

Jürgen Sutter

## **Alemannische Wikipedia – eine Online-Enzyklopädie in alemannischen Dialekten**

Die Alemannische Wikipedia (<http://als.wikipedia.org>) ist eine der rund 280 Sprachversionen der bekannten Internetenzyklopädie. Sie ist heute die umfangreichste Internetseite, die in alemannischen Dialekten geschrieben wird, und mit mittlerweile über 8 Millionen Wörtern wahrscheinlich auch das größte Textkorpus in Alemannisch.

### **Hintergrund und Geschichte**

*“Wikipedia is first and foremost an effort to create and distribute a free encyclopedia of the highest possible quality to every single person on the planet in their own language.”*

So beschrieb Wikipedia-Gründer Jimmy Wales im Jahr 2005 Ziel und Vision der Online-Enzyklopädie. Die Idee einer „Enzyklopädie in allen Sprachen“ führte im Lauf der Jahre dazu, dass nicht nur Versionen in verbreiteten Standardsprachen online gegangen sind, sondern auch solche in weiteren nicht-standardisierten Regional- und Minderheitensprachen, so war zum Beispiel die dritte Sprachversion überhaupt nach Englisch und Deutsch eine Ausgabe in Katalanisch.

Im September 2003 stellte der elsässische Informatiker Alexis Dufrenoy bei der US-amerikanischen Wikimedia Foundation einen Antrag auf Einrichtung einer Wikipedia auf Elsässisch. Nach zwei Monaten Online-Diskussionen ging die Elsässische Wikipedia schließlich am 13. November 2003 online. Dufrenoy zog dann aber nach Paris und fand aus beruflichen Gründen keine Zeit mehr, Artikel zu schreiben (SUTTER 2011). Nach einem Jahr entdeckten Dialektsprecher aus Südbaden, der Schweiz und dem Allgäu das mittlerweile eingeschlafene Projekt und schlugen vor, es zu einer Enzyklopädie aller alemannischen Dialekte zu erweitern. Heute, elf Jahre nach der Gründung, hat die Alemannische Wikipedia einen engeren Kreis von rund zehn Autoren, die regelmäßig schreiben, und etwa fünfzig, die seltener etwas beitragen. Dabei handelt es sich um Autoren aus allen sechs europäischen Ländern, in denen alemannische Dialekte gesprochen werden.<sup>1</sup>

---

<sup>1</sup> Zur regionalen Verteilung der Dialekte siehe unten Tab. 1 und 2



Abb. 1 Startseite der Alemannischen Wikipedia am 3. Oktober 2014

Neben der Alemannischen Wikipedia gibt es heute im Raum Deutschland-Schweiz-Österreich eine plattdeutsche, eine nordfriesische, eine saterfriesische, eine obersorbische, eine niedersorbische, eine ripuarische, eine luxemburgische, eine pfälzische, eine bairische und eine rätoromanische Ausgabe sowie vorläufige Testversionen in Ostfränkisch und Schlesisch.<sup>2</sup> Angrenzend an den deutschsprachigen Raum findet sich schließlich eine ganze Reihe von Wikipedien in weiteren germanischen und romanischen Regionalsprachen (z. B., Limburgisch, Franko-Provenzalisch, Lombardisch u. a.), siehe hierzu folgende Abbildung:

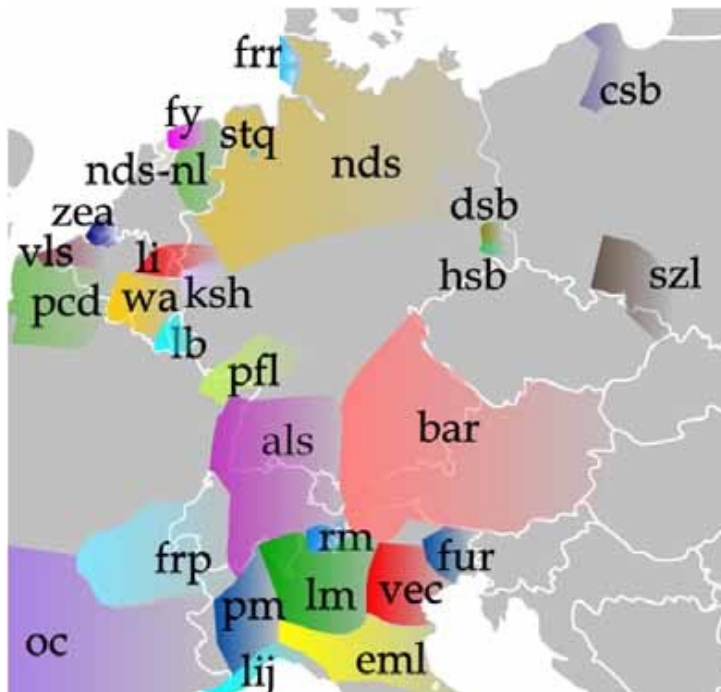


Abb. 2 Regionalsprachen und Dialekte in Mitteleuropa, in denen eine eigene Wikipedia-Sprachversion existiert. Die Buchstabenkombination bezeichnet den Code, mit dem die URL die jeweiligen Seite beginnt: <http://als.wikipedia.org>, <http://bar.wikipedia.org>, <http://nds.wikipedia.org> usw.

<sup>2</sup> Hinzu kommt eine Wikipedia in Pensilfaanisch.

## Inhalte und Textsorten

Zentraler Inhalt sind die enzyklopädischen Artikel, in denen sich der geschriebene Dialekt einer wissenschaftlichen Sachprosa nähert. Charakteristisch ist hier die Übernahme vieler Fachwörter aus der Standardsprache. Diese werden in der Regel lautlich angepasst an den jeweiligen Dialekt, in dem der Artikel geschrieben ist, z. B. *Büteldier* (Beuteltiere), *Riibluut* (Reibelaut), *Striichinstrumänt* (Streichinstrument), *Wüsseschaftstheorii* (Wissenschaftstheorie).

Der Satzbau in den Artikeln ist oft sehr stark am Satzbau standardsprachlicher geschriebener Texte orientiert:

*E Bezugssüsteem isch in dr Füsik e Gebild us Ruum und Zit, wo s brucht, zum äidütig und vollständig z beschriibe, wie sich ortsabhängigi Grössene verhalte. D Laag und d Beweegig vo füsikalische Körper chönne nume relativ zum ene Bezugssüsteem aagee wärde. Zum e Bezugssüsteem z definiere, weelt mä e Bezugspunkt, legt d Ruumrichdige fest und bestimmt e füsikalische Brozäss, für zum d Zit mässe.*

Eine kleine Auswahl an Artikel liegt auch in gesprochener Form vor, in der Regel vorgelesen vom Artikelautor.

Zu allen Artikeln in der Wikipedia kann eine eigene Diskussionsseite angelegt werden, auf der über Form und Inhalte diskutiert werden kann. Auch jedem Autor mit angemeldetem Account steht eine Diskussionsseite zur Verfügung, auf der er direkt angesprochen werden kann. Weiterhin gibt es eine Reihe von zentralen Forumsseiten, auf denen grundlegende Debatten über Projektziele, sprachliche Formen und Schreibweisen, Richtlinien usw. geführt werden.

Die Autoren der Alemannischen Wikipedia diskutieren auf diesen Seiten miteinander in aller Regel im Dialekt, auch Sprecher sehr unterschiedlicher alemannischer Dialekte (z. B. Elsässer, Schweizer oder Schwäbischsprecher). Die Sprache auf den Diskussionsseiten steht dem gesprochenen Dialekt meist sehr nahe, was auch in ähnlicher Weise bekannt ist aus Untersuchungen zur Sprache in Chats und Sozialen Netzwerken (SIEBENHAAR 2003, SIEBENHAAR 2005).

Wenn Nicht-Dialektsprecher in die Diskussion involviert sind, wechseln die Autoren auf Deutsch oder auch je nach Sprachkenntnissen der beteiligten Autoren auch auf Englisch, Französisch usw.

In kleinerem Umfang findet man in der Alemannischen Wikipedia auch *Dialektliteratur*. Da sämtliche Inhalte der Wikipedia unter einer freien Lizenz stehen, können hier nur solche Texte hochgeladen werden, deren Copyright abgelaufen ist, d. h. der Autor muss mindestens 70 Jahre tot sein. Vertreten sind unter anderem Werke der Schriftsteller Pietro Axerio Piazza, Michel Buck, Otto Gittinger, Paul Haller, Johann Peter Hebel, August Lustig, Sebastian Sailer, Rudolf von Tavel und Louis Zumstein.

Auf eigenen Projektseiten werden *Namen* gesammelt, neben dialektalen Ortsnamen auch Tier-, Pflanzen- und Pilznamen. Weiterhin gibt es Sammlungen mit Phraseologismen und auch kleinere *Wortschatzsammlungen*

Am 30. Juni 2014 hatte die Alemannische Wikipedia insgesamt 16.939 Artikel. Davon waren 15.301 geschrieben von insgesamt 518 verschiedenen „Benutzern“, d. h. Autoren, die sich mit einem eigenen Account angemeldet haben, und 1.638 geschrieben von insgesamt 864 verschiedenen sogenannten „IP-Adressen“, d. h. Autoren, die sich nicht mit einem Account angemeldet haben und deren Beiträge unter der IP-Adresse gespeichert werden, die ihnen von ihrem jeweiligen Internetprovider zugewiesen wurde.

Eine Auszählung des Artikelbestandes von Ende Juni 2014 ergab folgende Verteilung auf alemannische Dialekträume. Ausgewertet wurden hierfür 15.464 von den insgesamt 16.939 Artikeln (91,3 %):

Tab. 1 Anteile einzelner Dialekträumen am Artikelbestand der Alemannischen Wikipedia

Dialektraum	Anzahl Artikel	Anteil [%]
Südalemannisch	12.389	80,1
Oberrheinalemannisch	2.377	15,4
Schwäbisch	555	3,6
Bodenseeelemannisch	143	0,9

Eine Auszählung der Artikel nach Ländern ergab folgende Tabelle:

Land	Anzahl Artikel	Anteil [%]
Schweiz	7.578	49
Deutschland	5.835	37,7
Frankreich	1.947	12,6
Österreich	89	0,6
Italien	9	0,1
Liechtenstein	6	0,04

Tab. 2 Anteile einzelner Herkunftsländer am Artikelbestand der Alemannischen Wikipedia

## Mögliche sprachwissenschaftliche Fragestellungen

Bisher wurden die Texte der Alemannischen Wikipedia nur selten von Sprachwissenschaftlern analysiert. Eine Ausnahme bilden zum Beispiel Untersuchungen der Arbeitsgruppe um Yves Scherrer an der Universität Genf.<sup>3</sup>

Sämtliche Inhalte der Alemannische Wikipedia können in Form eines xml-Dumps heruntergeladen werden<sup>4</sup> und dann auf einem lokalen Rechner weiter analysiert werden. Hierbei gibt es die Möglichkeit, die Daten nur mit der jeweils aktuellsten Version aller Seiten herunterzuladen oder auch mit der gesamten Versionsgeschichte und somit mit allen Änderungen, die je gemacht wurden. Zur sprachlichen Analyse der Artikel und ihrer Versionsgeschichte wurden von Wikipedianern bereits viele Tools entwickelt, die zum freien Download angeboten werden.

Im Folgenden sollen einige mögliche sprachwissenschaftliche Fragestellungen aufgezeigt werden, die sich anhand des Textkorpus der Alemannischen Wikipedia untersuchen lassen.

Wer schreibt in der Alemannischen Wikipedia und warum? Wie auch die deutschsprachige Wikipedia wird das Projekt von Freiwilligen geschrieben. Welche Beweggründe veranlassen Dialektsprecher, an einer derartigen Enzyklopädie mitzuschreiben?

Wie werden die Dialekte verschriftet? Welche Diskussionen werden zur Schreibweise geführt?

Da alle gespeicherten Änderungen mit einer Zeitangabe versehen sind, kann auch bei einzelnen Autoren die Entwicklung der Schreibweise im Lauf der Zeit analysiert werden.

Gibt es im Lauf der Jahre Tendenzen zu einer Vereinheitlichung der Schreibweisen?

Welche Tendenzen zum Ausbau des Alemannischen lassen sich beobachten?

Welche Fachwörter werden aus der Standardsprache übernommen, wo werden eigene Dialektwörter verwendet?

Welche grammatikalischen Strukturen werden aus der Standardsprache übernommen?

Welche grammatikalischen Strukturen werden aus dem gesprochenen Dialekt übernommen?

Unterscheiden sich hier Übersetzungen und selbständig geschriebene Artikel? Rund ein Viertel der Artikel wurden aus der deutschsprachigen Wikipedia übersetzt.

Wie wirkt sich der kollektive Schreibprozess aus? In der Wikipedia schreiben vielfach mehrere

---

<sup>3</sup> Siehe z. B. SCHERRER/RAMBOW 2010, SCHERRER 2011, SCHERRER 2013. Andere Arbeiten, die die Alemannische Wikipedia als Textkorpus herangezogen haben sind z. B. BERNHARD/LIGOZAT 2013, HOLLENSTEIN/AEPLI 2014 oder auch das Neue Baseldeutsch Wörterbuch (CHRISTOPH MERIAN VERLAG 2010), das Belegsätze aus Artikeln der Wikipedia verwendet, diese interessanterweise dann aber in die Schreibweise des Wörterbuchs übertragen, zum Teil sogar Belegsätze in anderen Dialekten auf Baseldeutsch übersetzt hat.

<sup>4</sup> Die Dumps können unter <https://dumps.wikimedia.org/backup-index.html> heruntergeladen werden.

Autoren an einem Artikel. Welche Schreibweisen, welche Dialektformen bleiben konstant, welche werden geändert? Welche Wörter werden als zu standardnah eingestuft und durch andere, „dialektalere“ Wörter ersetzt. Welche Formen werden überhaupt als „falsch“ eingeschätzt und entsprechend korrigiert?

Was passiert, wenn Autoren einen Artikel ganz in ihren eigenen Dialekt umschreiben? Welche Formen werden hier geändert, welche bleiben dagegen erhalten? Hier lassen sich zum Beispiel Rückschlüsse darauf ziehen, welche Formen als abweichend wahrgenommen werden.

Da die Alemannische Wikipedia mittlerweile einen sehr großen Textkorpus bildet, lassen sich damit auch Untersuchungen zur statistischen Häufigkeit von bestimmten Wörtern und Formen durchführen. Hier wäre es interessant zu erforschen, ob sich die Unterschiede zur Häufigkeit von bestimmten Formen im Vergleich zu Korpora gesprochener alemannischer Dialekte ähnlich verhalten zu den Unterschieden zwischen geschriebener und gesprochener Standardsprache.

Auch inhaltliche Untersuchungen können interessante Aspekte beleuchten. Welche Themen werden dargestellt? Wie unterscheidet sich die Darstellung von bestimmten Themen von derjenigen der deutschsprachigen Wikipedia?

Ein Nachteil bei der Untersuchung von Texten der Alemannischen Wikipedia ist, dass in vielen Fällen nur wenige Informationen zu den jeweiligen Autoren vorliegen. Die meisten „Wikipedianer“ schreiben unter einem anonymen Benutzeraccount. Wie oben bereits beschrieben, kann zu jedem Autor Kontakt über dessen Benutzer-Diskussionsseite aufgenommen werden, viele haben auch eine Mailfunktion aktiviert, so dass sie auch per Mail angeschrieben werden können. Manche Autoren legen allerdings Wert auf ihre Anonymität.

Während die Sprache, die auf Diskussionsseiten verwendet wird, dem gesprochenen Dialekt meist sehr nahe steht, stellt sich natürlich die Frage, ob die Sprache, die in den enzyklopädischen Artikeln verwendet wird, überhaupt Dialekt ist. Nun, die Antwort auf diese Frage hängt davon ab, wie man den Begriff „Dialekt“ definiert. Wenn man unter Dialekt eine Varietät der gesprochenen Sprache mit einer lokalen bzw. regionalen Reichweite versteht, dann ist die Sprache in den Artikeln ähnlich wie diejenige der Dialektliteratur per definitionem kein Dialekt. Wenn man hingegen Dialekt definiert als Varietät mit einer lokalen bzw. regionalen Reichweite, dann zählen die Artikel der Alemannischen Wikipedia durchaus zu einer modernen Form des Gebrauchs von alemannischen Dialekten im Internet.

## Literatur

BERNHARD, DELPHINE; LIGOZAT, ANNE-LAURE: Es esch fäscht wie Ditsch, oder net? Étiquetage morphosyntaxique de l'alsacien en passant par l'allemand. Taln-Récital 2013, 17-21 Juni, Les Sables d'Olonne

BERNHARD, DELPHINE; LIGOZAT, ANNE-LAURE (2013): Hassle-free POS-Tagging for the Alsatian Dialects. In: Marcos Zampieri, Sascha Diwersy (Hrsg.): Non-Standard Data Sources in Corpus Based-Research. Shaker (ZSM-Studien, Schriften des Zentrums Sprachenvielfalt und Mehrsprachigkeit der Universität zu Köln 5)

CHRISTOPH MERIAN VERLAG (Hrsg.) (2010): Neues Baseldeutsch-Wörterbuch. Christoph Merian Verlag, Basel

HOLLSTEIN, NORA; AEPLI, NOEMI (2014): Compilation of a Swiss German Dialect Corpus and its Application to PoS Tagging. Proceedings of the 1st Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects. Dublin

SCHERRER, YVES (2013): Continuous variation in computational morphology. The example of Swiss German. In: TheoreticAl and Computational MOrphology: New Trends and Synergies (TACMO)

SCHERRER, YVES (2011): Morphology generation for Swiss German dialects, SFCM, Zürich

SCHERRER, YVES; RAMBOW, OWEN (2010): Word-based dialect identification with georeferenced rules. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 1151– 1161

SIEBENHAAR, BEAT (2003): Sprachgeographische Aspekte der Morphologie und Verschriftung in schweizerdeutschen Chats. In: Linguistik online 15, S. 125–139

SIEBENHAAR, BEAT (2005): Die dialektale Verankerung regionaler Chats in der deutschsprachigen Schweiz. In: Eggers, Eckhard; Schmidt, Jürgen Erich und Stellmacher, Dieter (Hg): Moderne Dialekte – Neue Dialektologie. Stuttgart: Steiner (ZDL-Beiheft 130), 691–717

SUTTER, JÜRGERN (2011): Alemannische Wikipedia. In: Alemannisch dunkt üs guet I/II, 13-14

Jürgen Sutter

Seeheimer Straße 27a

64297 Darmstadt

[sutter.juergen@gmail.com](mailto:sutter.juergen@gmail.com)