/Accessibilit/de/html/UncProxAnaly_learningObject4.html).
- Glaser, Elvira (2013): „Area formation in morphosyntax“. In: *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives*, S. 195–221.
- (2014): „Wandel und Variation in der Morphosyntax der schweizerdeutschen Dialekte“. In: *Taal en Tongval* 66.1, S. 21–64.
- Glaser, Elvira und Bart, Gabriela (2015): „Dialektsyntax des Schweizerdeutschen“. In: *Regionale Variation des Deutschen. Projekte und Perspektiven*, S. 79–105.
- Goebel, Hans (1982): *Dialektometrie*. Bd. 157. Verlag der Österreichischen Akademie der Wissenschaften.
- Gooskens, Charlotte (2004): „Norwegian dialect distances geographically explained“. In: *Language Variation in Europe. Papers from the Second International Conference on Language Variation in Europe ICLAVE*. Bd. 2, S. 12–14.

- Gower, John (1966): „Some distance properties of latent root and vector methods used in multivariate analysis“. In: *Biometrika* 53.3-4, S. 325–338.
- Green, Anne, Hogarth, Terence und Shackleton, Ruth (1999): „Longer distance commuting as a substitute for migration in Britain: a review of trends, issues and implications“. In: *International Journal of Population Geography* 5.1, S. 49.
- Haag, Karl (1898): „Die Mundarten des oberen Neckar- und Donaulandes“. In: *Buchdruckerei Egon Hutzler, Reutlingen*.
- Hamming, Richard (1950): „Error detecting and error correcting codes“. In: *Bell Labs Technical Journal* 29.2, S. 147–160.
- Heeringa, Wilbert und Nerbonne, John (2001): „Dialect areas and dialect continua“. In: *Language Variation and Change* 13.03, S. 375–400.
- Hernández-Campoy, Juan Manuel (2003): „Exposure to contact and the geographical adoption of standard features: Two complementary approaches“. In: *Language in Society* 32.02, S. 227–255.
- Herzog, Irmela (2013): „Berechnung von optimalen Wegen am Beispiel der Zeitstraße“. In: *Archäologische Informationen* 31.1&2, S. 67–96.
- Hotzenköcherle, Rudolf (1962): *Sprachatlas der deutschen Schweiz: Band 1*. Francke.
- Howell, David (2012): *Statistical methods for psychology*. Cengage Learning.
- Imhof, Eduard et al. (1950): *Gelände und Karte*.
- Ingraham, Christopher (2017): *Republicans and Democrats both try to gerrymander, but only one party is any good at it*. Zugriff: 06.06.2017. URL: [https://www.washingtonpost.com/news/wonk/wp/2017/06/06/republicans-and-democrats-both-try-to-gerrymander-but-only-one-of-them-is-any-good-at-it/?utm\\_term=.1a155c9199fa](https://www.washingtonpost.com/news/wonk/wp/2017/06/06/republicans-and-democrats-both-try-to-gerrymander-but-only-one-of-them-is-any-good-at-it/?utm_term=.1a155c9199fa).
- Jeszszky, Péter und Weibel, Robert (2014): „Correlating morphosyntactic dialect variation with geographic distance: Local beats global“. In: *Extended Abstract Proceedings of the GIScience*, S. 186–191.
- Jeszszky, Péter et al. (angenommen): „Exploring global and local patterns in the correlation of geographic distances and morphosyntactic variation in Swiss German“. In: *Journal of Linguistic Geography*.
- Johnston, Ron (2002): „Manipulating maps and winning elections: measuring the impact of malapportionment and gerrymandering“. In: *Political Geography* 21.1, S. 1–31.
- Jorio, Marco (2015): *Wiener Kongress, in: Historisches Lexikon der Schweiz (HLS)*. Version vom 03.02.2015, Zugriff: 09.05.2017. URL: <http://www.hls-dhs-dss.ch/textes/d/D8922.php>.
- Kellerhals, Sandra (2014): „Dialektometrische Analyse und Visualisierung von schweizerdeutschen Dialekten auf verschiedenen linguistischen Ebenen“. Magisterarb. Abteilung Geographie der Universität Zürich, Schweiz: Universität Zürich.
- Kendall, Maurice (1938): „A new measure of rank correlation“. In: *Biometrika* 30.1/2, S. 81–93.
- Kley, Andreas (2016): *Kantone, in: Historisches Lexikon der Schweiz (HLS)*. Version vom 13.04.2016, Zugriff: 09.05.2017. URL: <http://www.hls-dhs-dss.ch/textes/d/D26414.php>.
- Kolmogoroff, Alexander (1933): „Grundbegriffe der Wahrscheinlichkeitsrechnung“. In:

- Kürschner, Sebastian und Gooskens, Charlotte (2011): „Verstehen nah verwandter Varietäten über Staatsgrenzen hinweg.“ In:
- Lameli, Alfred (2013): *Strukturen im Sprachraum: Analysen zur arealtypologischen Komplexität der Dialekte in Deutschland*. Bd. 54. Walter de Gruyter.
- Lau, Thomas (2013): *Zweiter Villmergerkrieg*, in: *Historisches Lexikon der Schweiz (HLS)*. Version vom 28.02.2013, Zugriff: 09.05.2017. URL: <http://www.hls-dhs-dss.ch/textes/d/D8911.php>.
- (2014): *Erster Villmergerkrieg*, in: *Historisches Lexikon der Schweiz (HLS)*. Version vom 22.01.2014, Zugriff: 09.05.2017. URL: <http://www.hls-dhs-dss.ch/textes/d/D8910.php>.
- Lee, Jay und Kretschmar, William (1993): „Spatial analysis of linguistic data with GIS functions“. In: *International Journal of Geographical Information Science* 7.6, S. 541–560.
- Levenshtein, Vladimir (1966): „Binary codes capable of correcting deletions, insertions, and reversals“. In: *Soviet physics doklady*. Bd. 10. 8, S. 707–710.
- Lucas, Christopher (2015): „Contact-induced language change“. In: *Claire Bowerman & Bethwyn Evans (red.), The Routledge Handbook of Historical Linguistics*. London: Routledge, S. 519–536.
- Lyons, John (1968): *Introduction to theoretical linguistics*. Cambridge university press.
- Massey, Douglas, Rothwell, Jonathan und Domina, Thurston (2009): „The changing bases of segregation in the United States“. In: *The Annals of the American Academy of Political and Social Science* 626.1, S. 74–90.
- NASA, Landsat Program (1994): *Landsat ETM+ L7119502702720070715RGB, Switzerland*.
- Nass, Steven (2015). Zugriff: 06.06.2017. URL: <https://commons.wikimedia.org/w/index.php?curid=38869847>.
- Nerbonne, John und Kleiweg, Peter (2007): „Toward a dialectological yardstick“. In: *Journal of Quantitative Linguistics* 14.2-3, S. 148–166.
- Nerbonne, John und Wiersma, Wybo (2006): „A measure of aggregate syntactic distance“. In: *Proceedings of the Workshop on linguistic Distances*. Association for Computational Linguistics, S. 82–90.
- Nerbonne, John et al. (1996): „Phonetic distance between Dutch dialects“. In: *CLIN VI: Proceedings of the sixth CLIN meeting*, S. 185–202.
- Newman, Mark (2016): *Maps of the 2016 US presidential election results*. Zugriff: 06.06.2017. URL: <http://www-personal.umich.edu/~mejn/election/2016/>.
- Nichols, Johanna (2013): „The vertical archipelago: Adding the third dimension to linguistic geography“. In: *Space in language and linguistics: Geographical, interactional, and cognitive perspectives*, S. 38–60.
- O’Grady, William, Dobrovolsky, Michael und Katamba, Francis (1997): *Contemporary linguistics*. St. Martin’s.
- Openshaw, Stan und Taylor, Peter (1979): „A million or so correlation coefficients: three experiments on the modifiable areal unit problem“. In: *Statistical applications in the spatial sciences* 21, S. 127–144.
- Pickl, Simon (2013): *Probabilistische Geolinguistik: Geostatistische Analysen lexikalischer Variation in Bayerisch-Schwaben*. Franz Steiner Verlag.

- Pickl, Simon et al. (2014): „Linguistic distances in dialectometric intensity estimation“. In: *Journal of Linguistic Geography* 2.01, S. 25–40.
- Preparata, Franco und Shamos, Michael (2012): *Computational geometry: an introduction*. Springer Science & Business Media.
- Reichmann, Eva (2000): *Narrative Konstruktion nationaler Identität*. Röhrig.
- Rouwendaal, Jan und Rietveld, Piet (1994): „Changes in commuting distances of Dutch households“. In: *Urban studies* 31.9, S. 1545–1557.
- Schmidt, Jürgen Erich und Herrgen, Joachim (2001): „Digitaler Wenker-Atlas (DiWA)“. In: *Bearbeitet von Alfred Lameli, Tanja Giessler, Roland Kehrein, Alexandra Lenz, Karl-Heinz Müller, Jost Nickel, Christoph Purschke und Stefan Rabanus. Erste vollständige Ausgabe von Georg Wenkers „Sprachatlas des Deutschen Reichs*.
- Schnyder, Caroline (2013): *Die Reformation, in: Historisches Lexikon der Schweiz (HLS)*. Version vom 29.01.2013, Zugriff: 09.05.2017. URL: <http://www.hls-dhs-dss.ch/textes/d/D13328.php>.
- Schuler, Martin et al. (2005): *Eidgenössische Volkszählung 2000: Die Raumgliederung der Schweiz*. Neuchâtel, Switzerland: Bundesamt für Statistik BFS.
- Schwarz, Gideon (1978): „Estimating the dimension of a model“. In: *The annals of statistics* 6.2, S. 461–464.
- Serlo.org.
- Siegel, Jeff (2008): *The emergence of pidgin and creole languages*. Oxford University Press.
- Spearman, Charles (1904): „General Intelligence Objectively Determined and Measured“. In: *The American Journal of Psychology* 15.2, S. 201–292.
- Spruit, Marco René (2006): „Measuring syntactic variation in Dutch dialects“. In: *Literary and linguistic computing* 21.4, S. 493–506.
- Spruit, Marco René, Heeringa, Wilbert und Nerbonne, John (2009): „Associations among linguistic levels“. In: *Lingua* 119.11, S. 1624–1642.
- Sullivan, Gail und Feinn, Richard (2012): „Using effect size—or why the p value is not enough“. In: *Journal of graduate medical education* 4.3, S. 279–282.
- Szmrecsanyi, Benedikt (2011): „Corpus-based dialectometry: a methodological sketch“. In: *Corpora* 6.1, S. 45–76.
- (2012): „Geography is overrated“. In: *Dialectological and folk dialectological concepts of space*, S. 215–231.
- Thomason, Sarah (2001): *Language contact: an introduction*.
- Tobler, Waldo (1970): „A computer movie simulating urban growth in the Detroit region“. In: *Economic geography* 46.sup1, S. 234–240.
- (1989): „Frame independent spatial analysis“. In: *Accuracy of spatial databases*, S. 115–122.
- (1993): *Three presentations on geographical analysis and modeling*. NCGIA.
- Trudgill, Peter (1974): „Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography“. In: *Language in society* 3.2, S. 215–246.
- (2010): „Contact and Sociolinguistic Typology“. In: *The handbook of language contact*, S. 299.

- Universität Zürich (2016a): *Mann-Whitney U Test*. Zugriff: 13.09.2017. URL: <http://www.methodenberatung.uzh.ch/de/datenanalyse/unterschiede/zentral/mann.html>.
- (2016b): *Syntaktischer Atlas der Deutschen Schweiz SADS*. Zugriff: 13.09.2017. URL: <http://www.dialektsyntax.uzh.ch/de.html>.
- (2017a): *Einfache lineare Regression*. Zugriff: 12.06.2017. URL: <http://www.methodenberatung.uzh.ch/de/datenanalyse/zusammenhaenge/ereg.html>.
- (2017b): *Shapefile der Kantons Grenzen der Schweiz*. Datensatz vom Modul GEO372 des geographischen Instituts der Universität Zürich. URL: <http://www.geo.uzh.ch/microsite/geo372/>.
- van Etten, Jacob (2012): *R package gdistance: distances and routes on geographical grids (version 1.1-4)*.
- (2015): *gdistance: Distances and Routes on Geographical Grids*. R package version 1.1-9. URL: <https://CRAN.R-project.org/package=gdistance>.
- Volker, Martin (2006): „Reporting effect size estimates in school psychology research“. In: *Psychology in the Schools* 43.6, S. 653–672.
- Wong, David (2009): „The modifiable areal unit problem (MAUP)“. In: *The SAGE handbook of spatial analysis*, S. 105–123.
- Würgler, Andreas (2012): *Eidgenossenschaft*, in: *Historisches Lexikon der Schweiz (HLS)*. Version vom 08.02.2012, Zugriff: 09.05.2017. URL: <http://www.hls-dhs-dss.ch/textes/d/D26413.php>.



## PERSÖNLICHE ERKLÄRUNG

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

Unterschrift:

---