

## **Problemstellung/Stand der Forschung**

**Sprachatlanten** sind matrixartig organisierte, geolinguistische Datensammlungen, die die innerlinguistische Befindlichkeit einer untersuchten Sprachlandschaft im Sinn eines synchronen Schnittes dokumentieren. Dabei werden auf einer Anzahl  $p$  einzelner Sprachatlaskarten mit einer Anzahl  $N$  im Raum verteilter Messpunkte im Feld erhobene dialektale Reflexe zu vorgegebenen Konzepten abgebildet. In der Romanistik, der Mutterphilologie der Sprachgeographie, wird die Abbildung des innerlinguistischen Datenmaterials traditionellerweise über Volltextkarten in Form phonetischer Transkriptionen bewerkstelligt.

Einzelne **Sprachatlaskarten** können nach bestimmten innerlinguistischen (i.e. phonetischen, lexikalischen u.a.) Kriterien ausgewertet/taxiert werden, um allfällige, hinter den Einzeldaten liegende räumliche Strukturen herauszuarbeiten. In diesem Zusammenhang werden aus der Vielfalt divergierender Antworten pro Atlaskarte einige wenige Antworttypen/Klassen abstrahiert, die jeweils eine, für die Einzeltaxierung gewählte Merkmalsausprägung gemeinsam haben. Diese **Taxierung** geschah bislang bzw. geschieht auch heute noch meist händisch und zwar in der Form, dass der Datenbefund pro Messpunkt aus der Karte gelesen, kodiert und auf eine sogenannte "stumme Karte" übertragen wird. Durch Zusammenführen der Taxierungserträge mehrerer solcher Karten auf sogenannten Dichtekarten kann die Verteilung einzelner linguistischer Merkmale (bzw. Merkmalsausprägungen) im Raum sichtbar gemacht werden.

Die in Sprachatlanten vorhandene zweidimensionale **Datenmatrix** ( $N$  Messpunkte mal  $p$  Karten) kann – im Gegensatz zur oben angesprochenen händischen Auswertung einzelner, i.e. weniger Karten – aber auch dazu herangezogen werden, möglichst hochrangige Ordnungsstrukturen, Muster und Symmetrien im von der untersuchten Sprechergemeinschaft sprachlich bewirtschafteten Raum zu entdecken und zu visualisieren.

Daraus resultierende Ergebnisse können schließlich zur Gewinnung neuer bzw. zur Verifizierung altbekannter Hypothesen herangezogen, also interpretiert und traditionellen Forschungsergebnissen zur Seite bzw. gegenübergestellt werden. Aufgrund der mitunter sehr großen, zu verarbeitenden Datenmengen (idealiter aller in einem Sprachatlas zur Verfügung stehenden Karten) können die o.a. Arbeitsvorgänge jedoch nicht mehr händisch, sondern nur mit massiver EDV-Unterstützung bewerkstelligt werden. Der im Zusammenhang mit einer solchen, exhaustiven Auswertemethodologie angesprochene **Klassifikationsvorgang** besteht im wesentlichen aus drei Arbeitsschritten bzw. Arbeitszielen, nämlich der Datenreduktion, der Ähnlichkeits- bzw. Distanzmessung und der die Mustererkennung ermöglichen Klassenbildung. Seit den 70er-Jahren spricht man in diesem Kontext von *Dialektometrie* als einer Spielart von *Pattern Recognition* bzw. *Automatischer Klassifikation*.

Mit der Ende 1998 erfolgten Publikation des Sprachatlasses des Dolomitenladinischen und angrenzender Dialekte, I. Teil (ALD I), eines sprachkartographischen Werks zum südostschweizerischen und nordost-oberitalienischen bzw. zentralrätoromanischen Dialektraum (Graubünden, Lombardei, Trentino, Ladinien, Veneto, Friaul), liegt nun ein zur Gänze maschinenlesbares, aktuelles sprachgeographisches Corpus vor, dessen dialektometrische Auswertung im Rahmen des beantragten Projekts bewerkstelligt werden soll. Das Untersuchungsnetz des ALD I umfasst 217 Messpunkte, an denen zu rund 1.500 Konzepten basilektale Reflexe erhoben und in phonetischer Umschrift in den Fragebüchern notiert wurden. Alle Explorationsgespräche wurden zudem auf Band aufgezeichnet (und zum Teil digitalisiert), so daß neben dem transkriptorischen auch ein akustisches "Kontrollarchiv" zur Verfügung steht. Der in sieben gedruckten Bänden (4 Kartenbände, 3 Indexbände) publizierte ALD I liegt auch als elektronische Version auf drei CD-ROM vor. Dadurch ist eine der wesentlichsten Voraussetzungen zur dialektometrischen Auswertung bereits *a priori* gegeben.

(Auszug aus dem Projektantrag an den FWF, März 2000)