

Datengewinnung mittels Crowdsourcing im Dienste der Sprachwissenschaft - Die virtuelle Forschungsumgebung VerbaAlpina

Groß angelegte Sprachatlasprojekte liefern naturgemäß eine Momentaufnahme einer Sprache. Die Befragungen kontinuierlich weiterzuführen, erübrigt sich allein schon wegen des organisatorischen wie finanziellen Aufwands. Auch beschränken sich die erhobenen Daten jeweils auf isolierte geographische Räume. Mit Datenbank- und Webtechnologie bietet sich die Möglichkeit Daten aus verschiedenen Quellen zusammenzuführen und durch indirekte Sprachdatenerhebungen im Internet sowohl areale Lücken zu füllen als auch den Datenbestand in diachroner Hinsicht auszugleichen.

Das DFG-Projekt VerbaAlpina der Ludwig-Maximilians-Universität München (<https://www.verba-alpina.gwi.uni-muenchen.de>) entwickelt seit Ende 2014 eine virtuelle Forschungsumgebung mit dem Ziel, Sprachdaten des Alpenraums länderübergreifend zu sammeln und verfügbar zu machen. Seit dem Frühjahr 2017 wird dialektale Lexik zum Themenbereich Alm-/Milchwirtschaft mittels Crowdsourcing gesammelt. Über eine eigens für diesen Zweck angelegte Web-Applikation (https://www.verba-alpina.gwi.uni-muenchen.de/en/?page_id=1741) können Nutzer das Vorhaben unterstützen, indem sie ihre Belege zu bestimmten almwirtschaftlichen Konzepten beisteuern. Dieses Material wird dann in eine Datenbank eingespeist und mit Sprachdaten aus Sprachatlanten entsprechend verknüpft. In den weiteren Bearbeitungsschritten werden die Daten so angepasst, um eine weitgehende Vergleichbarkeit zu älteren Erhebungen herzustellen. Bei der Analyse bietet das Online-Portal differenzierte Möglichkeiten der Visualisierung.

Der Vortrag beleuchtet anhand konkreter Beispiele die verschiedenen Stadien, die für die Erhebung, Verarbeitung und Darstellung der Sprachbelege notwendig sind. Dabei werden neben den theoretischen sprachwissenschaftlichen Aspekten auch die technischen Herausforderungen und deren Lösung behandelt.

Den Anfang bildet die Betrachtung der speziellen Anforderungen, die eine Erhebung von Daten im Internet an die zu entwickelnde Anwendung stellt. Das sind einerseits wissenschaftliche Aspekte, vor allem welche Daten in welcher Form erhoben werden sollen, und andererseits möglichst große Nutzerfreundlichkeit und spezieller Zuschnitt auf die zu erwartende Zielgruppe.

Im Weiteren wird der Prozess der Eingliederung der neu erhobenen Daten in die bestehende Systematik in der Datenbank von VerbaAlpina besprochen. Hierbei wird ein kurzer Einblick in Modellierung und Strukturierung der Kerndaten des Projekts unter Bezug auf die Anforderungen, die sich durch die verschiedenen Datenquellen und Sprachen ergeben, gegeben.

Die Anzeige der Daten mithilfe einer interaktiven Karte stellt einen weiteren wichtigen Teil dar. Besonderer Fokus wird hierbei auf die verschiedenen Formen der Visualisierung gelegt, sowohl vom Standpunkt der Herangehensweise (onomasiologisch vs. semasiologisch) als auch dem der Darstellung (qualitativ vs. quantitativ bzw. topographisch vs. abstrahiert).

Ein letzter Abschnitt befasst sich mit der längerfristigen Verwendung der erhobenen Daten, was sowohl das Thema der Zitierbarkeit eines sich meist täglich ändernden Datenbestands als auch dessen Archivierung einschließt. Ebenfalls eingegangen wird auf der Problem der nachhaltigen Speicherung der Projektdaten, um eine möglichst hohe Nachnutzbarkeit zu gewährleisten (vgl. Thomas Krefeld & Stephan Lücke (2017): Nachhaltigkeit – aus der Sicht virtueller Forschungsumgebungen. Korpus im Text. Version 7 (10.03.2017, 12:27). url: <http://www.kit.gwi.uni-muenchen.de/?p=5773&v=7>).