

La digitalisation du dictionnaire en ligne *Pledari grond* et le recueil des textes romanches les plus importants – la *Chrestomathie* – ont permis la publication en open source de données langagières considérables. Ceci ouvre aussi de nouvelles perspectives pour l'enseignement du romanche. Il est apparu que des projets collaboratifs basés sur le web peuvent contribuer à la vivacité d'une langue minoritaire: une plateforme facile d'accès et efficace peut générer un fort engouement pour la langue, comme le montre par exemple l'expérience de *Pledari grond* avec près de 1'000 visites par jour.

## RUMANTSCHIA DIGITALA: DIGITALE RESSOURCEN DES BÜNDNER- ROMANISCHEN

### Einleitung: Das Bündnerromanische

Das Bündnerromanische wird in sprach- und sprachwissenschaftlicher, in historischer, kultureller und politischer Sicht als Einheit wahrgenommen. Es weckte das sprachwissenschaftliche Interesse der Romanistik vorwiegend in der Schweiz, in Deutschland, Österreich und Italien, dabei mitunter nicht frei von politischen Intentionen. Der Schwerpunkt des sprachwissenschaftlichen Interesses lag lange in den Bereichen der Phonetik und der Phonologie, des Lexikons und der Etymologie, der Morphologie, mit starker diachroner Fokussierung. Die Arbeit von Bernard Cathomas (1977) initiierte eine größere Zahl soziolinguistischer Untersuchungen. Wegweisend in ihrer methodischen Strenge ist das einführende Kapitel von Weinreich (1953, 1963) für Fragen des Sprachkontakts und der Konfrontation linguistischer Systeme.

All diese Arbeiten handeln in vorwiegend wissenschaftlich-akademischer Perspektive *über* Sprache und Sprecher, sie verhandeln nicht *mit* den Sprechern. Die Sprecher und ihre Sprache sind Gegenstand, nicht Subjekte. Dies hat sicher

auch mit den Medien zu tun, in denen die erwähnten Untersuchungen publiziert werden: Es sind Printmedien, und dies heisst auch: Sie sind nicht oder nur schwer modifizierbar, sie sind den Rezipienten – den Romanen in den Bergen Bündens, den ausgewanderten Romanen im Mittelland – nicht einfach zugänglich. Diese Situation hat sich mit der Einführung Neuer Medien vor dem Hintergrund des Internets grundlegend geändert. Die Neuen Medien sind interaktiv. Damit bieten sich neue, kollaborative Formen der Wissensgewinnung und der Wissensverbreitung an<sup>1</sup>. Gerade für die Gewinnung von Wissen über kleine Sprachen eröffnen sich damit neue Möglichkeiten. Die Sprecher dieser Sprachen sind nicht mehr nur (passive) Objekte der Wissensgewinnung, sondern haben daran aktiv Anteil. Die Auseinandersetzung mit der Sprache – nicht nur ihre Nutzung – hat den Effekt, kleinere und oftmals bedrohte Sprachen zu stärken. Die Ergebnisse interaktiver Arbeiten stehen zudem einem größeren Anwenderkreis unentgeltlich zur Verfügung. Crowd Sourcing ist insofern ein kosteneffizientes Mittel der Sprachbewahrung. Die Neuen Medi-

Jürgen Rolshoven | Köln ●

PIÙ ARTICOLI SU QUESTO TEMA:  
[WWW.BABYLONIA.CH](http://WWW.BABYLONIA.CH) > ARCHIVIO TEMATICO >  
[SCHEDA 3](#)

<sup>1</sup> Das bekannteste Beispiel dafür ist die Wikipedia.

### Jürgen Rolshoven

geb. 1952, Computerlinguist und Romanist, Professur für Linguistische Datenverarbeitung und Romanische Philologie 1987-1990 in Bamberg, seit 1990 Professur für Sprachliche Informationsverarbeitung an der Universität zu Köln, u.a. mehrere Drittmittelprojekte zum Bündnerromanischen.



# Die Neuen Medien sind interaktiv. Damit bieten sich neue, kollaborative Formen der Wissensgewinnung und der Wissensverbreitung an. Gerade für die Gewinnung von Wissen über kleine Sprachen eröffnen sich damit neue Möglichkeiten.

en stehen so im Kontrast zu den alten Printmedien: Die mediale Produktion und Distribution ist nicht von Auflagen abhängig und materiell an Papier gebunden. Neue Medien sind smart. Dies möchten wir in diesem Beitrag an den Beispielen der Rätoromanischen Chrestomathie und des interaktiven deutsch-romanischen, romanisch-deutschen Wörterbuchs *Pledari Grond* und seiner Derivate zeigen. Dies verweist nicht nur auf Veränderungen für Sprachen und ihre Sprecher, sondern auch auf Aufgabenverlagerungen für Sprachwissenschaft und Computerlinguistik. Die Computerlinguisten sind herausgefordert, die Informationstechnologien für die interaktive und webbasierte Text- und Wörterbucharbeit zu erstellen.

## Die Rätoromanische Chrestomathie

### Entstehung und Rezeption

Die Rätoromanische Chrestomathie von Caspar Decurtins ist die größte und wichtigste Textsammlung des Bündnerromanischen. Sie wurde zwischen 1896 und 1919 in den *Romanischen Forschungen* veröffentlicht. Decurtins, geboren 1855, 1916 gestorben, war Jurist, Philologe, Sprachwissenschaftler und Politiker. Gemeinsam mit Georges Python war er eine treibende Kraft bei der Gründung der Universität Freiburg, an der er von 1905 bis 1913 die Professur für Kulturgeschichte vertrat. Die Rätoromanische Chrestomathie umfasst 13 Bände. Darin sind die „verschiedenartigsten Quellengattungen vertreten, und die Texte stammen aus allen Bündnerromanischen Gegenden und aus allen [vier] Jahrhunderten...“ (Egloff & Mathieu, 1986: 7). Nach dem Tod von Decurtins veröffentlichten Decurtins Witwe Maria Decurtins Geronimi und Christian Caminada, der spätere Bischof von Chur, die letzten beiden Bände.

### DRC – die Digitale Rätoromanische Chrestomathie

Die Rätoromanische Chrestomathie fand jedoch nur recht begrenzt zu den Bündnerromanen. Um den kulturellen Schatz zugänglich zu machen, fasste die Societad Retorumantscha in den 1980er Jahren den Entschluss, die Chrestomathie als Reprint im Octopus-Verlag von Andreas Joos in Chur herauszugeben. Peter Egloff und Jon Mathieu erschlossen das Werk durch ei-

nen ausgezeichneten Registerband (Egloff & Mathieu, 1986).

Das deutsche DigiZeitschriften-Projekt ([www.digizeitschriften.de](http://www.digizeitschriften.de)) – vergleichbar dem Schweizer RetroSeals-Projekt ([www.retro.seals.ch](http://www.retro.seals.ch)) – hat die Rätoromanische Chrestomathie als Faksimile digital erfasst. Es fehlt dort jedoch ein elektronischer Volltextzugriff, d.h. eine tiefendigitalisierte Version. Ein webbasierter Volltextzugriff auf das gesamte Textkorpus schafft neue Grundlagen für die unterschiedlichen kulturwissenschaftlichen Disziplinen, aber auch für die schulische Arbeit mit Sprache(n). Durch die Tiefendigitalisierung wird die Chrestomathie Open Source und damit allen frei zugänglich.

Die Chrestomathie zeichnet sich nicht nur durch inhaltlichen Reichtum aus, sondern auch durch hohe sprachliche Varianz, durch unterschiedliche und veränderliche Verschriftungsprinzipien und durch veränderte Typographie in den mehr als 20 Jahren ihrer Veröffentlichung.

Dies macht eine fehlerfreie optische Zeichenerkennung (OCR) unmöglich. Wegen der hohen Fehlerrate müssen die Texte manuell nachkorrigiert werden. Da die Korrektur der mehr als 8'000 Seiten umfassenden Chrestomathie nicht von einer einzelnen Person geleistet werden kann, setzt die Lösung dieses Problems auf Vernetzung, Interaktivität und kollaboratives Arbeiten, ähnlich, wie es die Online-Enzyklopädie Wikipedia erfolgreich praktiziert. Die Sprecher der Sprachgemeinschaft übernehmen die Textkorrektur. Hierzu sind jedoch webbasierte kollaborative Werkzeuge notwendig. Es war die Aufgabe der Computerlinguisten (aus Köln), diese Softwarewerkzeuge zu erstellen. Damit sind Gegenstand und Aufgaben des Chrestomathieprojekts umrissen, das von der Deutschen Forschungsgemeinschaft (als Institution vergleichbar dem Schweizer Nationalfond) gefördert wurde.

Die Umsetzung von Wiki-Prinzipien erfordert, dass Korrektur- und Veränderungsschritte einsehbar und zurücksetzbar sind. Veränderungen sind zu dokumentieren. Die aktive Auseinandersetzung mit den Texten, die Kommentierung der Arbeit in der Web Community oder auch Fragen müssen durch Kommentarfelder ermöglicht werden.

Im Folgenden wird das System anhand obiger Abbildung aus der Sicht der Kor-

rektoren dargestellt. Es handelt sich um einen webbasierten Editor, der im linken mittleren gelb unterlegten Fenster den OCR-gelesenen Text darstellt; im rechten mittleren Fenster liegt das Faksimile oder digitalisierte Image, das als Grundlage für die OCR diente. Das Textbeispiel ist der älteste romanische Katechismus von Daniel Bonifaci. Der Text im linken Fenster zeigt den angeklickten Namen *Bonifaci* in grünen Lettern. Es geht um die Frage, ob das angeklickte Wort durch die OCR korrekt erkannt wurde. Dieses Wort erscheint im Digitalisat rechts grün eingerahmt. So können Bearbeiter leicht überprüfen, ob das Wort im Text korrekt erkannt wurde. Ist das nicht der Fall, kann der Text links händisch verbessert werden. Durch die optische Verknüpfung von Wörtern im Text mit dem Image des Digitalisats durch (grüne) Unterlegung wird die Arbeit der Korrektoren sehr erleichtert. Die Verknüpfung beruht auf der Auswertung von Positionsinformationen, die beim Einlesen des Textes durch die OCR-Software codiert sind.

Die weiteren Teilfenster unterstützen die Korrektoren bei ihrer Arbeit. Das Suchfenster präsentiert alle Fundstellen des Namens *Bonifaci*. Im Fenster rechts kommunizieren Korrektoren webbasiert miteinander. Hier können Lösungen reflektiert, erfragt, kommentiert oder begründet werden.

Das untere Teilfenster rechts dokumentiert die Korrekturgeschichte, das linke untere Teilfenster unterstützt die Verwendung von Sonderzeichen.

### ARC – die Annotierte Rätoromanische Chrestomathie

Der Wert digitaler Texte kann durch grammatikalische Annotation gesteigert werden. Eine solche Annotation versieht die Wörter des Textes mit grammatikalische Angaben, die über die Wortart, über Numerus, Kasus oder Tempus u.a.m. Auskunft geben. Bei der Annotation kommen unterschiedliche Verfahren zum Einsatz. Häufig wird händisch ein Mustertext annotiert; diese Annotationen werden dann mit Hilfe statistischer Verfahren ausgewertet und auf das gesamte Korpus übertragen. Damit besteht das Vorgehen aus zwei Phasen, einer Trainings- und einer Produktionsphase. Im vorliegenden Fall wurde dieser Weg mit Hilfe eines spezialisierten Programms für grammatische Auszeichnung, eines Taggers, beschritten. Auch Tagger liefern in



Abb. 1.

Der webbasierte Editor im DRC-Projekt

der Produktionsphase keine fehlerfreien Ergebnisse. Daher bedarf es händischer Eingabe für die Erstellung des Mustertextes in der Trainingsphase und für die Korrektur der in der Produktionsphase ausgezeichneten Texte.

Hierfür wurde der oben beschriebene webbasierte Korrektureditor in dem Folgeprojekt „Annotierte Rätoromanische Chrestomathie“ erweitert.

Er präsentiert sich den Korrektoren und Annotatoren in folgender Weise (Siehe Abb. 2).

Abb. 2: ARC-Editorfunktionen: Das blau unterlegte Wort *Catechismus* wird als Nomen (NN im Rahmen „Edit a word“) ausgezeichnet (getaggt).



### Das Pledari Grond

Die Digitale Rätoromanische Chrestomathie und die Annotierte Rätoromanische Chrestomathie haben gezeigt, dass kollaborative webbasierte Projekte für eine Minderheitssprache erfolgreich durchgeführt werden können. Sie führen neue Formen der Einbindung und

Da die Korrektur der mehr als 8000 Seiten umfassenden Chrestomathie nicht von einer einzelnen Person geleistet werden kann, setzt die Lösung dieses Problems auf Vernetzung, Interaktivität und kollaboratives Arbeiten, ähnlich, wie es die Online-Enzyklopädie Wikipedia erfolgreich praktiziert.

der Partizipation von Sprechern und Sprachgemeinschaften ein und unterstützen kleine kulturelle Gemeinschaften in ihrem Bestreben, ihre Sprachen zu schützen und zu fördern. Im Gegensatz zu traditionellen Massenmedien, deren kostenintensive Produktion einen großen Konsumentenkreis erfordert – eine Vorbedingung, die die geringe Sprecherzahl von Kleinsprachen nicht erfüllen kann – bieten die neuen kollaborativen Medien durch Crowd Sourcing die Chance, nahezu kostenfrei zu produzieren. Gleiches gilt auch für die Distribution via Web. Vor diesem Hintergrund liegt es nahe, Produktions- und Distributionsschemata der digitalen Chrestomathie auf Lexika zu übertragen. Dabei geht es um das Pledari Grond (PG), die umfangreichste Sammlung von Wörtern und Ausdrücken des Rumantsch Grischun und ihrer Überset-

zung ins Deutsche (bzw. auch aus dem Deutschen). Zunächst 1993 in Buchform veröffentlicht, ist das PG seit 2006 auch als Online-Wörterbuch verfügbar ([www.pledarigrond.ch](http://www.pledarigrond.ch)). In Zusammenarbeit von Lia Rumantscha und Sprachlicher Informationsverarbeitung wurde eine neue Version des Online-Wörterbuchs erarbeitet. Dabei werden die Möglichkeiten der Nutzer erweitert: sie sind aufgerufen, an der Entwicklung des PG teilzuhaben, indem sie Vorschläge zur Änderung, Erweiterung oder Neuaufnahme von Einträgen direkt in das Online-Wörterbuch eintragen. Nach einer redaktionellen Prüfung durch Mitarbeiter der Lia Rumantscha werden diese Vorschläge dauerhaft übernommen. Auf diese Weise kann sich die Sprechergemeinschaft aktiv und gemeinschaftlich am weiteren Auf- und Ausbau des PG beteiligen.

*Giovanni Segantini, La natura, 1898-99.*



Das PG wurde außerordentlich positiv aufgenommen; an Werktagen erfolgen fast 1000 Nutzungszugriffe täglich. Auch hat sich das Konzept der partizipativen Erweiterung als über Erwarten positiv ausgewirkt (s. dazu v.a. Neufeind, Telli 2016), eine Erhebung der Nutzungsdaten hat einen nahezu 30% Anteil der Nutzung durch *mobile devices* (Smartphones u.ä.) ergeben. Die hohe Akzeptanz dürfte auch in der für solche Geräte optimierten Benutzeroberfläche liegen und somit Zielgruppen ansprechen, bei denen diese Geräte mit positiven Konnotationen intensiv genutzt werden.

Das Konzept des auf dem Rumantsch Grischun beruhenden PG wird momentan erweitert für die Wörterbücher der Idiome, d.h. für das Sutselvische, das Engadinische, das Surselvische und das Surmeirische. Damit werden durch den Einsatz angepasster Informationstechnologien auch potentielle Konflikte über den Status des Rumantsch Grischun und der Idiome entschärft.

Das lexikographische Material der elektronischen Wörterbücher kann in weiteren Zusammenhängen genutzt werden: als Grundlage für Orthographiekorrektursysteme, die momentan für das Rumantsch Grischun und für die Idiome im Hinblick auf die gängigen Textverarbeitungssysteme wie Word und Open Office entwickelt werden.

### Sprache, Wortschatz, Sprachdidaktik

Mit der digitalen Rätomanischen Chrestomathie, der Annotierten Rätomanischen Chrestomathie und dem Pledari Grond stehen der Öffentlichkeit (und damit auch allen Bildungseinrichtungen) sprachliche Daten in beträchtlichem Umfang Open Source und ubiquitär zur Verfügung. Das erleichtert sicherlich textbasierte Unterrichtsvorbereitung. Daraus folgt auch die Möglichkeit einer weitergehenden und spezialisierten Verwendung. Die Nutzung dieser Daten wird durch einfache und effiziente Benutzer- und Zugriffsflächen erleichtert.

Die Chrestomathie ermöglicht so in ihrem sprachlichen und thematischen Reichtum neue und erweiterte Formen für den Sprach- und Sachunterricht. Ein zusätzlicher Anreiz könnte gerade für Jugendliche darin bestehen, dafür ihre Smartphones zu nutzen.

Die folgenden Überlegungen verstehen sich nicht als konkrete Anleitungen zur Unterrichtsplanung, sie bescheiden sich mit dem Ziel, eine Basis für eine Kommunikation zwischen Praktikern in den Schulen, an Universitäten und Volkshochschulen und Sprachwissenschaftlern und Computerlinguisten zu schaffen.

Bezüglich schulischer Lernziele einige Anregungen: In der Chrestomathie finden sich zahlreiche Beispiele für unterschiedliche Textgattungen (die in der ARC thematisch annotiert sind): Briefe, Gedichte, Novellen, Märchen, Legenden, Dramen. Im Bereich Schriftlichkeit bietet sich der Vergleich von Orthographien bei Texten z.B. des Surselvischen 1700 und nach 1850 an.

Bei Wortarten können beispielsweise anhand von Textvorkommen – etwa für *negin(a)* – Kriterien für eine Zuordnung erarbeitet werden.

Mit Hilfe des ARC-Editors kann eine solche Aufgabe auch in interaktiver Gruppenarbeit unter Einsatz der Kommentarfunktion im Rahmen von Unterrichtsprojekten durchgeführt werden.

Im fachübergreifenden Sachunterricht Biologie können die Texte der Chrestomathie nach Wildtieren in Märchen durchsucht werden, z.B. nach *uolp* (Fuchs), *tgaper* (Rabe), *uors* (urs) (Bär), *tschess* (Alpenadler). Als sprachliche Hilfe stehen dabei das Pledari Grond und seine Derivate zur Verfügung, auch im Hinblick auf Erweiterbarkeit.

### Digitale Desiderata

Traditionelle Printmedien sind vor allem bei kleinen Auflagen teuer. Die Kosten für Produktion, Material und Distribution sind hoch. Diese Kostenfaktoren entfallen aber zu einem guten Teil, wenn bei der Produktion Crowd Sourcing, bei der Distribution das Web genutzt werden. Materialkosten entfallen nahezu gänzlich. Daher wäre es für das Bündnerromanische wünschenswert, diese Prinzipien für möglichst viele sprachliche Daten anzuwenden. Das Romanische verfügt über kulturelle Schätze in seinen Zeitungen, in seiner Buchproduktion, auch in seinem *Dicziunari Rumantsch Grischun*. Diese Schätze sind vorhanden, aber verborgen, da dem Zugriff entzogen. Sie warten darauf, digital verfügbar gemacht zu werden.

### Literatur

Cathomas, B. (1977). *Erkundungen zur Zweisprachigkeit der Rätoromanen. Eine soziolinguistische und pragmatische Leitstudie*. Bern, Frankfurt am Main: Europäische Hochschulschriften I/183. Verlag Herbert Lang.

Egloff, P. & Mathieu, J. (1986). *Rätoromanische Chrestomathie*. Begründet von Caspar Decurtins, Band 15. Chur: Octopus-Verlag / Società Retorumantscha.

Lutz, F. & Rolshoven, J. (Im Druck). Rumantschia Digitala – Moderne bündnerromanische Lexikographie. *Atti del VIII Colloquium retoromanistich 2014*, Cormons, Società Filologica Friulana.

Mondaca, F. & Atanassov, M. (Im Druck). ARC: Annotierte Rätoromanische Chrestomathie. *Atti del VIII Colloquium retoromanistich 2014*, Cormons, Società Filologica Friulana.

Neufeind, C. & Telli, D. (Im Druck). Rumantschia Digitala – das Pledari Grond 2.0. *Atti del VIII Colloquium retoromanistich 2014*, Cormons, Società Filologica Friulana.

Rolshoven, J. (2012). Die Digitale Rätoromanische Chrestomathie. *Ladinia* XXXVI, 119-151.

Rolshoven, J. & Lutz, F. (2014). *Crestomazia Digitala. Akten des V. Rätoromanistischen Kolloquiums Lavin 2011*, 85-103.

Rolshoven, J. & Neufeind, C. (2014). Partizipation und Persistenz. Die Digitale Rätoromanische Chrestomathie. C. Haug & R. Thiele (Hgg.), *Buch - Bibliothek - Region. Wolfgang Schmitz zum 65. Geburtstag*. Wiesbaden: Harrassowitz, O., pp. 341-348.

Rolshoven, J. & Lutz, F. (2015). Die Rätoromanische Chrestomathie – digital und annotiert. In: T.Kahl, J.Kramer & E. Prifti (hgg.), *Romanica et Balcanica. Wolfgang Dahmen zum 65. Geburtstag. Jenaer Beiträge zur Romanistik*, Band 7, München: Akademische Verlagsgemeinschaft, pp. 265-280.

Weinreich, U. (1953, 1963). *Languages in contact. Findings and problems*. New York: Publications of the Linguistic Circle of New York 1 (1953); Den Haag (1963).