

# Sprachdistanzmessung in einem Forschungsprojekt zu den alemannischen Dialekten im Churer Rheintal in der Schweiz

VITTORIO DELL'AQUILA, FORKNINGSCENTRUM  
FÖR EUROPEISK FLESPRÅKIGHET<sup>1</sup>  
OSCAR ECKHARDT, INSTITUT  
FÜR KULTURFORSCHUNG GRAUBÜNDEN

## 1. Das Forschungsprojekt

Seit Januar 2012 läuft das Forschungsprojekt "Regionalisierung der Dialekte im Churer Rheintal". Das Projekt wird finanziert vom Schweizerischen Nationalfonds zur Förderung der wissenschaftlichen Forschung (SNF) und vom Institut für Kulturforschung Graubünden. Die Arbeit hat zum Ziel, die Veränderungen in der Sprachlandschaft rund um die Graubündner Hauptstadt Chur zu erfassen, zu dokumentieren und zu analysieren. Die Daten des Sprachatlasses der Deutschen Schweiz (SDS) werden für das Untersuchungsgebiet als Basisdialekte der einzelnen Ortschaften angesehen. Diesen werden die Daten aus je zehn Interviews mit Jugendlichen und jungen Erwachsenen gegenübergestellt. Das Projekt versucht unter anderem zu messen, wie gross die sprachlichen Distanzen zwischen den einzelnen Ortsmundarten und dem Churer Dialekt zur Zeit der SDS-Datenerhebung waren und wie gross diese Distanzen heute sind.

Für diese Art der Distanzmessung sind vier Berechnungen nötig.

**Berechnung 1** (horizontaler-arealer Vergleich): Churerdeutsch gemäss SDS // Dialekte der Ortsmundarten gemäss SDS.

**Berechnung 2** (vertikaler Vergleich): Churerdeutsch gemäss SDS // heutiger Churer Dialekt.

**Berechnung 3** (vertikaler Vergleich): Dialekte der Ortsmundarten gemäss SDS // heutige Dialekte der Ortsmundarten.

**Berechnung 4** (horizontal-arealer Vergleich): heutiger Churer Dialekt // heutige Dialekte der Ortsmundarten.

<sup>1</sup> *Centre d'Études Linguistiques pour l'Europe.*

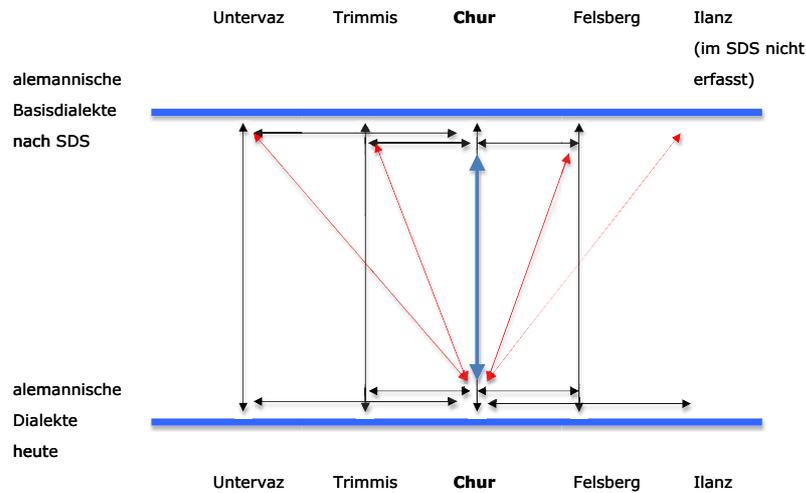


Abbildung 1: Struktur der verschiedenen Distanzmessungen.

Insgesamt weist unsere Studie mit zwölf SDS-Gemeinden zwölf Messpunkte in der Vergangenheit und mit 14 aktuell untersuchten Gemeinden 14 Messpunkte in der Gegenwart auf (SDS-Gemeinden plus Domat/Ems und Ilanz).

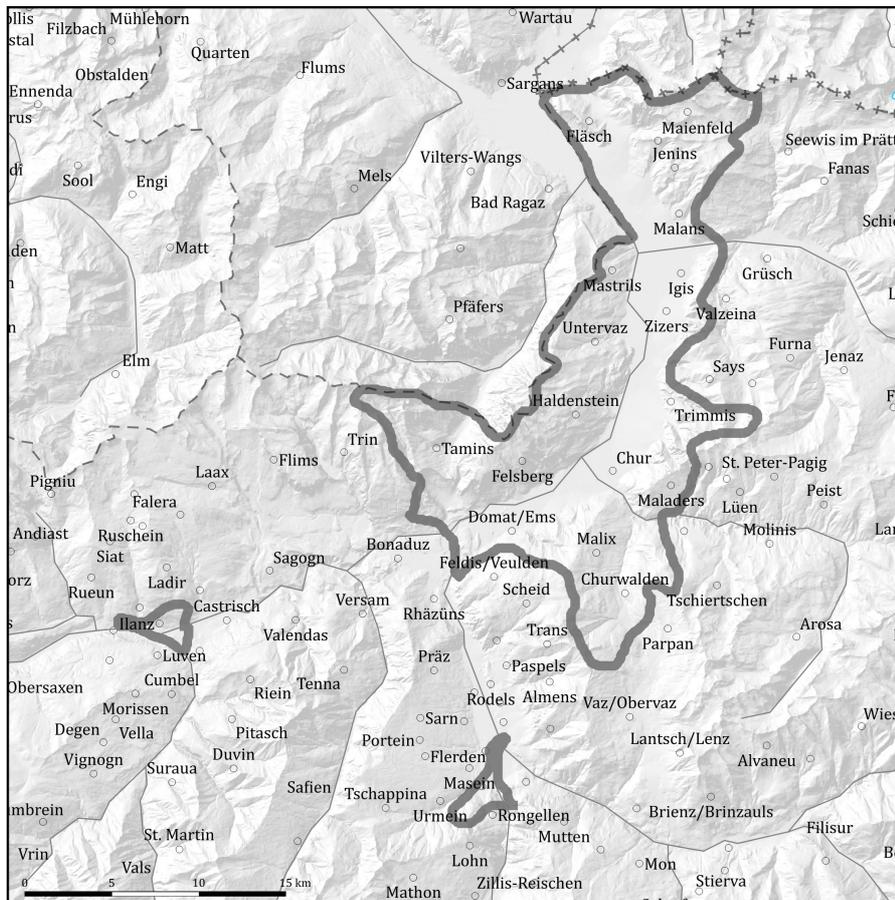


Abbildung 2: Karte des untersuchten Gebiet

## 2. Übersicht über Forschung zur Sprachdistanzmessung

Vorgängig zu unserem Projekt galt es abzuklären, welche Formen von Distanzmessung für das doch relativ kleinräumige Untersuchungsgebiet geeignet sind. In der Folge stellen wir kurz vor, welche hauptsächlichen Richtungen zur Distanzmessung bisher in der Dialektologie Anwendung gefunden haben. Selbstverständlich vereinfachen die folgenden Ausführungen die Bemühungen der Linguisten um die Distanzmessung extrem.

### 2.1 Dialektometrie nach der Salzburger Schule

Als wesentliche Forschungsrichtung ist natürlich die Dialektometrie nach Hans Goebel<sup>2</sup> zu nennen (Goebel 1982). Goebel und sein Team gehen von bereits existierenden Sprachatlanten oder ad-hoc-Sprachatlanten als Basis aus. Die dort vorhandenen Daten werden taxiert und der elektronischen Datenverarbeitung zugeführt. Ein wesentliches Ziel der Dialektometrie besteht darin, möglichst hochrangige Ordnungsstrukturen in sprachgeographischen Netzen aufzufinden.

Wie das Salzburger Team um Goebel selber schreibt, richtet sich das **Forschungsinteresse** in erster Linie auf die regionale Distribution von Dialektähnlichkeiten und betrifft damit “Dialektkerne und Übergangszonen, die durch eine mehr oder weniger geringe Dialektvarianz zwischen benachbarten Orten gekennzeichnet sind”<sup>3</sup>. Für die Dialektometrie werden verschiedene Verfahren benutzt, die aus der numerischen Klassifikation gut bekannt sind. Mit diesen Verfahren gelingt es, aus dem aus den Sprachatlanten gewonnenen Material Basismuster zu abstrahieren und mit verschiedenen bildgebenden Verfahren zu visualisieren.

“Als **Ergebnis** darf nicht eine einzige Klasseneinteilung erwartet werden; vielmehr können durch verschiedene Verfahren jeweils andere Aspekte des gesuchten Basismusters aufgedeckt werden. Prinzipiell sind wir an der Vielfalt der taxometrischen Verfahren und den dazu gehörigen Resultaten bzw. an den daraus erfließenden linguistischen Interpretationen interessiert”<sup>3</sup>.

Um die Informationen aus dem Material der Sprachatlanten zu extrahieren, setzen Goebel und Helfer ein Verfahren ein, das sie Taxierung nennen. “Dabei kann man die zu klassifizierenden Daten (Taxandum) jeweils nach phonetischen, morphosyntaktischen oder lexikalischen Kriterien gruppieren (taxieren). Das Ergeb-

<sup>2</sup> Eine vollständige, aktualisierte Liste der zahlreichen Publikationen Goebels zur Dialektometrie findet sich unter <[https://www.sbg.ac.at/rom/people/prof/goebel/dm\\_publi.htm](https://www.sbg.ac.at/rom/people/prof/goebel/dm_publi.htm)>, (13.2.2014. 8.40).

<sup>3</sup> <<http://www.dialectometry.com/dmdocs/index.html>>, (14.1.2014. 15 h).



Bei den Konsonanten wird im Prinzip ähnlich vorgegangen. Allerdings beträgt die maximale Distanz zwei Punkte, was mit einer grösseren Vorkommenshäufigkeit der Konsonanten im deutschen Sprachsystem begründet wird.

### 2.3 Der Levenshtein-Approach

Ein weiterer Zugang zur Sprachdistanzmessung basiert auf dem Levenshtein-Algorithmus. Die wichtigsten Theoretiker zu diesem Algorithmus sind Heeringa und Narbonne für dialektologische Forschungen und Wichmann für die Klassifikation der Sprachen im Allgemeinen. (vgl. z.B. Nerbonne – Heeringa 1997, Heeringa – Nerbonne 2013, Nerbonne – Kretzschmar 2013).

Der Approach nach Levenshtein benötigt im Gegensatz zur Datenbehandlung nach Goebel keine Taxierung der Daten und erlaubt zudem einen Vergleich der Daten sowohl auf phonetischer Ebene als auch auf anderen linguistischen Ebenen. Wie wir noch aufzeigen werden, haben wir für unsere Studie trotz des Levenshtein-Algorithmus zusätzlich Taxationen vorgenommen und eine Analyse mit taxierten Daten durchgeführt. In einem gewissen Sinne stellt die Salzburger Schule nach Goebel eine Weiterentwicklung der historischen Dialektologie mit den Mitteln der Datenverarbeitung dar. Die Anwendung des Levenshtein-Algorithmus hingegen fusst eher auf der typologischen Linguistik und der Glottochronologie von Swadesh (1971). Unsere Studie nimmt beide Ansätze auf. Wir gehen von einem erweiterten Verständnis der Dialektometrie aus und versuchen, mit dem Levenshtein-Algorithmus die Daten möglichst effizient zu verarbeiten, um zu strukturellen Ergebnissen zu gelangen.

## 3. Was ist sprachliche Distanz?

### 3.1 Sprachliche Distanz

Der Titel dieses Aufsatzes kündigt Aussagen zur “Sprachdistanzmessung” an. Unter Distanz versteht man gemeinhin die Entfernung zwischen zwei Punkten. In den einschlägigen Wörterbüchern ist der Begriff meist nicht genauer definiert, wird allenfalls durch das Synonym “Abstand” erklärt. Hilfreich erscheint uns die Erklärung aus einer Anleitung zum Statistik-Softwarepaket SPSS:

“Sowohl Distanz- als auch Ähnlichkeitsmaße dienen dazu, die Ähnlichkeit verschiedener Fälle oder Variablen zu quantifizieren. Beide Maße untersuchen, wie nahe die Werte zweier Fälle oder Variablen beieinanderliegen. Eine starke Ähnlichkeit zweier Fälle oder Variablen kommt in großen Werten eines Ähnlichkeitsmaßes und in kleinen Werten eines Distanzmaßes zum Ausdruck. Distanzmaßen liegt damit im Vergleich zu Ähnlichkeitsmaßen die entgegengesetzte Perspektive zugrunde, da sie zunächst nicht

versuchen, die Ähnlichkeit, sondern die Unähnlichkeit zweier Fälle oder Variablen zu quantifizieren. Aus diesem Grund werden Distanzmaße häufig auch als Unähnlichkeitsmaße bezeichnet. Es gibt mehrere statistische Prozeduren, in denen Distanz- und Ähnlichkeitsmaße verwendet werden, um eine größere Anzahl von Objekten (Fälle oder Variablen der Datendatei) derart in Gruppen zu unterteilen, daß innerhalb einer Gruppe Objekte mit besonders großer Ähnlichkeit zusammengefaßt werden, während sich die verschiedenen Gruppen möglichst deutlich voneinander unterscheiden" (Brosius 1998: Kapitel 27.1).

Diese Erklärung passt zu unseren Vorstellungen von Distanz, weil sie einerseits aufzeigt, dass Distanz immer auch mit Ähnlichkeit zusammenhängt, und andererseits weist sie darauf hin, dass mehrere Objekte auch zu Gruppen zusammengefasst werden können und es auch Distanz zwischen Gruppen geben kann. In unserem Fall messen wir die linguistischen Varietäten, die sowohl aus einzelnen Informantinnen und Informanten als auch aus Ortschaften (gleichzeitig oder zu verschiedenen Zeiten) bestehen können. Die im Vergleich stehenden Variablen bestehen in unserem Falle aus Wörtern in ihrer phonetischen und morphologischen Struktur.

Zur Messung der Distanz haben wir zunächst eine Berechnung vorgenommen, die wir als „Rein phonetische Messung“ bezeichnen. Auf der anderen Seite haben wir für die typologisierten Formen eine „Typologische Messung“ angestellt.

### 3.2 Sprachliche Distanz als Levenshtein-Distanz (LD)

Wie oben dargestellt, stellt der Levenshtein-Algorithmus den Ausgangspunkt aller unserer Berechnungen dar. Das Verfahren wird heute in verschiedenen Variationen in der Computertechnologie als weit verbreiteter Algorithmus zum Vergleich von (Text)strings eingesetzt. Die Levenshtein-Distanz gibt an, wie viele Einfüge-, Lösch- und Ersetzoperationen minimal notwendig sind, um eine Zeichenkette in eine andere Zeichenkette zu überführen. Beim Minimalpaar *Kind* : *Rind* 'Kind, Rind' beispielsweise ist eine Veränderung (die Veränderungen werden oft auch als *costs*, *Kosten* bezeichnet) nötig, um das erste Wort in das zweite zu überführen. Beim Paar *Bruader* : *Brüadera* 'Bruder, Brüder' sind zwei Änderungen nötig. Beim Paar *ksäha* : *ksee* 'gesehen, gesehen' braucht es drei Änderungen.

Vom Algorithmus der Levenshtein-Distanz ausgehend, hat sich die Damerau-Levenshtein-Distanz entwickelt, die gegenüber der LD eine erweiterte Funktionalität aufweist. Wenn beispielsweise im Lexem 'Frauen' in den alemannischen Dialekten die Varianten *Froue* und *Fruoe* auftauchen würden, erkennt die Damerau-Levenshtein-Distanz die vertauschten Zeichen und zählt die Distanz nur einfach statt doppelt. Für die Datenbank setzen wir die Datenbanksoftware FileMaker12 pro ein, für die für die Damerau-Levenshtein-Distanzberechnung ein öffentlicher Algorithmus besteht (<<http://www.briandunning.com/cf/965>>).

Für die Distanzberechnung wird folgendes Verfahren angewendet, das man sich bildlich so vorstellen kann. Die zu vergleichenden Wörter werden so untereinander platziert, dass die sich entsprechenden Segmente eines Wortes sich genau gegenüberstehen. Diese Gegenüberstellung und Ausrichtung wird in der Regel *Alignment* genannt. Für zwei Varianten von alemannischem ‘Fensterläden’ sieht es so aus:

```
[ f e n t s t ər l ɛ : d a ]
[ f ɛ n s t ər l ɛ d ə ]
  1 1 1 1 1
absolute Levenshtein-Distanz aLD= 5
```

Wenn man auf diese Weise zwei Wörter vergleicht, zeigt sich, dass bei längeren Strings die Distanz grösser ist als bei kürzeren Strings, denn je länger das Wort ist, desto grösser ist die Wahrscheinlichkeit, dass in den Dialekten (oder allgemein in der gesprochenen Sprache) Unterschiede auftauchen. Da diese Feststellung der Idee widerspricht, dass Lexeme unabhängig ihrer Länge linguistische Einheiten darstellen, wird die Levenshtein-Distanz meist durch die Länge der Strings dividiert. Im Beispiel oben ergibt sich also ein Wert von  $13 / 5 = 2.5$ . In einem Beispiel, das wir weiter unten wieder aufnehmen, sehen die Werte so aus:

```
[gaj ə] ‘gegangen’
[kajka] ‘gegangen’      Levenshtein-Distanz LD 3 / 5 = 0.6
```

Die Reine<sup>4</sup> Levenshtein-Distanz funktioniert auch bei Zahlenketten.

### 3.3 Gewichtete Levenshtein-Distanz (WLD)

Für unser Forschungsprojekt gehen wir von einer gewichteten Levenshtein-Distanz aus. Die Gewichtung selber kann auf unterschiedlichen Ebenen vorgenommen werden. Meist wird sogar jeder einzelne Laut gewichtet. Gemeinhin bezeichnet man dieses Vorgehen als gewichtete Levenshtein-Distanz, *Weighted Levenshtein Distance* oder kurz WLD-Algorithmus (Wichmann – Grant 2010, Wichmann – Holman – Bakker - Brown 2010, Wichmann 2013).

<sup>4</sup> Unter “reiner” Levenshtein-Distanz verstehen wir die Distanz zwischen *beliebigen* Zeichen. In der Folge werden wir nicht *beliebige*, sondern verschiedene *phonetische* und *typologische* Zeichenketten verwenden und berechnen.

Der Vergleich zwischen linguistischen Varietäten findet in der Regel auf der Basis mehrerer Wörter statt, d.h. man berechnet für jedes Wort eine eigene WLD und präsentiert am Schluss den Mittelwert der berechneten Distanzen, die Skalierung reicht dabei von 0 bis 1, wobei 0 aussagt, dass die verglichenen Formen identisch sind, 1 hingegen bezeichnet die grösstmögliche Distanz zwischen den Formen.

Wir sind der Ansicht, dass ein solcher gewichteter Vergleich nur dort statthaft ist, wo die einzelnen "Zeichen" zweier Strings aus verschiedenen Sprachen denselben phonetischen Wert haben. Man kann also nicht zwei Sprachen mit verschiedenen "Orthographien" vergleichen. Aber auch wenn man Transkriptionen zweier verschiedener phonetischer Texte (aus verschiedenen Sprachen) vergleicht, berücksichtigt der Levenshtein-Algorithmus die sprachinhärenten Unterschiede der betrachteten phonetischen Systeme nicht. [r] und [R] werden von Algorithmus unterschiedler, obwohl es sich in vielen Sprachen nur um Allophone handelt.

Die Levenshtein-Distanz berücksichtigt offensichtlich auch die Nähe und Ähnlichkeit von Lauten nicht: [a] ist gleichermassen unterschiedlich von [p] wie [p] von [b]. In der Realität wissen wir, dass beispielsweise [p] und [b] sich hauptsächlich durch das Merkmal Stimmhaftigkeit/Stimmlosigkeit unterscheiden, aber denselben Artikulationsort aufweisen und zu den Explosiva gehören. [p] und [z] hingegen unterscheiden sich durch die Merkmale Stimmhaftigkeit, Artikulationsort und Artikulationsweise. Um diese Abstufungen in der Distanz zu berücksichtigen, kann man die phonetischen Zeichen durch eine Matrix (*feature matrix*) ersetzen, die verschiedene, eindeutige distinktive Merkmale enthält. Heeringa und Bolognesi (Bolognesi – Heeringa 2002) gehen sogar so weit, dass sie einen *Codierstring* verwenden, der Parameter aus dem Spektrogramm der einzelnen Laute berücksichtigt.

Für unsere Studie haben wir hingegen jedem Phon einen dreichiffrigen Code zugeordnet, der die Artikulationsart, den Artikulationsort und die Stimmhaftigkeit definiert. Zusätzlich können wir mit einer oder zwei zusätzlichen Chiffren Suprasegmentalia, Längen, Silbenakzente, Nasalierungen, Aspirationen, Palatalisierungen und für die alemannischen Dialekte auch Lenisierungen festhalten. Die folgenden Tabellen zeigen auf, wie die Codierungen konzeptionell und im Einzelnen aussehen. Wichtig ist, dass jedem Symbol ein eigenständiger Wert zugeteilt ist.

## 3.3.1 Konzept für die Codierungen

*Vokale*

Bildungsort der Vokale

8 = hinten  
 0 = zentral  
 9 = vorne

Öffnungsgrade der Vokale

1 bis 7 = 1=offen, 7= geschlossen

Rundung der Vokale

) = gerundet  
 ] = ungerundet

*Konsonanten*

Artikulationsart

F = Frikativ  
 I = Flap  
 J = Approximant  
 L = lateral Approximant  
 M = lateral Frikativ  
 N = Nasal  
 R = Vibrant  
 S = Plosiv  
 Z = Affrikat

Artikulationsort

A = alveolar  
 E = post-alveolar  
 D = dental  
 U = uvular  
 T = retroflex  
 P = bilabial  
 K = velar  
 H = glottal  
 C = palatal  
 B = labio-dental  
 Q = pharyngal

Stimmhaftigkeit

X = stimmhaft  
 Y = stimmlos

Suprasegmentalia und Diverses

h = aspiriert  
 ‘ = betont  
 : = lang  
 \$ = lenisiert  
 ~ = nasaliert

## 3.3.2 Die Codierungen für die IPA-Symbole

sound	code	sound	code	sound	code	sound	code
ɹ	ɹ	ɣ	92)	β	FPX	m	NPX
ɹ	ɹ	ɹ	92]	ϕ	FPY	ŋ	NTX
ɹ	ɹ	ø	93)	ɸ	FQX	r	RDX
~	~	e	93]	ħ	FQY	ʃ	SCX
	—	œ	95)	z	FTX	ʒ	SCX
:	:	ε	95]	ʂ	FTY	c	SCY
ʈ	01)	æ	96]	ɸ	FUX	t	SDY
ɨ	01]	œ	97)	χ	FUY	d	SDY\$
ɵ	03)	a	97]	ʃ	FWY	ʔ	SHY
ə	03]	ɯ	JKX	h	h	k	SKY
ə	04]	ɹ	JTX	r	IDX	g	SKY\$
ɵ	05)	w	JWX	ɹ	ITX	p	SPY
ɜ	05]	z	FAX	ɹ	IUX	b	SPY\$
ɐ	06]	s	FAY	ʊ	JBX	d	STX
U	81)	v	FBX	j	JCX	t	STY
u	81]	f	FBY	ɹ	JDX	g	SUX
U	82)	ç	FCY	ʌ	LCX	q	SUY
O	83)	ð	FDX	l	LDX	w	w
ɣ	83]	θ	FDY	L	LKX	x	x
o	85)	ʒ	FEX	l	LTX	ɟ	ZDX
ʌ	85]	ʃ	FEY	ʒ	MDX	ʂ	ZDY
ɒ	87)	fi	FHX	ɸ	MDY	ɟ	ZEX
ɑ	87]	h	FHY	ŋ	NBX	ʃ	ZEY
Y	91)	ɣ	FKX	ɹ	NCX		
I	91]	x	FKY	n	NDX		

Durch die dreistellige Codierung ergeben sich somit die Möglichkeiten einer totalen Verschiedenheit, einer 2/3-Verschiedenheit, einer 1/3-Verschiedenheit und einer Identität von Lauten; und für die vierstellige Codierungen analog 3/4, 2/4 usw. Die Umformung der IPA-Symbole in phonetische Merkmale verfeinert die

Distanzberechnung, vermindert aber auch die Distanz zwischen den verglichenen Elementen.

[gaŋ ə] ‘gegangen’					
SKY\$ 97] NKX			04]		
[kaŋka] ‘gegangen’					
SKY 97] NKX	SKX 97]	LD		$3 / 5 = 0.6$	
				WLD	$6 / 15 = 0.4$

Es ist nun so, dass einige Parameter weniger “Optionen” haben als andere, weshalb sie frequenter und weniger distinktiv sind (stimmhaft/stimmlos).

### 3.3.3 “Spezifische Differenz” (SD) und am Korpus gewichtete Levenshtein-Distanz (DLWC)

Um diesem Problem entgegenzutreten, haben wir für die Distanzmessung eine weitere Gewichtung vorgenommen, indem wir die “spezifische Differenz SD”<sup>5</sup> (und damit verbunden auch die “spezifische Ähnlichkeit”) mitberücksichtigt haben. In Anlehnung an Wichman haben wir einen “internen Differenzindex” berechnet, d.h.: wir haben für eine Serie von beliebigen Wörtern aus unserem Korpus die DLWC-Distanz<sup>6</sup> errechnet. Konkret heisst das, dass wir 42 zufällige Wörter verschiedener Bedeutung aus dem 120 Formen umfassenden Questionnaire erfasst haben, drei für jeden Untersuchungsort. In der Folge haben wir die Distanz zwischen jedem dieser Wörter mit allen anderen berechnet. Den Mittelwert all dieser Berechnungen bezeichnen wir als SD. Um schliesslich den DLWC-Wert zu berechnen, haben wir die nicht gewichteten Werte durch die SD dividiert.

Im Falle des Dialektsystems/der Dialektsysteme im alemannischen Untersuchungsgebiet im Churer Rheintal beträgt die “Spezifische Differenz” 0.73, die Spezifische Ähnlichkeit 0.27. Das gewichtete Resultat lautet nun für das folgende Beispiel:

<sup>5</sup> Der Begriff “spezifische Differenz (von Sprache)” ist in einem gewissen Sinne dem “spezifischen Gewicht” von Materie vergleichbar, indem er als Universalgrösse die Differenz von beliebigen Wörtern einer Sprache oder auch aller Sprachen untereinander ausdrückt. In unserem Falle haben wir die spezifische Differenz nur für die alemannischen Wörter aus unserem Korpus berechnet.

<sup>6</sup> DLWC = Distance Levenshtein Wighted on Corpus

[gaŋ ə] 'gegangen'					
SKY\$ 97]	NKX		04]		
[kaŋka] 'gegangen'					
SKY 97]	NKX	SKX 97]	LD	$3 / 5 = 0.6$	
				WLD $6 / 15 = 0.4$	
				DLWC $0.4 / 0.73 = 0.55$	

Diese Art von Berechnung kann nur in einem System gemacht werden, das aus einem Erhebungspunkt oder einem Gebiet mit einem relativ einheitlichen Dialektkontinuum besteht. Je differenzierter ein System in seinem Inneren ist, umso höher ist konsequenterweise die SD. Entsprechend geringer ist der statistische Wert der resultierenden Ergebnisse und damit auch der linguistischen Aussagekraft.

Die obige Aussage gilt für jedes Wortpaar, das miteinander verglichen wird. Die Distanz zwischen zwei Varietäten (d.h. zwischen zwei Punkten, deren Distanz man berechnen will) wird am Schluss der dem Mittelwert der einzelnen gewichteten DLWC entsprechen. Die SD ist tendenziell in Systemen höher, die längere Phonketten und/oder längere Wörter aufweisen. (Vgl. Wichmann 2013).

Natürlich berechnen wir sämtliche oben vorgestellten Formen von Distanz, LD, DLW und DLWC, dies einerseits um die Ergebnisse miteinander vergleichen zu können und andererseits, um die Validität der Ergebnisse besser kontrollieren zu können (Bolognesi – Heeringa 2002).

#### 4. Bericht aus der Werkstatt

In der Praxis ging die Arbeit folgendermassen vonstatten: Für die Berechnung der Distanz zwischen den einzelnen Dialekten mussten die Daten unserer Datenbank zunächst so aufbereitet werden, dass sie sinnvoll miteinander verglichen werden können. Wir erarbeiteten sogenannte „Vergleichsformen“, die wir in der IPA-Transkription festgehalten haben.

- Für die SDS-Lautungen bedeutet das, dass wir die nur in Diethscher Umschrift vorliegenden Formen ins IPA-System umformen mussten. Ebenso hatten wir die in enger Umschrift vorliegenden Informationen aus dem SDS in das IPA-System zu transliterieren.
- Auf phonetischer Ebene galt es, gewisse Assimilationen aufzuheben, um vergleichbare Formen zu gewinnen. Obwohl das Berechnungssystem auch komplexere Fälle berechnen kann, haben wir uns entschieden, einige nicht phonologische Besonderheiten ([p<sup>h</sup>nø:] = [pnø:] 'Pneu') und Sandhi-Phänomene (Liaisonen) anzupassen ([si heŋk < >k<sup>h</sup>ai ...] 'sie haben keine' => [si hend] 'sie haben'). Oder wir mussten z.B. Inversionen auflösen und Per-

- sonalpronomen weglassen, um die Qualität des betonten Vokals genauer zu untersuchen. Beispiel ‘wir gehen’: [mɪər gø:nd], [mɪrɐr gø:nd], [mər gø:nd], [gø:mər]= [gø:nd]
- Auf morphosyntaktischer Ebene haben wir beispielsweise bei der Pluralbildung von ‘Bruder’ die Formen auf das reduziert, was die Pluralbildung kennzeichnet: Umlaut im Diphthong im Singular und Plural, Pluralmarker *-re* am Wortende. Varianten im betonten Diphthong oder im Nebensilbenvokalismus haben wir vereinheitlicht.

Um die Berechnungen homogener zu machen, haben wir auch bei Mehrfachnennungen bei den Informanten und im SDS bestimmte Formen ausgeschlossen. Bei Doppelformen eines Informanten oder des SDS haben wir die ältere Form beibehalten und die jüngere gestrichen. Bei Mehrfachformen haben wir entweder die älteste Form ausgewählt oder uns für die dominierende Form entschieden. Dieses Ausschlussverfahren reduziert zwar die sprachliche Wirklichkeit, ist aber bei Typisierungen von sprachlichen Gegebenheiten verbreitet. Auch schon bei der Erarbeitung der SDS-Karten mussten sich die Autoren oft gegen die Variantenvielfalt entscheiden. Im Grunde genommen haben wir mit dem Ausschluss der Doppelformen eine Art Basisdialekt konstruiert, den wir mit den SDS-Formen und den Formen der anderen Dialekte vergleichen.

Neben der rein phonetischen Distanzberechnung führen wir auch eine Distanzberechnung aufgrund verschiedener Klassifikationen aus, die wir vereinfacht als “typologische Distanzberechnung” bezeichnen. Es muss hier natürlich festgehalten werden, dass auch die phonetische Distanzberechnung grammatikalische und morphosyntaktische Elemente miteinbezieht (vgl. oben Unterschiede zwischen Bolognesi –Heeringa 2002 und Goebel 1981, Goebel 2002). Die zusätzlichen Berechnungen dienen dazu, die Differenzen in den einzelnen sprachlichen Kategorien Morphologie, Lexik und Phonologie vertieft auszuleuchten. D.h. dass wir entweder aus dem Primärmaterial oder den „Vergleichsformen“ weitere abstrahierte Informationen gewonnen haben, die wir als numerische Ketten codiert haben. – Beispiel: Die verschiedenen Antworten zur Konjunktivbildung haben wir gegliedert in die Typen: Infinitiv, Konjunktiv I, Konjunktiv II, würde-Konjunktiv, tun-Konjunktiv. Eine phonetische Distanzberechnung ergäbe hier wenig Sinn.

Aus dem Grundsatz, dass nur Vergleichbares miteinander verglichen werden soll, lässt sich leicht der Schluss ziehen, dass lexikalische Phänomene unter Umständen sinnvollerweise nicht mit einer phonetischen Distanzmessung verglichen werden sollen. Wenn beispielsweise für das Standarddeutsche ‘Fensterläden’ in den alemannischen Dialekten ein Wort wie *Pälgga* dem standardnahen *Fensterläda* gegenübersteht, ergibt sich allein schon aus der Wortlänge eine grosse phonetische Distanz zwischen den beiden Lexemen. Diese grosse Distanz über-

deckt aber die feinen Messungen, die z.B. die Vokalqualitäten in *Pälgga/Pelgga* festhalten. Aus diesem Grund haben wir für die lexikalische Distanz ein typologisches, eigenes Messgefäß eingerichtet.

Die oben dargestellte Problematik, dass wir Daten vereinheitlichen und typologisieren mussten und gewisse Daten bei gewissen Berechnungen nicht berücksichtigen konnten, stellt ein allgemeines Problem bei der Bearbeitung sprachlicher Daten dar. Roland Bauer schreibt dazu treffend:

“Wie jede Klassifikation, verdichtet auch die DM [Dialektometrie] durch *Datenreduktion*. Dies bedeutet, dass die Rohdaten im Rahmen der Auswertung [...] *simplifiziert* und *transformiert* werden, wobei es einerseits zu *Informationsverlusten* kommt, die aber andererseits im Interesse einer *höherrangigen* Erkenntnis (Mustererkennung) bewusst in Kauf genommen werden” (Bauer 2009, S. 11).

Der Unterschied in den Berechnungsmethoden besteht darin, dass wir bei der rein phonetischen Distanzberechnung im Prinzip alle transkribierten Formen einer bestimmten Ortschaft oder eines bestimmten Informanten mit den transkribierten Formen einer anderen Ortschaft oder eines Informanten vergleichen können. Konkret heisst das zum Beispiel, dass wir alle Formen unserer Informanten von Jenins mit den SDS-Formen von Jenins kreuzen, oder dass wir alle aktuellen Formen unserer Informanten von Jenins mit jenen von Chur vergleichen. Wie wir oben ausgeführt haben, ist es nicht sehr sinnvoll, wenn man auch lexikalische Formen in einen phonetischen Vergleich miteinbezieht. Dem Lexem *Pälgga* ‘Fensterläden’ stehen beispielsweise *Fenschterlääda*, *Prittli* und fälschlicherweise auch *Schtorra* gegenüber. Die Werte aus einem phonetischen Vergleich dieser Lexeme wären viel zu hoch und würden die feineren Differenzierungen, z. B. offene Vokalqualität vs. geschlossene Qualität, in den Hintergrund treten lassen.

Bei der nicht-phonetischen Distanzberechnung hingegen gehen wir ganz anders vor. Bereits bei der Auswahl der Marker sind wir davon ausgegangen, dass wir gewisse Formen auf bestimmte Sprachwandelerscheinungen hin zu untersuchen beabsichtigten. Beispielsweise wollten wir feststellen, ob der ältere Plural *Maana* ‘Mannen, Männer’ sich gegenüber dem neueren Plural *Mäner* ‘Männer’ halten konnte. Entsprechend haben wir dieses Unterscheidungsmerkmal als Hauptmerkmal fixiert und dreifach gewichtet. Bei der Analyse des effektiv vorhandenen Sprachmaterials haben wir festgestellt, dass parallel die Formen *Maana* und *Mana* bzw. *Määner* und *Mäner* auftreten. Hier haben wir entsprechend die Vokalqualität als zweitwichtigstes Unterscheidungsmerkmal definiert und doppelt gewichtet. Und als drittes Merkmal haben wir schliesslich bei *Mäner* / *Mener* die Vokalqualität als Unterscheidungskriterium fixiert und einfach gewichtet.

Im Codierungs-Formular sieht das folgendermassen aus:

[mɛnər] ‘Männer’      Klassifikation: Gramm

(typ) Form	Hauptmerkmal	Marker 2	Marker 3
	0 leer/nicht relevant 1 Plural ‘Mannen’ 2 Plural ‘Männer’	0 leer/nicht relevant 1 langer Haupttonvokal 2 kurzer Haupttonvokal	0 leer/nicht relevant 1 geschlossener Haupttonvokal 2 offener Haupttonvokal
typ ma:na	1	1	0
typ mana	1	2	0
typ menər	2	2	1
typ mɛ:nər	2	1	2
typ mɛn:ər	2	2	2
typ mɛnər	2	2	2

Die Levenshtein-Distanz funktioniert nach der Booleschen Logik, d.h. sie unterscheidet nur zwischen *wahr* und *unwahr*: der Unterschied in der Tabelle zwischen den Ziffern 0 und 2 ist 1, d.h. wahr, die Ähnlichkeit 0.

Während es sich in diesem Fall bei der Unterscheidung des Hauptmerkmals um ein morphologisches Phänomen handelt und bei den beiden anderen um phonetische Erscheinungen, kann in anderen Fällen das phonetische Kriterium zugleich ein morphologisches sein, etwa wenn es sich um Ablaute oder Umlaute handelt. Durch die Gewichtung im Verhältnis 3:2:1 aber haben in der Distanzberechnung strukturelle Unterscheidungen mehr Bedeutung als phonetische. Gewichtet haben wir, indem wir das Hauptmerkmal dreifach, Marker 2 doppelt und Marker 3 einfach notieren.

Mit dieser Methode gelingt es uns mit einem drei- bzw. gewichtet sechsstelligen Code die meisten Formen von einander zu unterscheiden und die Strukturunterschiede der Formen herauszuholen, ohne uns in terminologisch-methodische Abgrenzungsdiskussionen zu verstricken.

Die Distanzberechnung erfolgt auch hier nach dem Levenshtein-Algorithmus, indem wir berechnen, wie viele Veränderungen nötig sind, um vom String x zum String y zu gelangen.

Form typologisiert	Code	Code gewichtet	Differenz 1. Position	Differenz 2. Position	Differenz 3. Position	absolute Differenz total
<i>mana</i>	120	111 22 0				
<i>menər</i>	221	222 22 1	3	0	1	4

Diese absolute Differenz dividieren wir anschliessend durch die Anzahl Zeichen im Code, so dass im Beispiel der Dividend  $4:6=0.6666$  ausmacht.

Zusammenfassend halten wir fest: Die rein phonetische Distanzberechnung und die typologische Messung helfen uns, diverse Aspekte der sprachlichen Veränderungen darzustellen und hervorzuheben. Die folgende Tabelle stellt die wesentlichen Aspekte der beiden von uns verwendeten Messmethoden dar.

Messungen und Messmethoden		
Benennung	Rein phonetische Messung	Typologische Messung
Typologie	keine Typologie	Typologie auf Grundlage einer auf den Fall zugeschnittenen Matrix, die sich aus der Fragestellung und aus der Differenzierungsmöglichkeit des Sprachmaterials ergibt. Keine Matrix, die grundlegende distinktive Merkmale bereit hält und ein uniformes Codieren ermöglichen würde.
Gewichtung bei der Datenerfassung	- keine Gewichtung - systembedingte Gewichtung durch Transkription in IPA	- Unterscheidung in phonetische, morphosyntaktische und lexikalische Fragestellung - Die Hauptfragestellung wird dreifach gewichtet, die zweite Fragestellung doppelt und die dritte Fragestellung einfach.
Reinheit der Messung	- rein phonetische Messung - implizit auch morphosyntaktische Messung - Ausschluss von Formen, die die Messresultate verfälschen würden, v.a. lexikalische Mehrfachnennungen	- Die Hauptfragestellung ist immer einer der drei Kategorien Phonetik, Morphosyntax und Lexik zugeordnet. - Die anderen Codierungen dienen in der Regel der Differenzierung des Sprachmaterials, beziehen sich auf verschiedene Ebenen.
Gewichtung bei der Distanzberechnung	- die Distanz zweier nahe liegender Phone wird weniger gewichtet als die weiter auseinander liegender Phone. - Der errechnete Wert wird durch den Wert der spezifischen Distanz 0.73 dividiert.	- keine zusätzliche Gewichtung
Ziel, linguistische Fragestellung	gewichtete phonetische Distanz im (fast) vollständigen Korpus	gewichtete Distanz in den Kategorien Phonetik, Morphosyntax und Lexik aufgrund ausgewählter typologischer Kriterien

## 5. Beispiele

### 5.1 Vorbemerkungen

Wir präsentieren hier zwei Beispiele aus unseren Distanzberechnungen. Referenzpunkt ist die Stadt Chur, welche als regionales Zentrum verschiedene Dienstleistungsfunktionen inne hat und in der fast alle weiterführenden Schulen domiziliert sind.

Für Chur haben wir zunächst die für die Studie relevanten historischen Daten aus dem SDS extrahiert und diese in die Datenbank eingefügt. Das sind insgesamt rund 150 Datensätze. Anschliessend haben wir mit 10 Jugendlichen und jungen Erwachsenen Interviews geführt, diese transkribiert und ebenfalls in die Datenbank eingefügt. Im ersten Vergleich Chur aktuell // Chur SDS stehen rund 1500 aktuelle Datensätze rund 150 SDS-Datensätzen gegenüber.

Genau dasselbe Vorgehen trifft für die beiden Ortschaften Churwalden und Felsberg zu.

In einer zweiten Messung haben wir den Vergleich Churwalden aktuell // Churwalden SDS bzw. Felsberg aktuell // Felsberg SDS vorgenommen, in der wiederum 1500 aktuelle Datensätze 150 historischen SDS-Datensätzen gegenüberstehen.

Die dritte Messung stellt den Bezug zwischen Churwalden SDS bzw. Felsberg SDS und Chur SDS her. Hier stehen sich jeweils 150 Datensätze gegenüber.

Und in der vierten Messung vergleichen wir die Daten von Churwalden aktuell bzw. Felsberg aktuell mit den Daten von Chur aktuell. Hier berechnen wir die Distanz, indem wir je die 1500 Datensätze der beiden Ortschaften mit jenen 1500 von Chur vergleichen.

Wir führen diese Zahlen hier explizit aus, weil die Streuung natürlich grösser wird, je mehr Datensätze man miteinander kreuzt. 150 SDS-Datensätze verglichen mit weiteren 150 SDS-Datensätzen müssen eine geringere Differenz ergeben als 1500 aktuelle Datensätze verglichen mit 1500 weiteren aktuellen Datensätzen.

Die Ortschaft Churwalden liegt rund 11 Autokilometer von Chur entfernt und zeigt im SDS noch einige ausgeprägte Formen eines Walser-Dialektes auf. Felsberg ist mit dem Auto von Chur aus nach sechs Kilometern Fahrt zu erreichen. Aufgrund der Lage links vom Rhein lag Felsberg früher abseits der Verkehrswege und entwickelte einige sprachliche Besonderheiten, so etwa konsequente Vokalentrundungen, die heute nicht mehr nachweisbar sind.

Für die Darstellung in den Grafiken haben wir die Berechnungsdaten aus der Datenbank mit dem Faktor 100 multipliziert, um ganzzahlige Werte zu bekommen.

## 5.2 Beispiel Churwalden – Chur

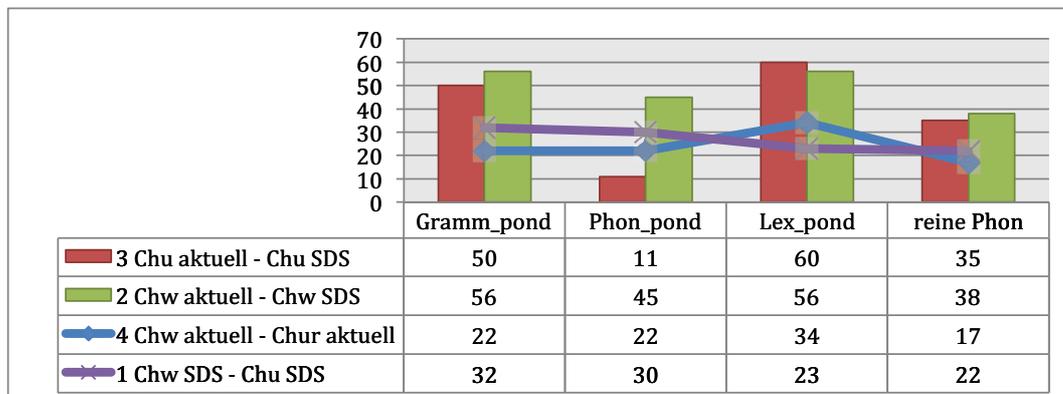


Abbildung 5: Beispiel Churwalden-Chur

Wir unterscheiden in unserer Darstellung zwischen einer Balkendiagramm-Grafik und einer Kurven-Grafik. Die Balkendiagramme zeigen die historische Distanz zwischen früher und heute auf. Die Kurvendiagramme halten die areale Distanz zwischen den Ortschaften wiederum in den Dimensionen früher und heute fest.

In den Balkendiagrammen erkennen wir in der ersten Spalte *Gramm\_pond*, dass die Veränderungen von früher zu heute im Bereich der Morphosyntax relativ hoch sind, dies sowohl in Chur als auch in Churwalden. Offensichtlich hat hier ein bedeutender Sprachwandel stattgefunden. Der relativ hohe Wert ist aber auch darauf zurückzuführen, dass unsere Untersuchung darauf angelegt war, Sprachwandelerscheinungen festzuhalten.

Die zweite Spalte *Phon\_pond* hält die Veränderung im Bereich der Phonetik fest. Gegenüber Chur stellen wir für Churwalden einen bedeutend höheren Wert fest.

Für die dritte Spalte *Lex\_pond* wurde der Wandel in der Lexik berechnet. Hier registrieren wir für beide Ortschaften – erwartungsgemäss – einen relativ hohen Wandel.

Die vierte Spalte schliesslich gibt den Wert der “reinen Phonetik” an. Für diese Daten haben wir das grösste Sample für den Vergleich. Bei den Werten ist zu berücksichtigen, dass die Werte mit der spezifischen Distanz gewichtet wurden und deshalb schon strukturell tiefer sind als bei den vorhergehenden Spalten. Das Beispiel bestätigt wie in Spalte 1, dass sich beide Ortschaften in etwa gleich stark vom SDS-Basisdialekt in Richtung aktuelle Mundart verändert haben.

In der ersten Spalte der Kurvengrafik stellen wir fest, dass die Distanz im Bereich der Morphosyntax zwischen Churwalden und Chur zur Zeit des SDS markant höher war als heute. Hier haben sich die Dialekte im Lauf der Zeit offensichtlich angenähert. Dasselbe trifft auch für die Bereiche Phonetik und reine

Phonetik zu. Im Bereich der Lexik hingegen ist die Distanz kleiner geworden. Der Dialekt von Chur hat sich offenbar stärker verändert als der von Churwalden, wodurch der Churwaldner Dialekt konservativer wirkt.

Insgesamt können wir festhalten, dass der Dialekt von Churwalden mit einigen salienten Merkmalen sich offensichtlich relativ stark gewandelt und sich dem Dialekt von Chur angenähert hat.

### 5.3 Beispiel Felsberg – Chur

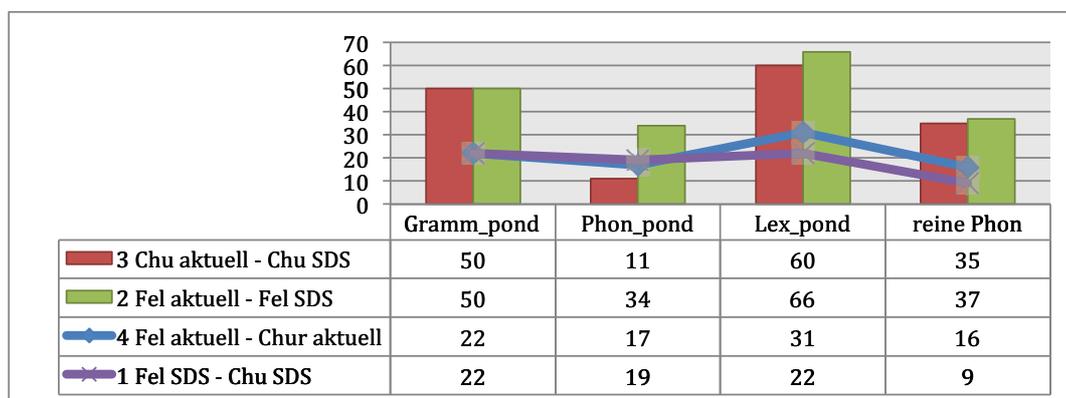


Abbildung 6: Beispiel Felsberg-Chur

Im Vergleich zum obigen Beispiel weist der Dialekt von Felsberg einige Besonderheiten auf. Im Bereich der Morphosyntax (Gramm\_pond) weisen die Dialekte von Felsberg und Chur eine genau gleich grosse Distanz auf zu den Formen des SDS. Und parallel dazu zeigen die beiden Ortschaften auch in der arealen Distanz den genau gleichen Wert auf. Beide Dialekte haben sich also gleich viel von ihren Basisdialekten entfernt, die Distanz zwischen den Dialekten ist gleich geblieben. Der saliente Churwaldner Dialekt hingegen hat an dieser Position sich der Sprache der Hauptstadt und der Region angeglichen.

Im Bereich der Phonetik (Phon\_pond) hat sich der Churer Dialekt weniger verändert als der Felsberger Dialekt. Die Distanz zwischen den beiden Dialekten ist fast gleich geblieben. Das bedeutet, dass sich der Dialekt von Felsberg dem Churer Dialekt angeglichen haben muss.

Im Bereich der Lexik (Lex\_pond) hat sich der Dialekt von Felsberg tendenziell mehr dem Churer Dialekt genähert als der Churwaldner Dialekt. Die Distanzwerte zwischen den Dialekten sind nahezu identisch.

Und in der letzten Spalte schliesslich fällt auf, dass in beiden Grafiken die Werte sich fast entsprechen, ausser dass der Wert der Distanz Fel SDS – Chu SDS bei 9 liegt (Churwalden 22). Offensichtlich waren sich die Dialekte von Felsberg und Chur früher ähnlicher als heute. Berücksichtigt man aber, dass die Streuung der Berechnung bei 1500 mal 1500 Datensätzen grösser ist als bei 150 mal 150

Datensätzen, kann man sagen, dass die beiden Dialekte immer noch relativ nahe bei einander liegen.

#### 5.4 Generalisierung der Aussagen

Es ist uns ein Anliegen festzuhalten, dass die vorliegende Interpretation der Grafiken vorläufigen Charakter haben. Wichtig ist, dass die Daten nur sinnvoll in einer Gesamtschau des ganzen Untersuchungsgebietes Aussagekraft besitzen. Die Daten werden nun in eine separate Datenbank aufgenommen und anschliessend noch genauer analysiert.

Um es nochmals zu wiederholen: Wir haben die Marker ausgewählt, mit der absicht, mit ihnen den Sprachwandel zu erfassen. Die Resultate unserer Studie sind deshalb also insgesamt drastischer, als wenn wir ein beliebiges Korpus erfasst hätten.

Dennoch können wir aufgrund der vorliegenden Resultate die folgenden Schlüsse ziehen.

A. Die Spalte reine Phonetik (reine Phon) repräsentiert am ehesten den allgemeinen sprachlichen Zustand, weil die Auswahl der Marker sich hier am wenigsten auswirkt.

B. Die Veränderungen in der zeitlichen Dimension sind tendenziell grösser als die Veränderungen in der räumlichen Dimension. D.h., die Veränderungen früher/heute sind alle relativ hoch, die Distanzen von Ort zu Ort sind geringer.

C. Die Distanz von Ort zu Ort verändert sich tendenziell wenig, trotz der grossen Veränderung innerhalb der einzelnen Orte.

D. Die Differenz/Distanz Phon\_pond zeigt die augenmerklichsten Unterschiede zwischen Chur und den einzelnen Ortschaften und zwar in dem Sinne, dass die einzelnen Ortschaften die phonetische/phonematische Struktur verändern, auch wenn Chur das nicht macht.

E. Die Spalte "reine Phonetik" repräsentiert vordergründig die phonetische Oberflächenstruktur aber auch die Morphologie, während phon\_pond die Synthese aus charakteristischen phonetischen Erscheinungen mit phonologischem Wert darstellt. Dies erlaubt die Schlussfolgerung, dass die Ortschaften rund um Chur zwar ihre sprachlichen Besonderheiten/Merkmale eher bewahrt haben, sich aber in der phonologischen Struktur soweit erneuert haben, dass sie sich dem Dialekt von Chur angenähert haben.

## Bibliographie

BAUER R. (2009), *Dialektometrische Einsichten. Sprachklassifikatorische Oberflächenmuster und Tiefenstrukturen im lombardo-venedischen Dialektraum und in der Rätioromania*, San Martin de Tor, Istitut Ladin Micurà de Rü. (= Ladinia monografica 01).

- BOLOGNESI, R., HEERINGA W. (2002), *De invloed van dominante talen op het lexicon en de fonologie van Sardische dialecten*, in BAKKER D., SANDERS T., SCHOONEN R., VAN DER WIJST P. (eds.) (2002), *Tijdschrift voor taalwetenschap*, 9 (1), Nijmegen University Press, Nijmegen, pp. 45-84.
- BROSIUS F. (1998), *SPSS 8: Professionelle Statistik unter Windows*, Bonn, International Thomson Publishing.
- DELL'AQUILA V. (2011), *GIS and sociolinguistics*, in LAMELI A., KEHREIN R., RABANUS S. *Language and space. An International Handbook of Linguistic Variation*, vol 2, Berlin & New York, Mouton De Gruyter, pp. 458-476.
- DELL'AQUILA V., IANNACCARO G. (2013) *Quelques considérations sur la cartographie des données linguistiques*, in «La Bretagne Linguistique», 17.
- GOEBL H. (1981), *Éléments d'analyse dialectométrique (avec application à l'ALS)*, «Revue de Linguistique Romane», 45, pp. 349-420.
- GOEBL H. (1982), *Dialektometrie. Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie*, Wien, Verlag der Öst. Akademie der Wissenschaften.
- GOEBL H. (2002), *Analyse dialectométrique des structures de profondeur de l'ALF*, in «Revue de Linguistique Romane», 66, pp. 5-63.
- HEERINGA W., NERBONNE J. (2013) *Dialectometry*, in HINSKENS F., TAELEMAN J. (eds.) (2013), *Language and Space. An International Handbook of Linguistic Variation*, Berlin & New York, Mouton de Gruyter, vol 3, pp.624-646.
- HOLMAN E.W., WICHMANN S., BROWN C.H., VELUPILLAI V., MÜLLER A., BAKKER D. (2008) *Explorations in automated language classification*, in «Folia Linguistica», 42.2, pp. 331-354.
- LAMELI A. (2004), *Standard und Substandard. Regionalismen im diachronen Längsschnitt*. Stuttgart, Franz Steiner Verlag. (= ZDL-Beiheft 128).
- LÉONARD J. L., DELL'AQUILA V. & GAILLARD-CORVAGLIA A. (2012), *The ALMaz (Atlas Lingüístico Mazateco): from geolinguistic data processing to typological traits*, in «STUF», 6, pp. 78 -94.
- NERBONNE J., HEERINGA W. (1997), *Measuring Dialect Distance Phonetically*, in COLEMAN J. (ed.), *Workshop on Computational Phonology, Special Interest Group of the Association for Computational Linguistics*, Madrid, pp. 11-18.
- NERBONNE J. & KRETZSCHMAR W. (2013), *Dialectometry++*, in «LLC: Journal of Digital Scholarship in the Humanities», 28(1), pp. 2-12.
- SDS = BAUMGARTNER H., HOTZENKÖCHERLE R. & AL. (1962-2003) *Sprachatlas der deutschen Schweiz*, 8 voll., Bern, Francke.
- SWADESH M. (1971), *The origin on diversification of language*. Chicago, Aldine.
- WICHMANN S. (2013), *Internal language classification*, in LURAGHI S., BUBENIK V. (eds.) «The Bloomsbury Companion to Historical Linguistics», pp. 70-86. Bloomsbury Publishing.
- WICHMANN S., GRANT A.P. (eds.) (2010), *Special Issue: Quantitative Approaches to Linguistic Diversity: Commemorating the Centenary of the Birth of Morris Swadesh*, «Diachronica» 27.2.
- WICHMANN S., HOLMAN E.W., BAKKER D. & BROWN C.H. (2010), *Evaluating linguistic distance measures*, in «Physica», 389, pp. 3632-3639.