

MAGNUS BREDER BIRKENES

PHILIPPS-UNIVERSITÄT MARBURG

N-GRAMM-BASIERTE ÄHNLICHKEITSMESSUNGEN ALS DIALEKTOMETRISCHE METHODE: DIE FRAGEBÖGEN DES WENKER-ATLAS

Vortrag am 31.10.18 um 18 Uhr c.t.
in Raum S004 / Schellingstraße 3 (VG)

N-Gramme sind Sequenzen von n Einheiten, z.B. Buchstaben oder Wörtern, und gehören seit geraumer Zeit zum Standardrepertoire der Computerlinguistik. In der Dialektologie oder Dialektometrie repräsentieren sie aber einen nur selten genutzten, wenn nicht gar völlig neuen Zugang (siehe jedoch Dipper/Schrader 2008 für solche Analysen zum Mittelhochdeutschen). Der Vortrag widmet sich einem solchen Ansatz, seinen Stärken und Schwächen, und zwar anhand eines Parallelkorpus, der Marburger Wenkermaterialien.

Mithilfe eines geographisch gleichverteilten Samples von ca. 2.000 transliterierten, maschinenlesbaren Fragebögen – Übersetzungen von jeweils 40 Sätzen – werden Buchstaben-n-Gramme, in diesem Fall Sequenzen von jeweils drei Buchstaben (Trigramme), erstellt und in jedem Fragebogen ausgezählt. Die Frequenzprofile der Bögen werden sodann über ein Ähnlichkeitsmaß miteinander verglichen und mithilfe unterschiedlicher dimensionsreduzierender Verfahren werden Strukturen in den Daten gesucht, um Dialektareale aufzudecken.

Getestet wird die Methode zunächst an einer kleineren Sprache, die in den Wenkermaterialien gut repräsentiert ist, dem Nordfriesischen, bei dem sich bekanntlich auf engem Raum besonders viel Variation findet und zu dem bisher nur wenige quantitative Analysen vorliegen. Dabei zeigt sich eine gewisse Bestätigung für die traditionelle Dialektgliederung des Nordfriesischen (vgl. Århammar 1968), wobei die Situation auf dem Festland weniger eindeutig ist als auf den Inseln. Mithilfe eines Assoziationsmaßes werden prominente Dialektmerkmale für die wichtigsten Dialektverbände automatisch extrahiert. Hier zeigt sich, dass sich der nordfriesische Sprachraum mithilfe zweier Erscheinungen, der Apokope und der Senkung, relativ gut eingrenzen lässt.

Im zweiten Teil des Vortrags werden erste Ergebnisse für die Dialekte des Deutschen gezeigt. Auch hier ergibt sich eine Bestätigung für die am meisten genutzte Dialekteinteilung von Wiesinger (1983), wobei in bestimmten Gebieten (z.B. dem Nordniederdeutschen) Untergruppierungen, die bei Wiesinger nicht vorhanden sind, klar hervortreten, während zwischen anderen wie dem Nordhessischen/Osthessischen und dem Thüringischen die Unterschiede geringfügiger als gedacht zu sein scheinen (s. aber Wiesinger 1970, der das Nord- und Osthessische zum Thüringischen stellt).

Literatur:

Århammar, Nils. 1968. Friesische Dialektologie. In Ludwig Erich Schmitt (ed.), Germanische Dialektologie (ZDL-Beihefte 5), 264–317. Steiner.

Dipper, Stefanie & Bettina Schrader. 2008. Computing Distance and Relatedness of Medieval Text Variants from German. In Text Resources and Lexical Knowledge. Selected Papers from the 9th Conference on Natural Language Processing (KONVENS-08), 39–51. Berlin: de Gruyter.

Wiesinger, Peter. 1970. Phonetisch-phonologische Untersuchungen zur Vokalentwicklung in den deutschen Dialekten: Bd.1 Die Langvokale im Hochdeutschen/Bd. 2 Die Diphthonge im Hochdeutschen. In Ludwig Erich Schmitt & Stefan Sonderegger (Hgg.): Deutscher Sprachatlas: Gesamtdarstellungen: Vokalismus ½ (Studie Linguistica Germanica.) Berlin: De Gruyter u. a.

Wiesinger, Peter. 1983. Die Einteilung der deutschen Dialekte. In: Besch, Werner, Ulrich Knoop, Wolfgang Putschke & Herbert E. Wiegand (Hgg.). 1983. Dialektologie: ein Handbuch zur deutschen und allgemeinen Dialektforschung. Zweiter Halbband (Handbücher zur Sprach- und Kommunikationswissenschaft 1.2.) Berlin & New York: De Gruyter, 807–900.



Vortrag im Rahmen des Linguistischen Kolloquiums
<http://www.lmu.de/linguistisches-kolloquium>