

Tiroler Ortsdialekte online. Von handschriftlichen Aufzeichnungen zum digitalen Sprachatlas

Das Projekt „Tiroler Ortsdialekte online“ des Tiroler Dialektarchivs hat zum Ziel, analog vorhandene, bis in die 1970er-Jahre zurückreichende handschriftliche Dialektaufzeichnungen aus 120 Tiroler Gemeinden zu digitalisieren und aufzubereiten. Daraus ausgewählte Sprachdaten sollen letztlich in einem Online-Sprachatlas zugänglich gemacht werden.¹ Die Heterogenität des Ausgangsmaterials, die Komplexität der Kategorisierung sowie die Limitierung, die das Kartierungstool hervorruft, stellt das Projektteam vor eine Reihe von Herausforderungen.

Produktion: Digitalisierung und Archivierung

In einem ersten Schritt wurden zur Langzeitarchivierung Scans der analogen Bestände angefertigt. Die eigentliche Digitalisierung des von Hand geschriebenen, in der Lautschrift Teuthonista aufgezeichneten Materials erfolgt nun durch die Abschrift in eine Sequenz von Unicodezeichen. Diese Zeichenfolge wird an der Oberfläche durch einen sich derzeit in Entwicklung befindlichen, an der Schreibweise der Exploratoren orientierten Font dargestellt, dessen Zeicheninventar aufgrund der verbleibenden Unvorhersehbarkeit des analogen Materials offen bleiben muss. Zudem gilt es, über zehn verschiedene, während der Befragung der Gewährspersonen simultan mitgeschriebene Aufzeichnungen zu interpretieren und zu nivellieren. So gibt es beispielsweise Fälle, in denen gleiche Zeichen für verschiedene Laute stehen (<v> für /v/ oder /ʋ/) und umgekehrt ein Laut durch verschiedene Zeichen dargestellt wird (/v/ als <v> oder <f>). Alle Sprach- und Metadaten laufen in einer auf Semantic MediaWiki basierenden Datenbank zusammen, die sowohl deren Speicherung als auch kartographische Darstellung ermöglicht.

Analyse: Kategorisierung

Der Kartengenerierung gehen mehrere Abstrahierungsschritte voraus: Grundlage ist der Originalbeleg, der den analog vorhandenen Beleg digital so exakt wie möglich wiedergibt. Zur besseren Lesbarkeit – unter anderem für Laien – wird daraus eine vereinfachte Form mit nur w-nigen, bekannten Sonderzeichen abgeleitet. Dies geschieht allerdings (noch) nicht automatisiert. Die aus diesem Prozess sukzessiv erarbeiteten Vereinfachungsregeln decken den Großteil der auftretenden Fälle ab, Sonderfälle bereiten aber immer wieder Schwierigkeiten, so etwa die Vereinfachung von ð̃nājə ‚annähen‘ zu oʰnaaje, oo(u)(n)naaie etc. Es ist muss hier ein Kompromiss zwischen Lesbarkeit und Lauttreue gefunden werden. In weiterer Folge wird jeder Beleg auf zwei Ebenen lemmatisiert, indem das belegnahe Stichwort (BSW), das unmittelbar mit dem Originalbeleg bzw. dem vereinfachten Beleg verknüpft ist, einem standardisierten Stichwort (SSW) zugewiesen wird. Diese doppelte Lemmatisierung macht es möglich, Karten unterschiedlicher Abstraktionsgrade zu erstellen. So können phonetisch-phonologische, morphologische, morphosyntaktische und lexikalische Phänomene differenziert abgebildet werden. Dem übergeordnet steht die Bedeutung. Diese vereinfachende Abbildung der komplexen Sprachrealitäten führt allerdings zu Problemen bei der Abgrenzung der Ebenen. Beispielsweise stellt sich die Frage, ob die vereinfachten Belege *ferménte*, *hurmenta* und *uremente* ‚Murmeltier‘ als separate Lexeme oder als phonetische Realisierungen ein und desselben Lexems zu betrachten sind. Abhängig von dieser Entscheidung ergeben sich andere Kartenbilder. Dabei zieht jede nachträgliche Neubewertung oder Regeländerung weitreichende Konsequenzen nach sich.

Visualisierung: Kartendarstellung

Die kartographische Darstellung der Sprachdaten mittels des standardmäßig in Semantic MediaWiki implementierten Google-Maps-Tools genügte anfänglich unseren Anforderungen, Karten schnell und einfach zu erstellen. Mit fortschreitender Eingabe kristallisierte sich allerdings dessen beschränkter Funktionsumfang heraus, der es etwa nicht zulässt, Symbole in unterschiedlichen Farben und Formen frei zuzuweisen oder Flächenkarten zu erstellen, die nicht auf einem Gemeindegrenzen-Layer basieren (sondern etwa auf einem Layer, der den Dauersiedlungsraum darstellt). Deshalb ist es in naher Zukunft angedacht, funktionsfähigere Kartentools zu testen und andere Möglichkeiten der Darstellung abzuwägen.

1 Vgl. ähnliche Projekte wie

- BayDat (<http://www.baydat.uni-wuerzburg.de:8080/cocoon/baydat/>),
- ALD – Atlante linguistico del ladino dolomitico e dei dialetti limitrofi (https://www.sbg.ac.at/rom/people/proj/ald/ald_home.htm),
- VIVALDI – Vivaio Acustico delle Lingue e dei Dialetti d'Italia (<https://www2.hu-berlin.de/vivaldi>)
- SyHD: Syntax hessischer Dialekte (www.syhd.info)

(alle vier ebenfalls basierend auf noch nicht in Sprachatlanten veröffentlichtem Material),

- VerbaAlpina (<https://www.verba-alpina.gwi.uni-muenchen.de>, basierend auf bereits veröffentlichtem Sprachatlas-Material und Material aus Neuerhebungen mit Hilfe von social software) oder
- Sprachatlas der deutschen Schweiz – Digitalisierte Dialektkarten (<http://www.dialektkarten.ch>, basierend auf bereits veröffentlichtem Sprachatlas-Material)

[alle eingesehen am 28.07.2017]