

Dialektdynamik in Kalabrien

Statistisches Praktikum WS 08/09

Institut für Statistik
LMU München

Teilnehmer: Christian Doll, Marcus Groß, Daniel Krklec

Projektpartner: Prof. Dr. Thomas Krefeld, Dr. Stephan Lücke

Betreuer: Prof. Dr. Helmut Küchenhoff

Inhaltsverzeichnis

1. Einleitung.....	3
2. Studie und Design.....	4
3. Ziel der Untersuchung.....	6
4. Untersuchung des Dialektes und Bildung von Indikatoren.....	7
4.1 Ähnlichkeits-Score.....	7
4.2 Indices aus den vier Teilkategorien.....	9
4.2.1 Complementi verbali, Verwendung des Infinitivs.....	10
4.2.2 Frasi avverbiali, Verwendung des Infinitivs.....	12
4.2.3 Tempi verbali, Zeitformen.....	13
4.2.4 Proclisi/enclisi, Stellung des Possessivpronomens.....	15
5. Zusammenfassung erster Ergebnisse.....	16
6. Gesamtindex.....	17
7. Clusteranalyse.....	18
8. Untersuchung der Dialektdynamik.....	19
8.1 Ambiguo-Score.....	20
8.2 Redefinition des Ähnlichkeits-Scores.....	21
8.3 Kategorien-Gesamtindex.....	23
8.4 Gesamtindex.....	25
8.5 Deskriptive Analyse der Dialektdynamik.....	26
8.6 Modelldiagnose.....	31
8.7 Alternative Modelle.....	33
8.8 Alternativer Indikator.....	36
8.9 Weitere Probleme.....	38
9. Zusammenfassung.....	39

1. Einleitung

Anhand umfangreicher Sprachdaten untersucht das Forschungsprojekt AsiCa die Dialekte, die in verschiedenen Orten Kalabriens gesprochen werden. Besonderes Interesse gilt dabei der Frage, ob und inwiefern sich neben der geographischen Herkunft auch die Biographie einer Person auf ihre Sprache auswirkt.

Welchen Einfluss haben Migration, Bildung oder Geschlecht auf die Art und Weise, wie ein Mensch spricht?

Mithilfe verschiedener generalisierter Regressionsverfahren sind wir dieser Frage nachgegangen. Die verwendeten Methoden sowie die Ergebnisse unserer Analysen werden in diesem Bericht vorgestellt.

2. Studie und Design

Grundlage der Folgenden Analysen liefern im DFG-Projekt ASICA (Atlante sintattico della Calabria) der Ludwig-Maximilians-Universität München gewonnene Daten. Ziel dieses Projektes ist die Analyse der sprachlichen Eigenheiten der südlichen italienischen Provinz Kalabrien. Für diese Untersuchung wurde ein zweistufiges Erhebungsdesign entworfen indem zunächst im Fragebogen Sätze aus der Italienischen Hochsprache in ihren Dialekt übersetzt wurden. Daraufhin fand ein ca. 30-minütiges freies Gespräch statt, welches aufgezeichnet und transskribiert wurde.

Befragt wurden acht Familien aus acht verschiedenen Orten in Kalabrien. Die eine Hälfte der Familie war im jeweiligen Ort in Kalabrien und die andere Hälfte in Deutschland wohnhaft. Die Auswahl der Personen erfolgte so, dass jeder Familienteil aus zwei männlichen und zwei weiblichen Personen aus zwei Generationen bestand. Da die Familien aus den Orten Bagnara Calabria und Polistena nur aus sechs bzw. aus fünf und nicht wie die übrigen Familien aus 8 Personen bestanden, nahmen 59 Personen an der Erhebung teil. Die Untersuchungseinheiten setzten sich somit aus 29 weiblichen Personen, 30 männlichen Personen, 31 Personen aus der ersten und 28 Personen aus der zweiten Generation, 29 in Deutschland lebenden, 30 in Italien lebenden Personen zusammen. Das Alter war von 14 bis 76 Jahren verteilt. Das Hauptaugenmerk lag auf Merkmalen der verwendeten Grammatik, jedoch wurden soziodemographische Daten wie Alter und Bildungsniveau ebenfalls erhoben. Wir werden uns im Folgenden nur mit den aus dem Fragebogen gewonnenem Material beschäftigen, da die Auswertung des freien Gesprächs bislang keine reliable Datengrundlage statistischer Analysen liefert. Die interessierenden grammatikalischen Merkmale wurden aus den übersetzten Sätzen gebildet und in 67 Variablen überführt.

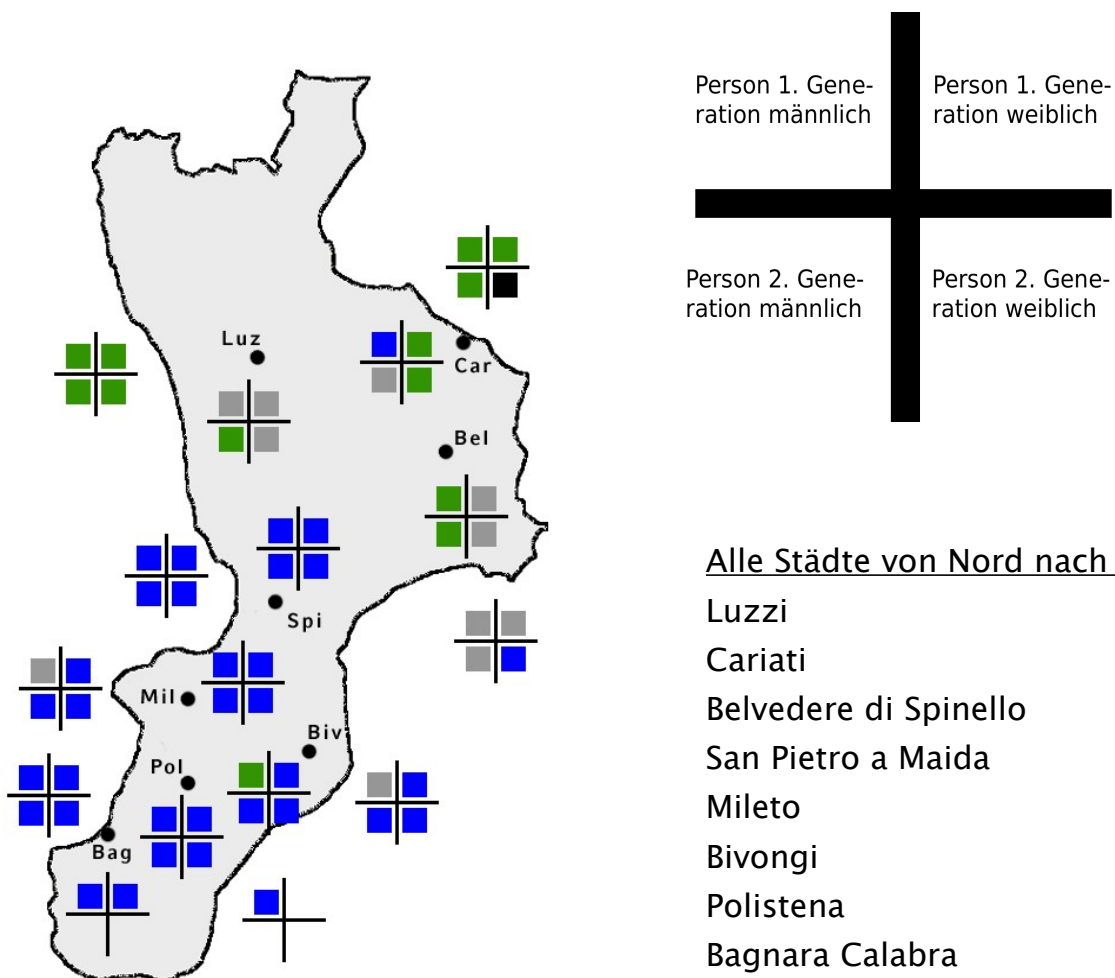
Die 54 Fragen wurden inhaltlich in vier große Frageblöcke aufgeteilt. Die Blockbildung in „complementi verbali“ (Verwendung des Infinitivs), „frasi avverbiali“ (Verwendung des Infinitivs), „tempi verbali“ (Zeitformen) und „proclisi/enclisi“ (Stellung des Possesivpronomens), erfolgte so, dass in den Blöcken jeweils die gleichen Antwortkategorien möglich waren.

Beispielsweise sollte in Frage 16 (Frageblock „complementi verbali“) der Satz „Si vergogna di uscire di casa“ in den Dialekt übersetzt werden wobei in diesem Fall die Verwendung des Infinitivs des Verbs „vergognare“ das grammatikalische Merkmal war. Da es sowohl fehlende, sowie uneindeutige Antworten gab, wurden die Antworten in vier Gruppen („mit Infinitiv“, „ohne Infinitiv“, „fehlender Wert“ und „Ambiguo“ (uneindeutige Antwort)) kategorisiert. Das Beispiel ist in folgender Graphik illustriert:

Merkmalsausprägungen sind farblich gekennzeichnet:

grün	=	mit Infinitiv
blau	=	ohne Infinitiv
schwarz	=	beides (Ambiguo)
grau	=	fehlende Werte

Die Variablen wurden in mehrere Kategorien eingeteilt. Verwendung des Infinitivs, Verwendung von Hilfsverben usw.



3. Ziel der Untersuchung

Ziel der Untersuchung war es zum einen, Unterschiede in den Dialekten bezüglich der Variablen Geschlecht, Generation, Wohnsitz, Herkunftsort und Region zu eruieren und zum anderen, Zusammenhänge zwischen diesen Variablen im Kontext mit der Veränderung der sprachlichen Gepflogenheiten in Verbindung zu bringen.

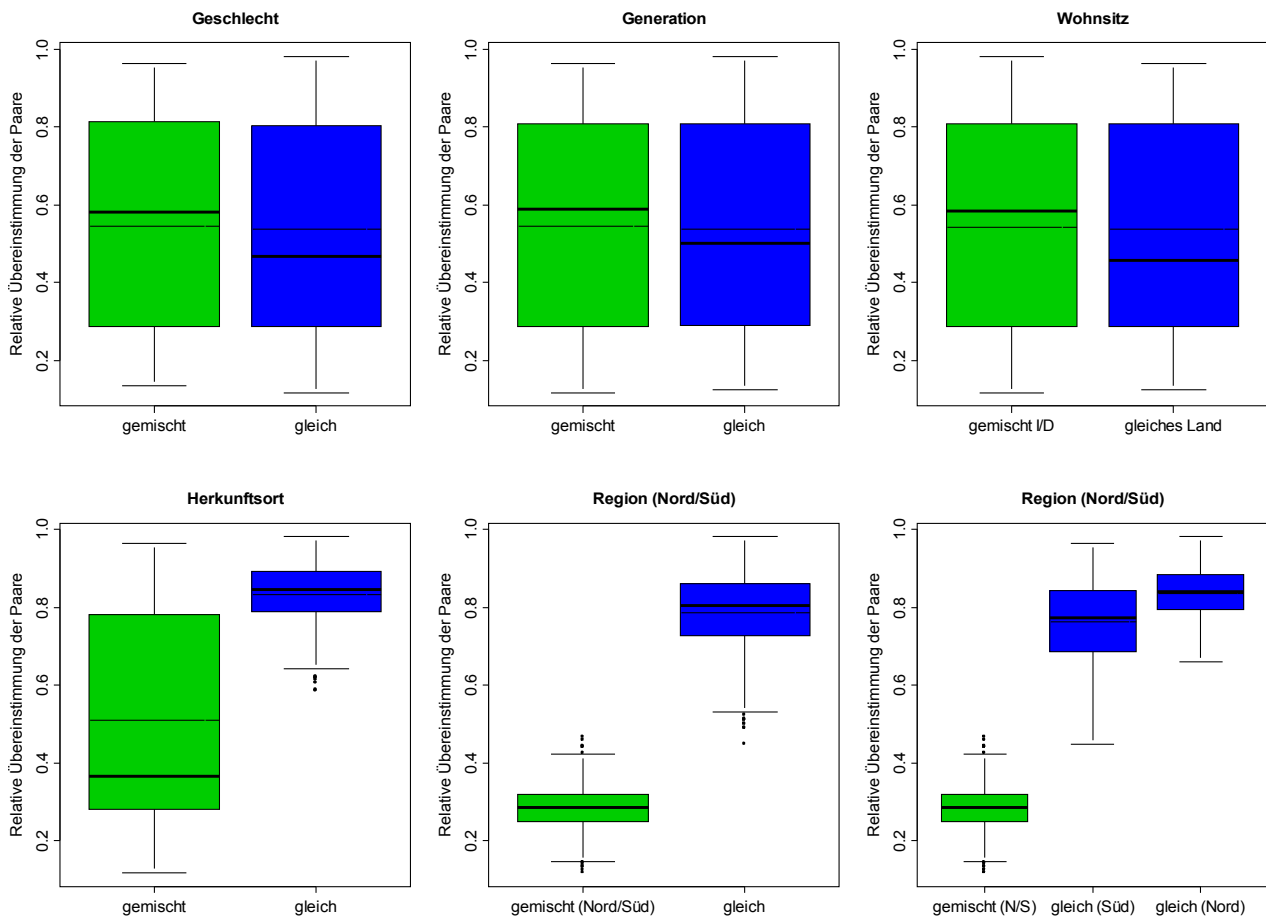
Uns war dabei durchaus bewusst, mit welchen Schwierigkeiten wir bei der Aufgabe, ein Konstrukt wie Dialektdynamik messbar machen zu wollen, konfrontiert werden würden. Angesichts der verwendeten Indikatoren und Verfahren jedoch, können wir guten Gewissens behaupten dieser keineswegs einfachen Aufgabe zumindest einigermaßen gerecht geworden zu sein.

Bevor wir uns jedoch der Untersuchung der Dialektdynamik widmen, werden in den nächsten Kapiteln zunächst Merkmale der Dialekte Kalabriens untersucht. Hierzu bildeten wir folgende Indikatoren.

4. Untersuchung des Dialektes und Bildung von Indikatoren

4.1 Ähnlichkeits-Score

Der erste Ansatz war die Bildung eines Ähnlichkeits-Scores. Dieser zählt Übereinstimmungen der Antworten aller möglichen Personenpaare und bildet relative Anteile. Dieser Ansatz verschafft einen ersten Eindruck über die Einflüsse der Kovariablen. Sollte sich etwa herausstellen, dass Personen mit niedrigerem oder Personen mit höherem Bildungsniveau in ihrer Kategorie homogenere Antworten geben als die Durchschnittsperson, könnte dies auf einen Einfluss des Bildungsniveaus auf den Dialekt hindeuten, der gegebenenfalls genauer untersucht werden könnte. Hierzu eine deskriptive Auswertung des Ähnlichkeits-Scores für die Kovariablen in Boxplots. Ein Boxplot dient dabei zur grafischen Veranschaulichung der Verteilung eines Merkmals. Die Box markiert die mittleren 50% der Daten. Extremere Werte, also die ersten 25% der Daten befinden sich oberhalb, die letzten 25% unterhalb der Box. Die Zäune außerhalb der Box markieren die Spannweite, also Minimum und Maximum. Besonders starke Ausreißer werden als gesondert als Punkte gekennzeichnet. Die dicke Linie innerhalb der Box stellt den Median, die dünne Linie das arithmetische Mittel dar.



Die obere Hälfte der Graphik (Boxplots zu den Variablen Geschlecht, Generation und Wohnsitz) bringt keine offensichtlichen Unterschiede im Ähnlichkeits-Score der beiden Gruppen, gemischt und gleich, zum Vorschein. Median, Mittelwert und Variabilität (Länge des grünen bzw. blauen Bereichs) des Scores unterscheiden sich von Gruppe zu Gruppe nicht bemerkenswert.

Im unteren Teil der Graphik ist augenscheinlich ein regionaler Unterschied erkennbar. Kommen Personen aus dem gleichen Ort (was gleichbedeutend mit gleicher Familie ist), so sind bei geringerer Streubreite weitaus höhere Übereinstimmungen in den Antworten zu beobachten.

Die letzten beiden Boxplots zeigen die Auswertungen unserer ersten Hypothese. Wir vermuteten einen gravierenden Unterschied bei der Aufteilung Kalabriens in Nord (die drei nördlichen Städte) und Süd (die fünf südlichen Städte).

Die Graphiken zeigen tatsächlich beträchtliche Unterschiede. Wird die Region (Nord, Süd) gleich gehalten, so liegen Median und Mittelwert bei weit geringerer Variabilität wesentlich höher. Der letzte Boxplot veranschaulicht zudem, dass die drei nördlichen Städte noch homogenere Antworten in ihrer Gruppe geben (höherer Mittelwert/Median und geringere Streuung).

4.2 Indices aus den vier Teilkategorien

Den zweiten Ansatz lieferte die Auswertung der vier Teilkategorien. Dabei wurden alle Antworten einer Kategorie zu einem auf $[0,1]$ normierten Index zusammengefasst. Mithilfe des EM-Algorithmus wurden fehlende Werte durch eine Regression auf den Ort ersetzt. Wir gingen von zufällig und nicht systematisch fehlenden Werten aus (Missing at Random). Diese Annahme wurde auch durch eine Regression auf die fehlenden Werte, welche keine Auffälligkeiten ergab, bestätigt.

Den Einflüssen der Variablen gingen wir mit Hilfe der Bildung Generalisierter linearer bzw. additiver Regressionsmodelle nach. Allgemein untersucht ein Regressionsmodell den Zusammenhang zwischen einer abhängigen, hier der Index einer Teilkategorie, und mehrerer unabhängiger Variablen, wie Geschlecht, Alter, Bildung usw. Ist ein Einfluss positiv, so wird ein positiver Zusammenhang zwischen einer unabhängigen und der abhängigen Variable angenommen, also z.B. mit steigendem Alter steigt die Wahrscheinlichkeit den Infinitiv zu verwenden. Wir weisen ausdrücklich darauf hin, dass dabei Drittvariablenkontrolle betrieben wird. Das bedeutet, dass die Einflüsse aller anderen Kovariablen „herausgerechnet“ werden. Dies bedeutet, dass beispielsweise das Alter, bei ansonsten gleicher Bildung, gleichem Geschlecht usw. steigt. Ob ein Zusammenhang signifikant und nicht nur zufällig ist, wird anhand des sogenannten p-Wertes, der Irrtumswahrscheinlichkeit, festgestellt. Ist dieser kleiner als 0,05, die Wahrscheinlichkeit, dass der beobachtete Wert in einer Stichprobe wie dieser auftritt, ist kleiner als 5%, so sprechen wir von einem signifikanten Zusammenhang. Üblicherweise wird das Signifikanzniveau auf 5% gesetzt. Jedoch sind Niveaus zu 10% oder 1% ebenfalls häufig anzutreffen. Eine hohe Signifikanz, das heißt ein geringer p-Wert bedeutet jedoch nicht zwangsläufig, dass ein starker Einfluss besteht. Die Entscheidung darüber ob eine Variable ins Modell aufgenommen wurde wurde nach GCV-Kriterium (Kreuzvalidierung) gefällt.

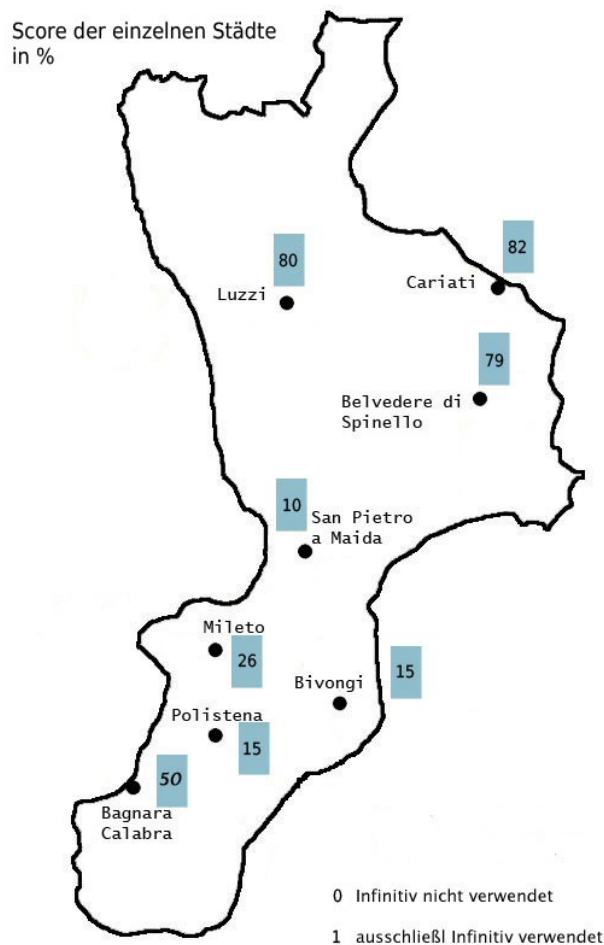
In den nachstehenden Abbildungen sind die auf $[0;1]$ normierten Indices der einzelnen Städte in Prozent abgebildet. Beim Frageblock „complementi verbali“ ergab sich beispielsweise bei der Familie in Cariatì ein Index von 0,82, was bedeutet, dass die Personen in Cariatì in 82 % der Fälle die Infinitivbildung bevorzugten.

4.2.1 Complementi verbali, Verwendung des Infinitivs

Regionale Unterschiede werden auf den ersten Blick deutlich. Als beste Modellierung erwies sich ein generalisiertes additives Modell mit quasi-binomialer Varianz- und Mittelwertstruktur. Der Index sollte durch die Variablen Wohnort, Herkunftsort, Alter und Bildung erklärt werden.

Modell: GAM (quasi-binomial)

Index \sim Wohnort + Herkunftsort + s(Alter) + Bildung



Als Kennwert des Zusammenhangs ergab sich ein R^2 von 0,92. Dass 92% der Streuung im Modell durch den Ort erklärt werden stellt ein eindeutiges Ergebnis dar.

	Estimate	Std. Error	T-value	Pr(> t)	
(Intercept)	1.43554	0.19102	-7.515	1.58e-09	***
Italien	-0.23173	0.13465	1.721	0.0920	.
Alter	-	-	-	0.179	
Cariati	0.31750	0.26891	-1.181	0.2438	
Bel. Di Spin.	-0.02082	0.24045	0.087	0.9314	
San Pietro	- 3.66505	0.29398	12.467	3.49e-16	***
Mileto	- 2.34949	0.23761	9.888	6.05e-13	***
Bivongi	- 3.14175	0.27713	11.337	6.83e-15	***
Polistena	- 3.05095	0.30723	9.930	5.29e-13	***
Bagn. Calabria	-1.50945	0.24318	6.207	1.43e-07	***
ungebildet	0.75850	0.37559	-2.019	0.0493	*

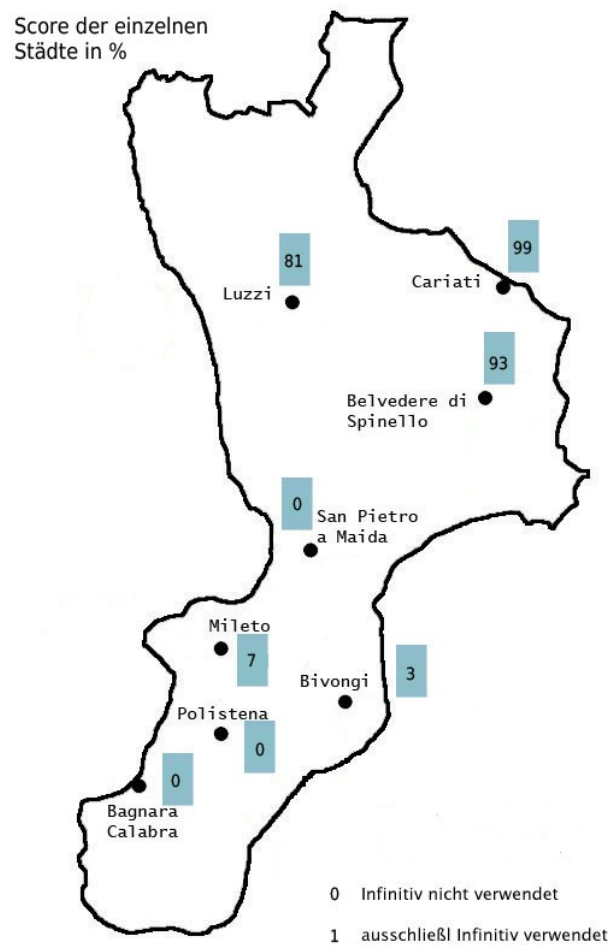
Der Faktor „Schulbildung“ wirkte sich signifikant negativ auf die Verwendung des Infinitivs aus. Der Faktor Wohnort in Italien wirkt sich hingegen positiv, jedoch knapp nicht signifikant, auf die Verwendung des Infinitivs aus. Interpretiert werden können die geschätzten Einflüsse folgendermaßen: Der exponentierte Schätzer gibt den Faktor der Erhöhung bzw. Verringerung der Chance das Infinitiv zu verwenden, wenn die jeweilige Einflussvariable um 1 steigt. Liegt etwa Schulbildung vor, so erhöht sich die Chance der Verwendung des Infinitivs um den Faktor $\exp(0,75850)=2,135$ gegenüber keiner Schulbildung.

4.2.2 Frasi avverbiali, Verwendung des Infinitivs

Ein ähnliches Bild ergibt sich im zweiten Block. Hier werden regionale Unterschiede besonders deutlich. Im Norden wird fast ausschließlich der Infinitiv verwendet, im Süden sehr selten. Hier diene ein generalisiertes lineares Modell mit Wohnort als Einflußgröße.

Modell: GLM (quasi-binomial)

Index ~ Wohnort



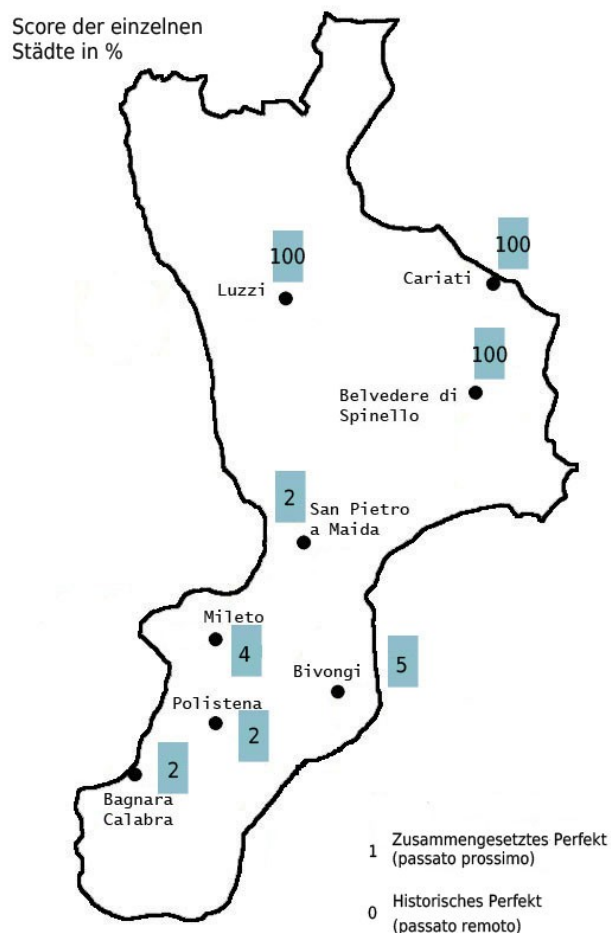
Der Wohnort erklärt 92% der Variabilität. Weitere signifikante Einflüsse konnten nicht festgestellt werden.

4.2.3 Tempi verbali, Zeitformen

Auch hier sind wieder regionale Unterschiede hervorzuheben. Das „passato prossimo“ (zusammengesetztes Perfekt) wird häufig im Norden und das „passato remoto“ (historisches Perfekt) wird häufig im Süden verwendet. Hier diene ein generalisiertes lineares Modell mit den Einflussgrößen Wohnort, Herkunftsort, Geschlecht, Alter, Alter (quadratisch) und Bildung.

Modell: GLM (quasi-binomial)

Index \sim Wohnort + Geschlecht + Herkunftsort + Alter + Alter² + Bildung



Die Regression des Indexes nur auf den Herkunftsort, ergibt ein R² von ca. 0,97*.

	Estimate	Std. Error	T-value	Pr(> t)	
(Intercept)	-1.27124	0.41244	-3.082	0.00337	**
Italien	-0.6417	0.22489	-2.852	0.00634	**
Cariati	1.94115	0.39223	4.949	9.21e-06	***
Bel. Di Spin.	0.11223	0.46056	0.244	0.80850	
San Pietro	-2.74495	1.33253	-2.060	0.04473	*
Mileto	1.13527	0.40333	2.815	0.00701	**
Bivongi	-1.43394	0.64227	-2.233	0.03018	*
Polistena	2.91166	0.43348	6.717	1.82e-08	***
Bagn. Calabria	4.47720	0.53306	8.399	4.72e-11	***
Bildung in Jahren	-0.07988	0.02629	-3.038	0.00381	**

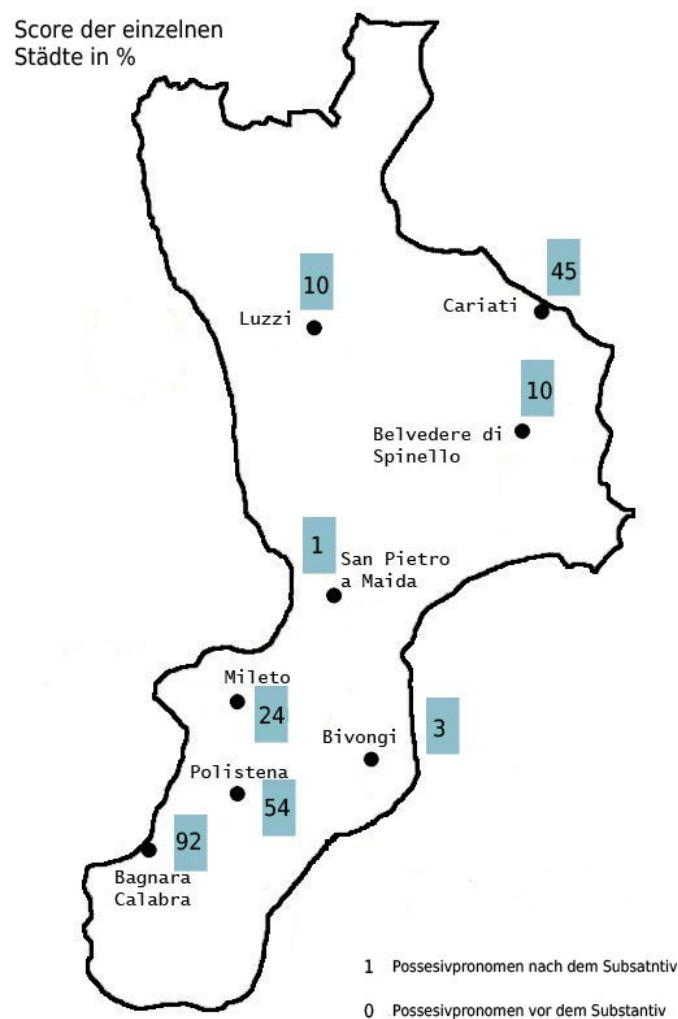
Außer dem Ort wirken sich die Faktoren „männlich“ und „keine Schulbildung“ signifikant positiv aus. Das Alter scheint einen quadratischen Einfluss mit Maximum bei ca. 45 Jahren zu haben. Alle Faktoren sind zum 5%-Niveau signifikant.

4.2.4 Proclisi/enclisi, Stellung des Possessivpronomens

Im letzten Block sind nicht so gravierende regionale Unterschiede zu beobachten wie in den Graphiken zuvor. Hier war die Stellung des Possessivpronomens vor bzw. nach dem Substantiv gefragt.

Modell: GLM (quasi-binomial)

Index \sim Wohnort + Herkunftsort + Bildung in Jahren



Hier erklärt der Herkunftsort alleine ca. 83 % der Devianz.

	Estimate	Std. Error	T-value	Pr(> t)	
(Intercept)	-1.27124	0.41244	-3.082	0.00337	**
Italien	-0.6417	0.22489	-2.852	0.00634	**
Cariati	1.94115	0.39223	4.949	9.21e-06	***
Bel. Di Spin.	0.11223	0.46056	0.244	0.80850	
San Pietro	-2.74495	1.33253	-2.060	0.04473	*
Mileto	1.13527	0.40333	2.815	0.00701	**
Bivongi	-1.43394	0.64227	-2.233	0.03018	*
Polistena	2.91166	0.43348	6.717	1.82e-08	***
Bagn. Calabria	4.47720	0.53306	8.399	4.72e-11	***
Bildung in Jahren	-0.07988	0.02629	-3.038	0.00381	**

Daneben wirken sich die Faktoren „Wohnort in Italien“ und „Bildung in Jahren“ signifikant negativ auf die Verwendung eines Possessivpronomens vor dem Substantiv aus (beides zum 1%-Niveau).

5. Zusammenfassung erster Ergebnisse

Die Anwendung deskriptiver Analysen lies einen Unterschied zwischen Nord- und Südkalabrien vermuten. Sowohl bei der Auswertung des Ähnlichkeits-Scores als auch bei der Auswertung der Teilkategorien schienen die Personen der drei nördlichen Städte Luzzi, Cariati und Belvedere di Spinello und der fünf südlichen Städte San Pietro a Maida, Mileto, Bivongi, Polistena und Bagnara Calabria in ihrer Gruppe homogene Antworten zu geben.

Als problematisch erwies sich in den Modellen jedoch die große Anzahl an Ausreißern. Viele Beobachtungen wiesen eine Cooks-Distance von über eins auf, daher sollten alle Einflüsse außer dem Ort mit großer Vorsicht betrachtet werden, auch wenn sie signifikant sind. Dazu muss erwähnt werden, dass die Anwendung des EM-Algorithmus den Einfluss des Ortes geringfügig überschätzt.

6. Gesamtindex

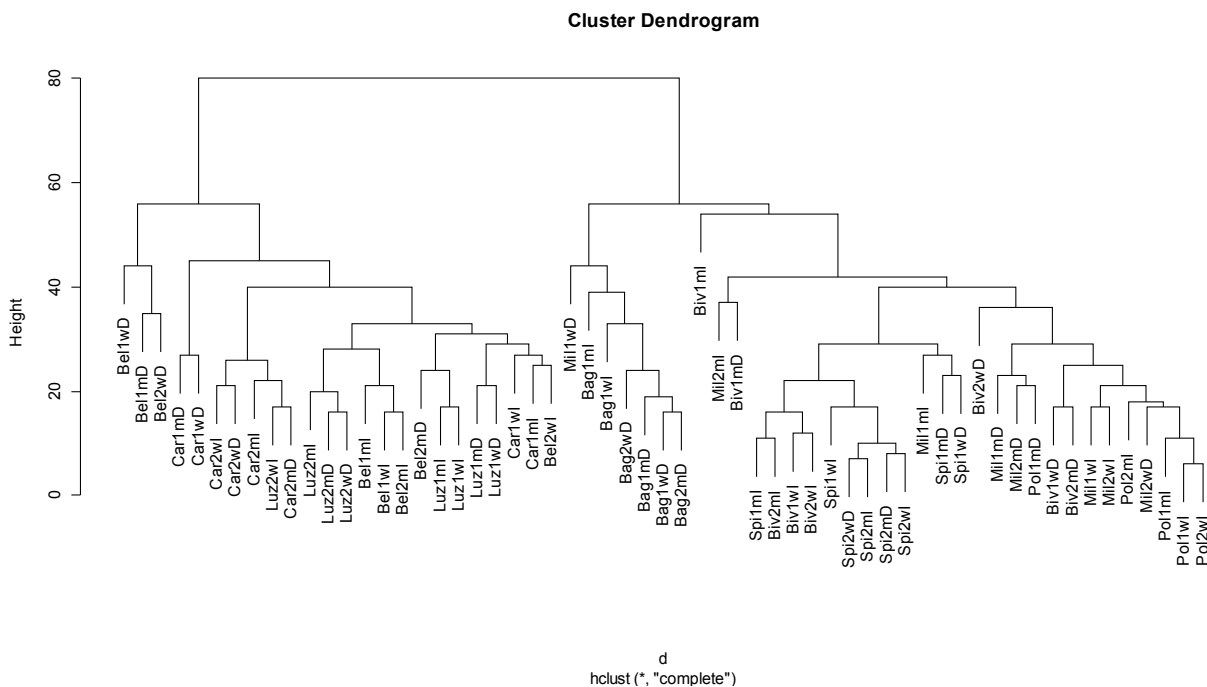
Neben der Untersuchung der einzelnen Frageblöcke erschien auch eine Gesamtauswertung des Indexes über alle Frageblöcke interessant. Wie ähnlich antworten z.B. eine Person aus Cariati und eine aus Belvedere di Spinello im Durchschnitt? Die folgende Kreuztabelle gibt diese Verhältnisse in einer Distanz-Ähnlichkeits-Matrix wieder.

	Luzzi	Cariati	Belv. Di Sp.	San Pietro	Mileto	Bivongi	Polistena	Bagn. Cal.
Luzzi	0.898	0.821	0.838	0.313	0.297	0.299	0.269	0.344
Cariati	0.821	0.816	0.824	0.231	0.254	0.244	0.243	0.321
Belv. Di Sp.	0.838	0.824	0.844	0.278	0.286	0.290	0.284	0.352
San Pietro	0.313	0.231	0.278	0.906	0.774	0.796	0.818	0.671
Mileto	0.297	0.254	0.286	0.774	0.764	0.754	0.805	0.673
Bivongi	0.299	0.244	0.290	0.796	0.754	0.799	0.799	0.633
Polistena	0.269	0.243	0.284	0.818	0.805	0.799	0.913	0.729
Bagn. Cal.	0.344	0.321	0.352	0.671	0.673	0.633	0.729	0.848

Die Zahlen in der Hauptdiagonalen zeigen die durchschnittliche Übereinstimmung einer Person mit seinen Familienmitgliedern, also dem jeweiligen Ort. Die übrigen Einträge der symmetrischen Matrix stellen die durchschnittliche Übereinstimmung einer Person eines Ortes mit einer beliebigen Person eines anderen Ortes dar. Der Index 0.821 zwischen Cariati und Luzzi bedeutet, dass die Antworten einer beliebigen Person aus Cariati und einer beliebigen Person aus Luzzi in ca. 82% der Fälle übereinstimmen. In der Graphik wird zum einen das bereits vermutete Nord-Süd-Gefälle und zum anderen die Sonderstellung Bagnara Calabrias deutlich. Personen aus Bagnara Calabria haben, im Vergleich zu Personen aus anderen südlichen Städten, geringere Übereinstimmungen mit Personen aus derselben Region.

7. Clusteranalyse

Um den Unterschieden zwischen Nord- und Südkalabrien nachzugehen, führten wir eine Clusteranalyse mit complete-Linkage durch. Die durchgeführte Clusteranalyse erhärtete unseren Verdacht. Wie in folgender Graphik ersichtlich wird, befinden sich die Personen der drei nördlichen Städte und die Personen der fünf südlichen Städte jeweils auf einem Ast des Dendrogramms. Die Kürzel am Ende jedes Astes symbolisieren die einzelnen Personen. "Bag1wl" ist z.B. die Person aus Bagnara Calabria (Bag) aus der ersten Generation (1), weiblich (w) mit Wohnsitz Italien (I).



Es ließe sich eventuell eine dritte Gruppe bilden die, bis auf eine, alle Personen aus Bagnara Calabria umfasst. Bemerkenswert war ebenfalls, dass die Personen aus Mileto im Gegensatz zu den anderen Familien, weit über den zweiten Ast verstreut sind. Zusammenfassend lässt sich sagen, dass die Dialekte in Kalabrien sich grob in zwei regionale Gruppen, Nord und Süd, unterteilen lassen, wobei der südlichste Ort Bagnara Calabria anscheinend eine Sonderstellung innerhalb der zweiten Gruppe einnimmt.

8. Untersuchung der Dialektdynamik

Wie lässt sich nun Dialektdynamik messen?

Leider lagen keine Zeitreihendaten vor. Demnach waren sprachliche Veränderungen einer Person bzw. einer Personengruppe in einem bestimmten Zeitraum nicht analysierbar. Es erschien sinnvoll, folgende Indikatoren der Dialektdynamik herauszuarbeiten. Bei den nachfolgenden Regressionsmodellen ist eine direkte Interpretation der Regressionskoeffizienten nur schwer möglich, da das Konstrukt Dialektdynamik einheitenlos ist. Daher sollte ausschließlich das Vorzeichen, die Signifikanz und die absolute Größe im Vergleich mit den anderen Koeffizienten bewertet werden.

8.1 Ambiguo-Score

Der Ambiguo-Score gibt die Anzahl der uneindeutigen Antworten einer Person bzw. Personengruppe wider. Er misst die Anzahl der „Ambiguo“ (Antworten die nicht genau einer Kategorie zugeordnet werden konnten) und drückt somit die Unsicherheit des oder der Befragten aus. Diese Unsicherheit kann auf eine Veränderung der Sprache deuten. Somit ergibt sich die einfache und plausible Hypothese: Je höher die Unsicherheit, desto höher ist die Dynamik. Die zu erwartende Anzahl von uneindeutigen Antworten, die jede Person im Übersetzungsbogen gegeben hat, wurde durch ein Generalisiertes Lineares Modell (GLM) unter Annahme einer negativ-binomial-verteilten Zielgröße und mit logarithmischer Linkfunktion modelliert:

	Estimate	Std. Error	T-value	Pr(> z)	
(Intercept)	-1.638214	0.866834	-1.890	0.05877	.
Italien	-0.362949	0.205125	-1.769	0.07683	.
Alter	0.116707	0.040636	2.872	0.00408	**
Alter²	-0.001288	0.000493	-2.612	0.00900	**
Cariati	0.604372	0.404989	1.492	0.13562	
Bel. Di Spin.	0.681531	0.393817	1.731	0.08353	.
San Pietro	-0.256467	0.471630	-0.544	0.58659	
Mileto	1.038825	0.381543	2.723	0.00648	**
Bivongi	-0.031817	0.438692	-0.073	0.94218	
Polistena	0.443804	0.475195	0.934	0.35033	
Bagn. Calabria	0.118547	0.450893	0.263	0.79262	
ungebildet	1.348363	0.559921	2.408	0.01603	*

Als hoch signifikante Einflussgrößen (Signifikanzniveau $\alpha = 5\%$) erweisen sich das Alter, der Herkunftsort und das Vorhandensein von Schulbildung. Der quadratische Einfluss des Alters legt nahe, dass die Wahrscheinlichkeit für uneindeutige Formulierungen bei Personen im mittleren Alter am höchsten war, während jüngere und ältere Probanden weniger dazu neigten. Analphabetismus (bzw. das Fehlen von Schulbildung) wirkt sich positiv auf das Auftreten von Ambiguo-Fällen aus.

Um ein Maß für die Wirkung zu erhalten, muss der Schätzwert (hier: 1,348363) in die Formel $\text{Faktor} = \exp(\text{Schätzwert})$ eingesetzt werden. Demnach geben Analphabeten (bei sonst gleichen Bedingungen) 3,9 Mal so oft uneindeutige Antworten wie Personen mit Schulbildung. Unter den Herkunftsorten sticht Mileto heraus. Die Wahrscheinlichkeit für Ambiguo-Fälle war für die Mitglieder der entsprechenden Familie um den Faktor 2,8 größer als im Referenzort Luzzi. Nur schwach signifikant (Signifikanzniveau $\alpha = 10\%$) ist der Einfluss des Wohnsitzes: Die ausgewanderten Personen antworteten rund 40% häufiger uneindeutig als ihre Familienmitglieder in Italien.

Dieses Modell, in das das Alter in Jahren eingeht, besitzt etwas mehr Erklärungskraft als ein entsprechendes Modell mit Generation als Einflussgröße. Letzteres ergibt allerdings einen signifikant niedrigeren Erwartungswert für die junge Generation im Vergleich zur älteren.

8.2 Redefinition des Ähnlichkeits-Scores

Der in Kapitel 4.1 vorgestellte Ähnlichkeits-Score wurde dahingehend geändert, dass nicht die durchschnittliche Übereinstimmung einer Person mit einer beliebigen Person, sondern die durchschnittliche Übereinstimmung mit den Familienmitgliedern gemessen wird. Die Übereinstimmung mit den Familienmitgliedern ist gleichbedeutend mit der Übereinstimmung mit Personen aus dem gleichen Ort, da jede Familie aus einem Ort stammt und diese Familie ihren Ort repräsentiert.

Uns war auch hier durchaus bewusst, dass eine Familie nicht unbedingt repräsentativ für den jeweiligen Ort sein muss.

Nichtsdestotrotz lautete hier die Hypothese: Je niedriger die Übereinstimmung mit der Familie, desto höher ist die Dialektdynamik.

Die durchschnittliche Übereinstimmung jeder Person mit ihren Familienmitgliedern wurde in Bezug auf die gleichen Kovariablen analysiert. In diesem Fall erwies sich ein GLM mit gammaverteilter Zielgröße (log-link) als sinnvollste Variante:

	Estimate	Std. Error	T-value	Pr(> t)	
(Intercept)	-0.4129561	0.2923564	-1.413	0.164388	
Italien	-0.1268850	0.0817485	-1.552	0.127337	
Alter	0.0261112	0.0140780	1.855	0.069910	.
Alter²	-0.0002854	0.0001756	-1.625	0.110758	
Cariati	0.7362755	0.1594494	4.618	3.03e-05	***
Bel.di Spin.	0.5960364	0.1515176	3.934	0.000274	***
San Pietro	-0.0830571	0.1530033	-0.543	0.589802	
Mileto	1.0789601	0.1532068	7.043	7.05e-09	***
Bivongi	0.7071757	0.1614045	4.381	6.57e-05	***
Polistena	-0.1687677	0.1753577	-0.962	0.340764	
Bagn. Calabria	0.4802551	0.1667588	2.880	0.005972	**
ungebildet	0.5647061	0.2254914	2.504	0.015796	*

Wieder ergeben sich die Einflüsse der Kovariablen als $\exp(\text{Schätzwert})$. In diesem Modell kann die Auswirkung des Alters und des Wohnsitzes nicht mehr als signifikant angesehen werden (auch der Einfluss der Generation in einem alternativen Modell erreicht nicht das Signifikanzniveau von 5%). Analphabetismus hingegen wirkte sich auch auf die Ähnlichkeit der Antworten signifikant aus. Die Probanden ohne Schulbildung wichen in ihren Übersetzungen (bereinigt um den Einfluss der anderen Größen) 1,8 mal so oft von den Versionen ihrer Familienmitglieder ab wie die übrigen Personen. Die Unterschiede zwischen den Herkunftsorten dürfen nicht vernachlässigt werden: In Cariati, Belvedere di Spinello, Mileto, Bivongi und Bagnara Calabria waren Übereinstimmungen in der verwendeten Grammatik signifikant seltener als im Referenzort Luzzi.

8.3 Kategorien-Gesamtindex

Der Kategorien-Gesamtindex setzt sich aus den in Kapitel 4.2 vorgestellten Kategorien zusammen. Hierbei wurden aus den Regressionsmodellen zu den vier Teilkategorien die betragsmäßigen Residuen extrahiert.

Hier lautet die Hypothese: Je höher die Abweichung vom Erwartungswert, desto höher ist die Dialektdynamik.

Eine Regression des Residuenindex – wieder wurde der Erwartungswert durch ein GLM mit Gammaverteilung modelliert (log-link) – bietet ein ähnliches Bild wie zuvor schon die Analyse des Ambiguo-Scores:

	Estimate	Std. Error	T-value	Pr(> t)	
(Intercept)	-1.4169512	0.4227521	-3.352	0.001592	**
Italien	-0.3227855	0.1182096	-2.731	0.008872	**
Alter	0.0505301	0.0203570	2.482	0.016687	*
Alter²	-0.0005725	0.0002539	-2.255	0.028843	*
Cariati	1.3347977	0.2305665	5.789	5.60e-07	***
Bel. Di Spin.	0.8174802	0.2190968	3.731	0.000513	***
San Pietro	-0.3483291	0.2212453	-1.574	0.122103	
Mileto	1.4881376	0.2215395	6.717	2.20e-08	***
Bivongi	0.4782712	0.2333936	2.049	0.046047	*
Polistena	0.7714177	0.2535701	3.042	0.003835	**
Bivongi	0.5678388	0.2411359	2.355	0.022758	*
ungebildet	0.7934410	0.3260642	2.433	0.018815	*

Der Wohnsitz in Italien wirkte sich signifikant negativ auf diesen Indikator für Dialektdynamik aus, d.h. die Vorhersagemodelle für die Verwendung der einzelnen grammatikalischen Formen waren für die entsprechenden Personen zuverlässiger, während es unter den emigrierten Teilnehmern durchschnittlich zu größeren Abweichungen kam. Der quadratische Einfluss der Lebensjahre legt auch hier ein Maximum an Dialektdynamik im mittleren Alter nahe.

Ein alternatives Modell mit der Einflussgröße Generation erwies sich als deutlich schlechter und ließ auch keinen signifikanten Einfluss der Generationszugehörigkeit erkennen. Der Herkunftsort hat wiederum einen signifikanten Einfluss: Außer San Pietro a Maida (kein signifikantes Ergebnis) haben alle anderen Orte höhere durchschnittliche Residuen als Referenzort Luzzi, wobei besonders Cariati und Mileto mit extremen Werten auffallen. Zuletzt besitzt die Schulbildung Erklärungskraft für die Höhe der Residuen: Die Syntax der Analphabeten unter den Probanden war mit den vier Modellen offenbar am Schwersten vorherzusagen,

auch dies möglicherweise ein Hinweis auf die besondere Dialektdynamik von Personen, die keinen Zugang zur Schriftsprache haben.

8.4 Gesamtindex

Schließlich wurde aus den bereits vorgestellten Indikatoren ein Gesamtindikator gebildet.

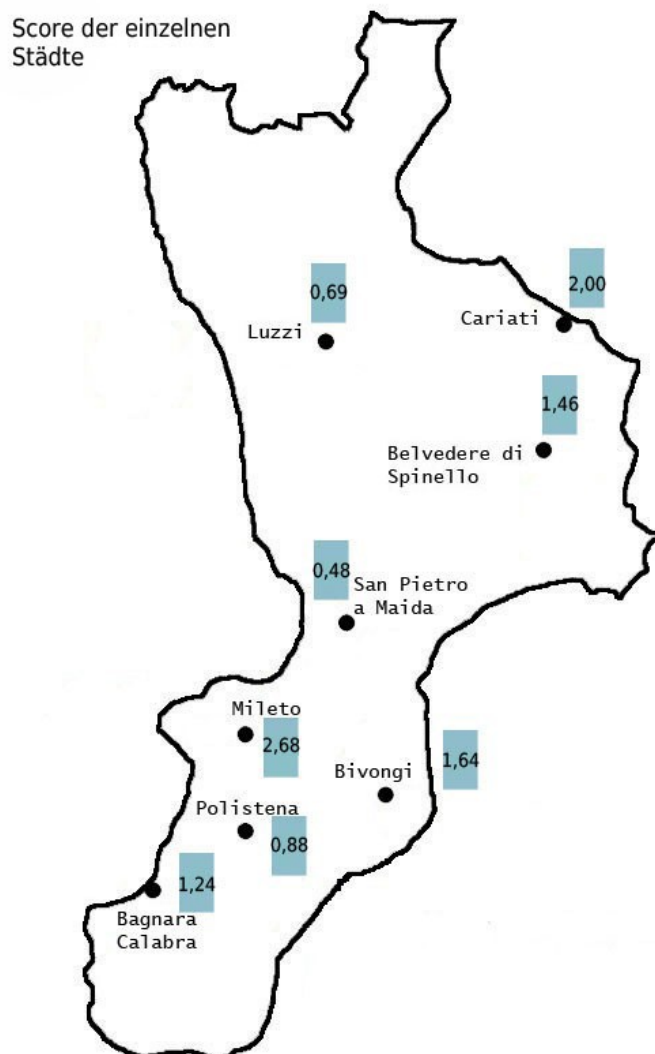
Wünschenswert ist eine Zusammenfassung der drei vorgestellten Maßzahlen zu einem einzigen Indikator für Dialektdynamik. Wie schon die Ergebnisse der vorangegangenen Regressionsanalysen nahe legen, scheinen sich die drei Indizes durchaus ähnlich zu verhalten. Bei Betrachtung der Korrelationen bestätigt sich dies: Zwischen den Teilindizes besteht ein mittlerer bis großer linearer Zusammenhang, sie scheinen also grundsätzlich die selbe Größe zu messen und sich daher für eine Zusammenfassung zu eignen:

	Ambiguo-Index	Kategorien-Index	Ähnl.-Score
Ambiguo-Index	1.0000000	0.4821761	0.7452833
Kategorien-Index	0.4821761	1.0000000	0.6459023
Ähnl.-Score	0.7452833	0.6459023	1.0000000

Ein Indiz für die Reliabilität einer solchen zusammengefassten Messgröße stellt Cronbachs Alpha dar, wobei Werte zwischen 0,7 und 1 gemeinhin als hinreichend reliabel betrachtet werden. Unsere Maßzahlen erzielen einen Wert von 0,83 und eignen sich somit zur Bildung eines Gesamtindikators. Zu diesem Zweck wurden die einzelnen Teilindizes gemäß ihrem durchschnittlichen Wert und ihrer Varianz standardisiert, um vergleichbare Skalen zu erhalten, und anschließend aufaddiert.

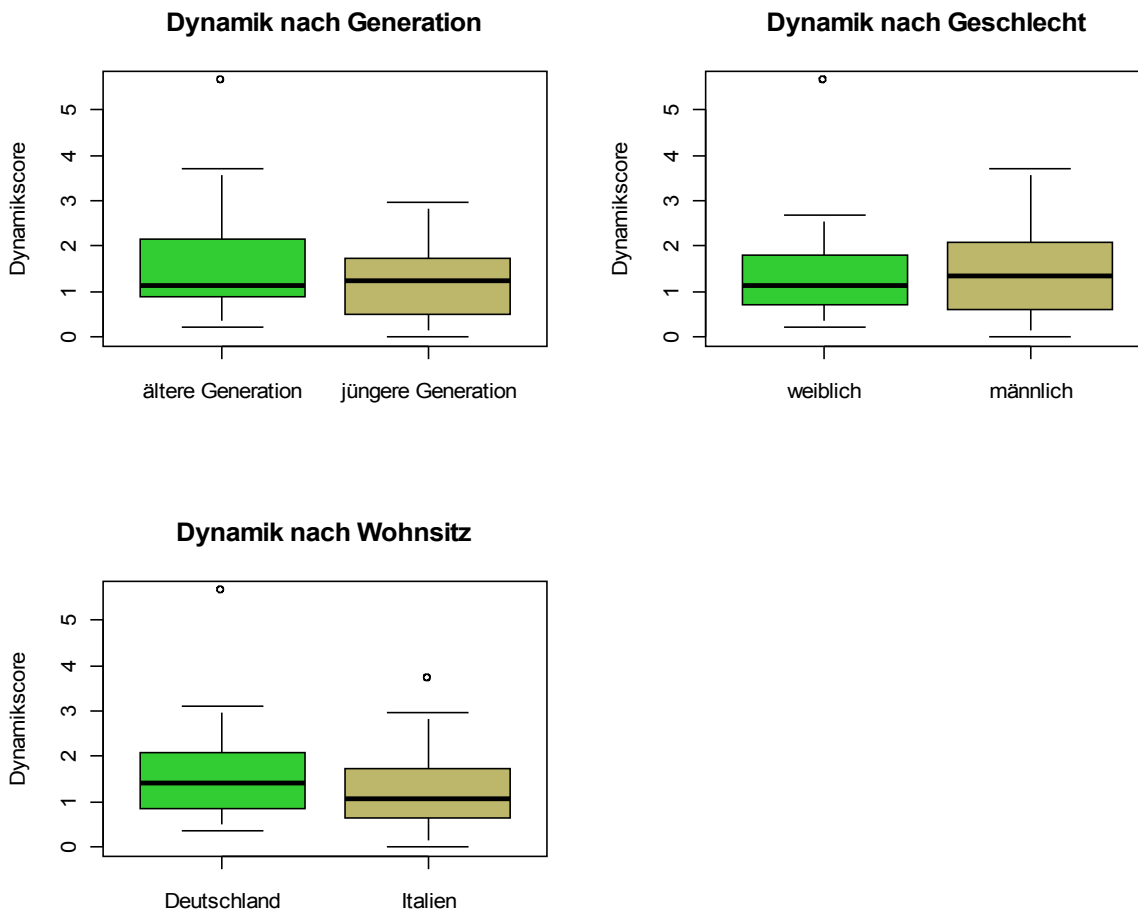
8.5 Deskriptive Analyse der Dialektdynamik

Nachdem die theoretische Größe Dialektdynamik nun messbar gemacht worden ist, lässt sie sich auch im Zusammenhang mit den verschiedenen Einflussgrößen analysieren. Eine deskriptive Untersuchung der durchschnittlichen Dynamikwerte der acht Familien erfolgt in kartographischer Darstellung:

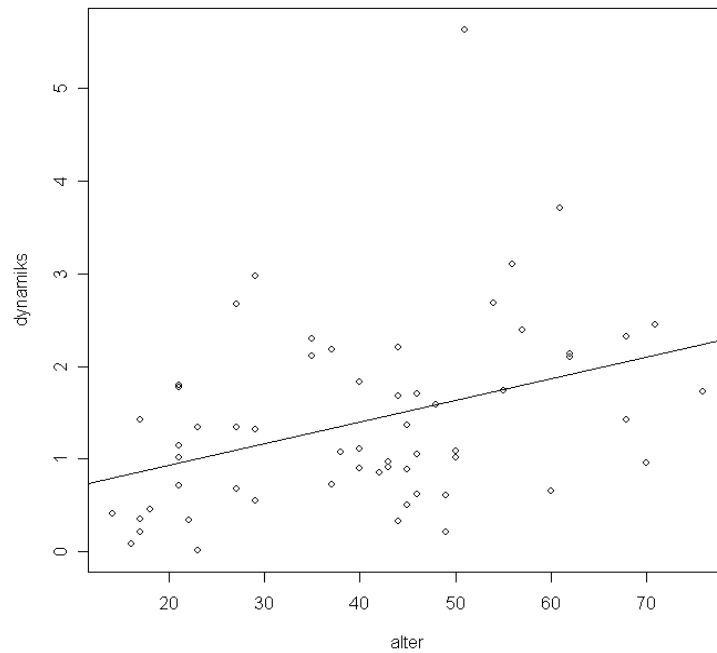


Die Untersuchungsteilnehmer aus den Orten San Pietro a Maida, Luzzi und Polistena weisen demnach eine besonders geringe Dynamik auf. Dass sie in Mileto

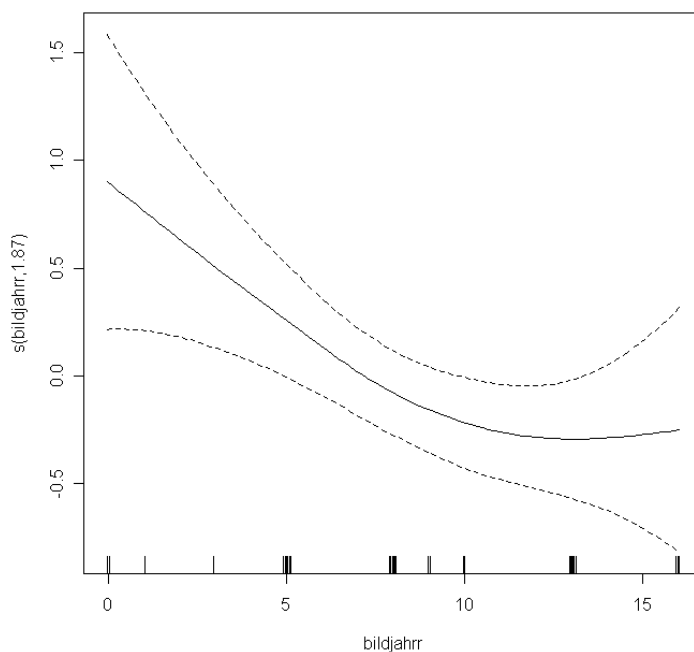
bei Weitem am höchsten ausfällt, ist angesichts der vorangegangenen Regressionsergebnisse sowie der Clusteranalyse der übereinstimmenden Antworten keine große Überraschung. Die folgenden Gegenüberstellungen von Boxplots für Generation, Geschlecht und Wohnsitz enthüllen hingegen keine gravierenden Unterschiede.



Die mittlere Sprachdynamik scheint bei den in Italien wohnhaften Personen etwas geringer zu sein als bei denjenigen, die in Deutschland leben. Frauen und Männer bzw. Alte und Junge unterscheiden sich allenfalls in der Varianz der Dynamik, kaum aber in ihrer mittleren Höhe. Trägt man die Dynamikwerte der 59 Probanden gegen ihre Lebensjahre ab, offenbart sich eine steigende Tendenz mit zunehmendem Alter. In einem linearen Regressionsmodell (eingezeichnete Gerade) unter Vernachlässigung aller anderen Variablen stellt sich dieser Zusammenhang als hochsignifikant heraus.



Der R^2 -Wert von 0,13 macht jedoch deutlich, dass das Alter alleine noch keine große Aussagekraft über die Dialektdynamik einer Person hat; nur 13% der Varianz in den Dynamikwerten lassen sich in diesem Modell auf den Einfluss des Alters zurückführen. Ein linearer Zusammenhang lässt sich für die genossene Ausbildung in Jahren nicht feststellen. Stattdessen wird in einem linearen Regressionsmodell mit flexibler Anpassung (penalisierter B-Spline) ein signifikanter monoton fallender Zusammenhang sichtbar:



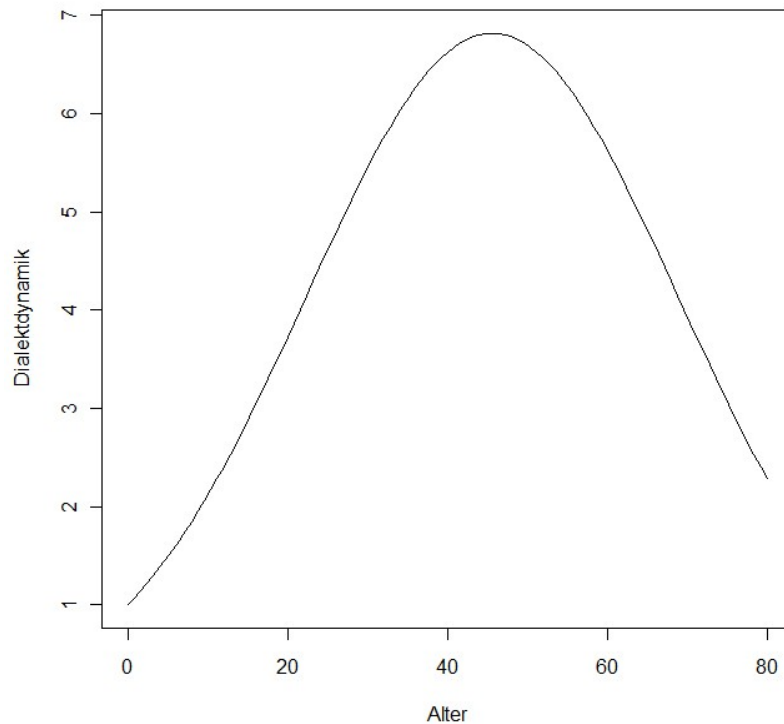
Ungeachtet aller anderen Variablen ist die Dialektdynamik bei Personen mit keiner oder geringfügiger Schulbildung stärker ausgeprägt als bei solchen, die eine höhere Schule oder Universität besucht haben. Offenbar ist die Wirkung eines weiteren Ausbildungsjahres im Anfangsstadium der Ausbildung noch gravierender als nach langjährigem Besuch einer Schule, da sich die Kurve in ihrem Verlauf verflacht. Die Regression ausschließlich auf die Dauer der Ausbildung vermag allerdings nur 11% der Streuung in der Dialektdynamik zu erklären.

Welche der deskriptiv gewonnenen Erkenntnisse über den Zusammenhang der verschiedenen Kovariablen mit der Messgröße für Dialektdynamik lassen sich in einer Regressionsanalyse erhärten? In einem GLM mit gammaverteilter Zielgröße (log-link) erwiesen sich alle Kovariablen außer der Generationszugehörigkeit als signifikant:

	Estimate	Std.Error	T-value	Pr(> t)	
(Intercept)	-1.9201190	0.4531252	-4.238	0.000107	***
Italien	-0.3454209	0.1263513	-2.734	0.008857	**
Alter	0.0841905	0.0217647	3.868	0.000343	***
Alter²	-0.0009231	0.0002715	-3.400	0.001401	**
Cariati	1.1446637	0.2460822	4.652	2.80e-05	***
Bel.di Spin.	0.9513253	0.2338321	4.068	0.000184	***
San Pietro	-0.2582042	0.2361301	-1.093	0.279875	
Mileto	1.4710114	0.2364418	6.221	1.35e-07	***
Bivongi	0.6164404	0.2492629	2.473	0.017152	*
Polistena	0.5291959	0.2710842	1.952	0.057023	.
Bagn. Calabria	0.6954791	0.2574588	2.701	0.009637	**
ungebildet	1.0581775	0.3519512	3.007	0.004270	**
männlich	-0.2498004	0.1233649	-2.025	0.048706	*

Unter den Herkunftsorten weist Mileto die stärkste Tendenz zu Dialektdynamik auf, San Pietro a Maida und Luzzi die geringste. Bei den Männern ist die selbe Tendenz etwas schwächer ausgeprägt als bei den Frauen, bei den „Auswanderern“ etwas stärker als bei den in Italien Wohnenden. In diesem vollen Regressionsmodell zeigt sich, dass der Einfluss des Alters im Zusammenspiel mit den anderen Kovariablen kein linearer, sondern ein quadratischer ist.

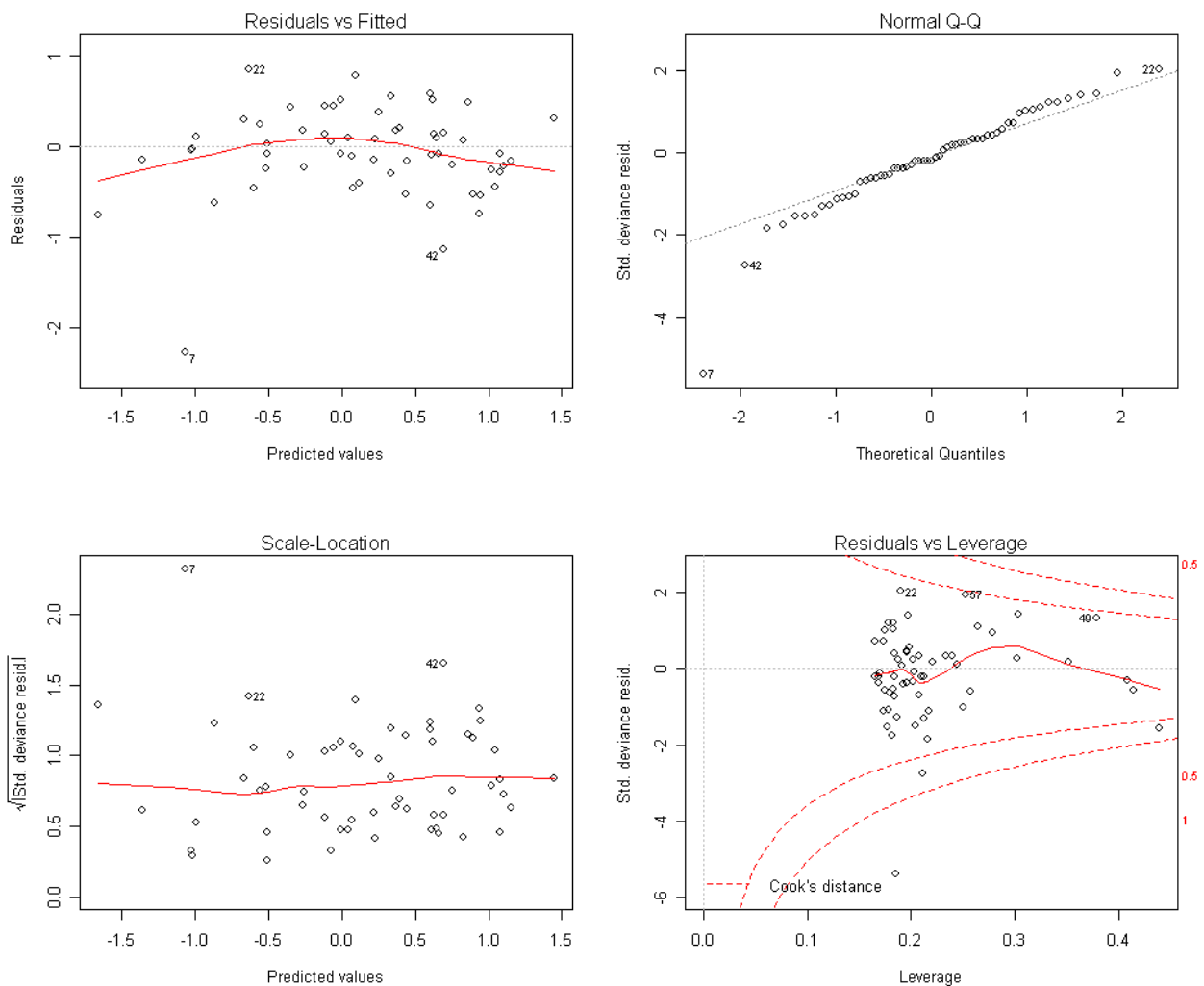
Die Neigung zur Dialektdynamik nimmt mit den Lebensjahren also nicht monoton zu, sondern erreicht einen Hochpunkt bei etwa 45 Jahren, um dann wieder abzufallen:



Bei der Interpretation muss man allerdings beachten, dass dieser Effekt auf einer Analyse von Querschnittsdaten zu einem Zeitpunkt mit einer bestimmten Altersverteilung basiert, also nicht zwingend einen zeitlichen Verlauf im Leben *eines* Menschen widerspiegelt. Entsprechend der Feststellung, dass der Einfluss der Schulbildung mit Zunahme der Ausbildungsjahre geringer wird, genügte es, diese Variable als binäre Größe (Schulbildung vorhanden/nicht vorhanden) ins Modell aufzunehmen. Konkret bedeutet das, dass der negative Einfluss, den der Bildungsgrad auf die Dialektdynamik zu haben scheint, zu einem großen Teil auf das Sprachverhalten der vier Analphabeten zurückzuführen ist. Diese erreichen im Schnitt etwa 2,9 Mal so hohe Dynamikwerte wie die übrigen Personen.

8.6 Modelldiagnose

Die Ergebnisse der Regressionsanalyse entsprechen recht gut den Vermutungen, die sich schon aus der deskriptiven Betrachtung der Daten ergeben haben. Doch wie gut passt das verwendete GLM mit Gammaverteilungsannahme zur tatsächlichen Struktur der Daten? Dies lässt sich in verschiedenen sogenannten Diagnostik-Plots anschaulich machen. Diese zeigen, dass das Modell grundsätzlich recht gut an die konkreten Datenwerte angepasst ist und sich auch die Verteilungsannahme aufrecht erhalten lässt:

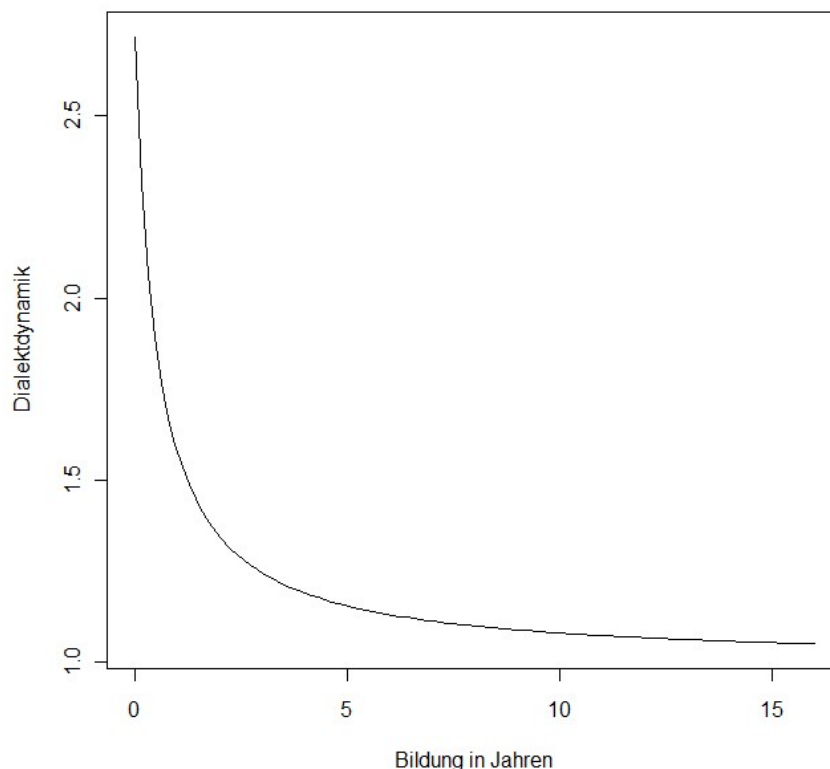


Probleme bereitet nur Beobachtung 7 aus unserem Datensatz, die in allen vier Graphiken markiert ist (es handelt sich um den männlichen Teilnehmer der jüngeren Generation aus Luzzi mit Wohnsitz in Italien). Ansonsten scheinen die Modellannahmen grundsätzlich erfüllt zu sein.

Bei keiner anderen Person unterscheidet sich der tatsächliche Dynamikwert annähernd so gravierend von der Vorhersage durch das Modell wie in diesem Fall. Zugleich nimmt die entsprechende Beobachtung dadurch auch einen unangemessen großen Einfluss auf die Anpassung des ganzen Modells. Es war daher ratsam, das selbe Modell noch einmal unter Auslassung dieser einen Person zu berechnen. Die Modellannahmen erschienen anschließend noch schlüssiger und die Aussagen des vorherigen Modells blieben im Kern erhalten. Erwartungsgemäß fielen die Schätzungen für die Variableneffekte nun betragsmäßig etwas geringer aus, da sie zuvor durch den Ausreißer übertrieben worden waren. Nur der Einfluss des Geschlechts, der auch im ersten Modell schon recht gering war, konnte in der Version ohne Ausreißer nicht mehr als signifikant angesehen werden. Ob die Männer in der Studie grundsätzlich eine größere Tendenz zur Dialektdynamik aufweisen als Frauen, bleibt also unsicher.

8.7 Alternative Modelle

Auch wenn das gewählte Modell die Daten gut zu beschreiben scheint, gibt es natürlich Möglichkeiten es zu modifizieren, insbesondere die Auswahl und Modellierung der Einflussgrößen zu alternieren. Die Behandlung der Variable Schulbildung als binäre Größe mag kritikwürdig erscheinen. Daher wurde zusätzlich eine Alternative berechnet, die die Dauer der Schulbildung in Ausbildungsjahren berücksichtigt. Im deskriptiven Teil wurde bereits deutlich, dass ein monoton fallender Zusammenhang zwischen der Ausbildungsdauer und der Stärke der Sprachdynamik mit abnehmender Effektstärke plausibler erscheint als ein linearer Zusammenhang. Daher wurde konkret die transformierte Größe $\frac{1}{1 + \text{Ausbildungsjahre}}$ ins Modell aufgenommen, die ein Abflachen des Effekts bereits einschließt. Grafisch äußert sich der geschätzte Einfluss folgendermaßen:

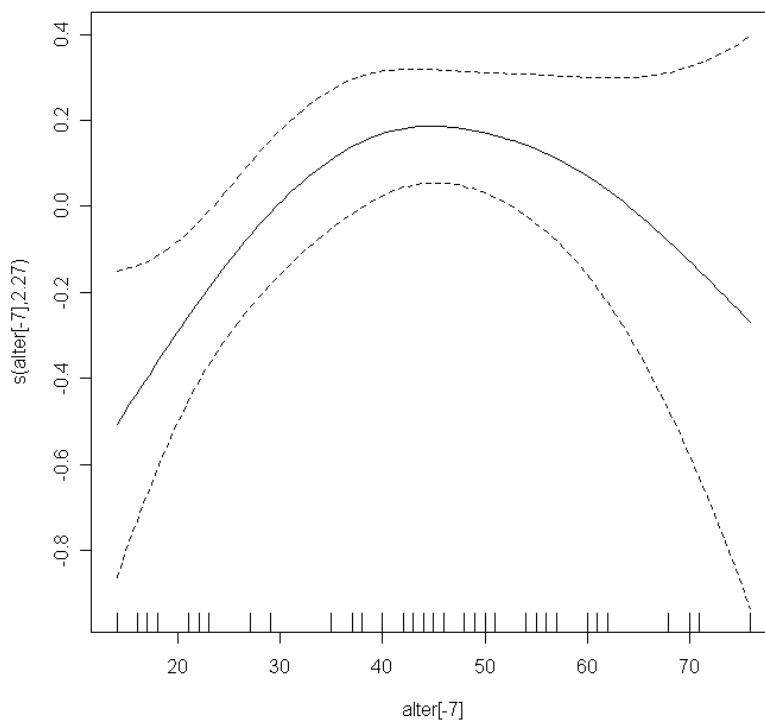


Auch in dieser Form ist der Einfluss der Ausbildung signifikant und das neue Modell unterscheidet sich für die anderen Variablen unwesentlich von dem mit binärer Bildungsvariable. Allerdings besitzt auch das alternative Modell keine größere Erklärungskraft (bzgl. des AIC-Maßes) als das ursprüngliche.

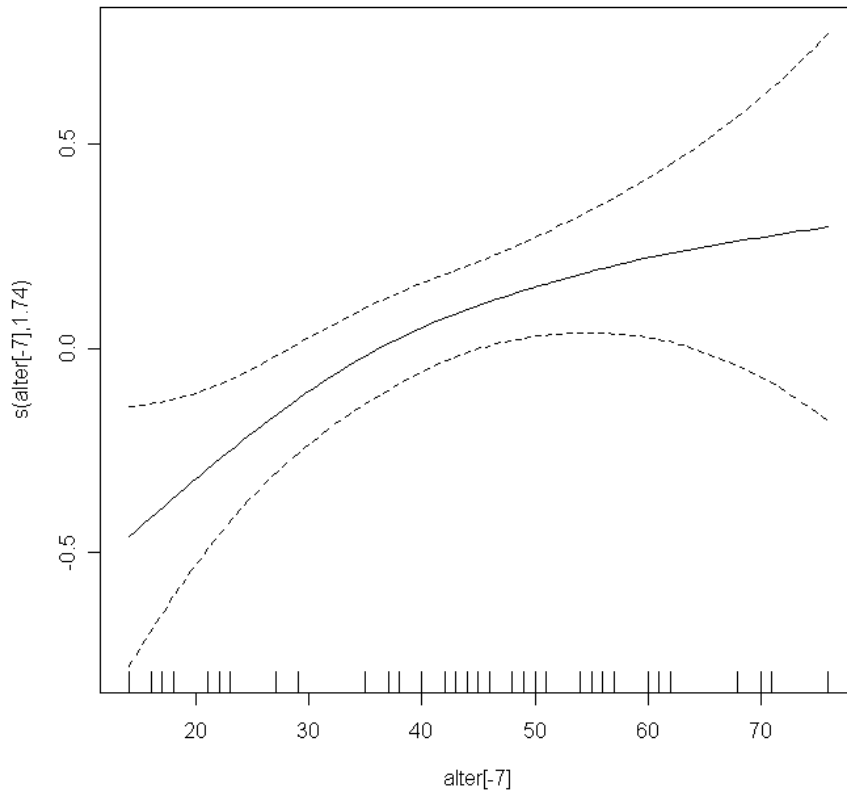
Die genaue Kenntnis der Ausbildungsdauer hat also gegenüber der bloßen Unterscheidung zwischen vorhandener und nicht vorhandener Schulbildung keinen wesentlichen Vorteil bei der Vorhersage der Dialektdynamik einer Person. Außerdem ging der ohnehin fragliche Einfluss des Geschlechts auch bei diesem Modell wieder als signifikanter Effekt verloren.

Darüber hinaus lässt sich das obige Modell auch dahingehend verändern, dass statt des Alters wieder die Generationszugehörigkeit berücksichtigt wird. Die entsprechende Regressionsanalyse zeigt auch tatsächlich eine signifikant größere Dialektdynamik bei der älteren Generation und ansonsten ähnliche Einflüsse wie vorher, das Modell ist jedoch deutlich weniger aussagekräftig (bzgl. AIC) als das ursprüngliche mit Alterseffekt. Auch hier ist wieder keine signifikante Auswirkung des Geschlechts auszumachen.

Die Betrachtung der Lebensjahre ist also gegenüber der Unterscheidung zwischen zwei Generationen vorzuziehen. Allerdings lässt sich der Einfluss des Alters mit Hilfe eines alternativen Generalisierten Additiven Modells (GAM) nonparametrisch und somit in seinem Verlauf flexibler modellieren. Die folgende Abbildung zeigt den Verlauf des auf diese Weise geschätzten Einflusses. Mit einem Maximum der Dialektdynamik bei etwa 45 Jahren stimmt auch diese nonparametrische Schätzung recht gut mit der ursprünglichen Modellierung als quadratischer Größe überein:



Allerdings verdeutlicht das eingezeichnete Konfidenzband, dass die sinkende Dynamik im höheren Alter ungewiss ist. Eine Stagnation auf hohem Niveau wäre ebenfalls denkbar. Die folgende Abbildung zeigt dagegen den Einfluss des Alters, wenn das Ausbildungsniveau gänzlich aus dem Modell entfernt wird:



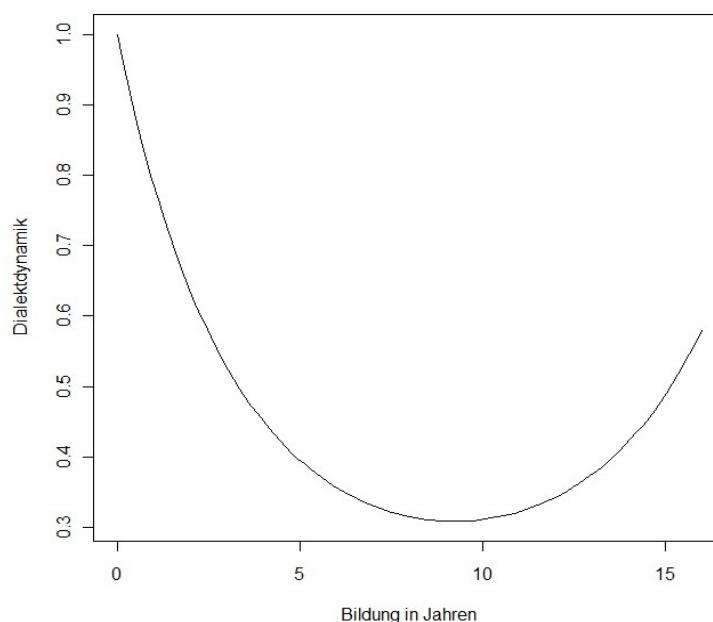
Der Zusammenhang zwischen Alter und Dialektdynamik scheint nun ein monoton steigender zu sein. Dies verdeutlicht noch einmal den großen Einfluss der vier Analphabeten in der Studie auf sämtliche Modelle: Da sie allesamt über 60 Jahre alt waren, verzerren sie nun möglicherweise den Verlauf der Einflusskurve des Alters. Die Problematik zeigt sich auch darin, dass die Größen Ausbildungsdauer und Alter der Personen sehr stark negativ korrelieren (Korrelationskoeffizient $r = -0,8$), die älteren Personen in der Studie eine durchschnittlich geringere Bildung besitzen als die jüngeren. Die Effekte beider Einflussgrößen sind also ohnehin schwer zu unterscheiden.

8.8 Alternativer Indikator

Ein weiteres grundsätzliches Problem ergibt sich aus der bisher verwendeten Methodik. In den konstruierten Indikator für Dialektdynamik fließen neben der Häufigkeit von Mehrdeutigkeiten (Ambiguo-Score) und der Übereinstimmung mit den Familienmitgliedern (Ähnlichkeits-Score) auch die betragsmäßigen Residuen aus den vier Vorhersagemodellen für die grammatikalischen Kategorien ein. Diese Modelle beruhen aber auf einer Regression auf die selben Kovariablen, die später auch in die Regressionsanalyse des Gesamtindikators für Dialektdynamik eingehen. Dadurch könnten unerwünschte Abhängigkeiten entstehen, die die Ergebnisse verfälschen. Dieses Problem lässt sich durch einen alternativen Indikator umgehen. Dieser setzt sich einerseits aus dem Ähnlichkeitsindex bezogen auf die Familienmitglieder (wie oben) sowie andererseits aus vier reduzierten Residuenscores zusammen. Letztere unterscheiden sich von den ursprünglichen Residuenscores dadurch, dass sie aus jenen Vorhersagemodellen hervorgehen, die statt sämtlicher Kovariablen nur den Herkunftsort als Einflussgröße enthalten. Auch diese vereinfachten Vorhersagemodelle verfügten bereits über eine große Erklärungskraft. Um bedenkliche Abhängigkeiten zu umgehen, darf eine Regressionsanalyse des neu gebildeten Indikators für Dialektdynamik den Herkunftsort nun nicht als erklärende Größe enthalten. Einerseits bietet sich dieses Vorgehen an, da die Herkunft einer Person leicht als ausschlaggebend für ihr Sprechverhalten ausgemacht werden konnte. Andererseits hat es den Nachteil, dass so keine Unterschiede in der Dialektdynamik zwischen den einzelnen Orten mehr aufgedeckt werden können. Das resultierende Modell (GLM mit Gammaverteilung) ähnelt stark den schon bekannten Ergebnissen:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.4943071	0.8935252	-0.553	0.5825	
Italien	-0.4750557	0.2266335	-2.096	0.0410	*
Alter	0.0811849	0.0367537	2.209	0.0316	*
Alter²	-0.0007987	0.0004675	-1.709	0.0935	.
Bildungsjahre	-0.2540171	0.1254328	-2.025	0.0480	*
Bildungsjahre²	0.0137821	0.0062704	2.198	0.0324	*
männlich	-0.1196459	0.2180341	-0.549	0.5855	

Wieder zeigen sich die in Deutschland wohnenden Personen anfälliger für Dialektveränderungen, der Einfluss des Alters zeigt den bekannten quadratischen Verlauf, der Einfluss des Geschlechts ist nicht signifikant. Ein wesentlicher Unterschied besteht aber im Effekt des Bildungsgrades (gemessen in Jahren), der nun ebenfalls als quadratisch erscheint. Die Parameterschätzungen legen nahe, dass die Neigung zu dialektdynamischem Verhalten mit zunehmender Bildung nicht kontinuierlich fällt, sondern möglicherweise zum Abitur hin oder bei Aufnahme eines Studiums wieder ansteigt. Es bleibt bei der Interpretation zu beachten, dass ein verdeckter Einfluss des Herkunftsorts hinter derartigen Effekten stehen könnte, da geographische Unterschiede in diesem Modell keine Berücksichtigung finden.



8.9 Weitere Probleme

Neben den Schwierigkeiten bei der Wahl des am besten geeigneten Modells für die Beschreibung des Auftretens von Dialektdynamik gibt es auch solche Aspekte zu beachten, die sich der statistischen Analyse entziehen. Zum einen steht nicht mit Sicherheit fest, was genau der verwendete Indikator für Dialektdynamik eigentlich misst. Als hinreichend reliabel hat er sich erwiesen, aber wie sieht es mit der Validität aus? Es kann nicht garantiert werden, dass eine niedrige sprachliche Übereinstimmung mit den eigenen Familienmitgliedern oder eine gewisse Unsicherheit bei der Verwendung von grammatikalischen Formen tatsächlich für eine Veränderung des Dialekts stehen. Eine solche Veränderung ließe sich mit großer Sicherheit nur in einer Langzeitstudie beobachten. Zum anderen muss bei jeder Interpretation der Ergebnisse das besondere Design der Studie berücksichtigt werden. Insbesondere kann die Tatsache, dass jeder der acht betrachteten kalabrischen Orte jeweils ausschließlich durch Mitglieder einer Familie repräsentiert war, eine Verallgemeinerung der Ergebnisse erschweren. Es besteht immer die Möglichkeit, dass Aussagen, die über die sprachlichen Besonderheiten einzelner Orte gemacht wurden, in Wahrheit nur auf familienspezifischen Phänomenen beruhen.

9. Zusammenfassung

Zur statistischen Analyse der Dialektdynamik in Kalabrien wurden Sätze herangezogen, die Probanden aus dem Hochitalienischen in ihren jeweiligen Dialekt übertragen sollten. Dabei zielte jeder dieser Sätze auf den Umgang mit einem oder mehreren grammatikalischen Problemen ab. Schon bei der deskriptiven Untersuchung der vier grammatikalischen Kategorien *complementi verbali*, *frasi avverbiali*, *tempi verbali* und *proclisi/enclisi* stellte sich heraus, dass der Dialekt einer Person wesentlich durch ihren Herkunftsort bestimmt wird. Konkret scheint sogar eine sprachliche Grenze zwischen den drei nördlichen und den fünf südlichen der vertretenen Orte zu verlaufen.

Das Phänomen der Dialektdynamik wurde durch die Zusammenfassung verschiedener Größen zu einer Indikatorvariablen messbar gemacht. Diese fiel für jede Person umso höher aus, je häufiger sie mehrdeutige syntaktische Lösungen anbot, je geringer ihre sprachliche Übereinstimmung mit den übrigen Repräsentanten ihres Herkunftsortes war, und je gravierender sie in der

Verwendung bestimmter grammatikalischer Formen von entsprechenden statistischen Vorhersagen abwich.

In verschiedenen Regressionsanalysen zeigte sich, dass verschiedene Einflussgrößen die Höhe des Indikators für Dialektdynamik bestimmen. Unter allen Herkunftsorten sticht Mileto als derjenige mit der größten durchschnittlichen Dialektdynamik hervor. Der Einfluss des Alters zeigte einen in etwa quadratischen Verlauf, wobei die Personen im mittleren Alter höhere Werte aufwiesen als junge und alte Probanden. Entscheidend war auch der Effekt des Bildungsniveaus: eine geringe Ausbildungsdauer ging mit einer deutlich erhöhten Dialektdynamik einher, wobei der Unterschied zwischen den Analphabeten und den übrigen Personen in der Studie am gravierendsten war. Möglicherweise neigten aber auch Personen am oberen Ende der Ausbildungsskala stärker zu dialektdynamischem Verhalten als Personen mit mittlerem Ausbildungsniveau. Diejenigen Untersuchungsteilnehmer mit Wohnsitz in Deutschland erreichten durchschnittlich etwas höhere Indikatorwerte als die in Italien lebenden. Ein Einfluss des Geschlechts wurde dagegen nicht ausreichend belegt, es gab allerdings Hinweise auf eine etwas stärker ausgeprägte Dialektdynamik bei Frauen.

Die Interpretation der Ergebnisse wird erschwert durch die sehr indirekte Messung des untersuchten Phänomens Dialektdynamik. Zusätzlich sind Aussagen bezüglich des Herkunftsortes nur mit großer Vorsicht zu treffen, da jeder Ort nur durch Mitglieder je einer Familie vertreten war.