Erstellung eines bayerischen Dialekt-Lexikons mithilfe von Facebook-Sprachdaten und Crowdsourcing

thema:
: Erstellung eines bayerischen Dialekt-Lexikons mithilfe von Facebook-Sprachdaten und Crowdsourcing

art:
: BA

betreuerIn:
: Manuel Burghardt

bearbeiterIn:
: Daniel Granvogl

professor:
: Christian Wolff

status:
: abgeschlossen

stichworte:
: Facebook, Graph API, Korpuslinguistik, Crowdsourcing

angelegt:
: 2013-02-11

Antrittsvortrag:
: 2014-05-26

ZweitgutachterIn:
: N.N.

Textlizenz:
: Unbekannt

Codelizenz:
: Unbekannt

Hintergrund

Bei der Untersuchung von Dialektsprachen wie dem Bayerischen, sind maschinenlesbare Sprachdaten rar: Da Dialekte vornehmlich auf den Bereich gesprochener Sprache beschränkt sind, ist die Erstellung von untersuchbaren Daten dementsprechend aufwendig, da die Sprache zunächst (unter natürlichen Bedingungen) aufgenommen, und anschließend transkribiert werden muss. Außerdem gibt es eine enorme Vielfalt von bayerischen Dialekten, welche eine einheitliche Transkribierung zusätzlich erschweren.

Mit sozialen Netzwerken wie Facebook stehen seit einiger Zeit große Mengen an geschriebener, bayerischer Dialektsprache zur Verfügung. Besonders in thematischen Gruppen wie etwa „Du bist ein echter Chamer wenn…“ (analoge Facebook-Gruppen gibt es für viele bayerische Städte) finden sich mehrere tausend Individuen eines Sprachraums, die vornehmlich im jeweiligen Dialekt über das soziale Netzwerk kommunizieren.

Zielsetzung der Arbeit

Das Ziel der Arbeit ist die Erstellung eines Korpus des Bayerischen: Das Korpus enthält nach Region und Häufigkeit sortierte Dialektwörtern, welche durch Crowdsourcing in das Hochdeutsche übersetzt wurden. Diese Korpus kann dann in Folgeprojekten zum Beispiel zur Erstellung eines bayerischen Sprachatlas herangezogen werden.

Konkrete Aufgaben

Ziel der Arbeit ist es im ersten Schritt über die Facebook Graph-API Sprachdaten aus unterschiedlichen Sprachräumen (Cham, Straubing, etc.) zu sammeln und quantitativ auszuwerten, um etwa herauszufinden welche spezifischen Wörter in bestimmen Regionen besonders häufig vorkommen (z.B. mit Voyant oder den TAPoR-Tools).

Im zweiten Schritt werden mit einem Crowdsourcing-Ansatz die häufigsten Wörter einer Region von den Sprechern „übersetzt“. Die Idee ist hierbei abermals die thematischen Facebook-Seiten „Du bist ein echter XY wenn…“ zu benutzen, und die Gruppenmitglieder spielerisch dazu zu animieren, typische Wörter in das Hochdeutsche zu übersetzen. Motivation für die Benutzer: Durch das Übersetzen typischer Dialektwörter beweisen sie, dass sie z.B. ein „echter Chamer“ sind. Die Übersetzungskomponente sollte als interaktive HTML-Seite realisiert werden, welche dann in der jeweiligen regionalen Facebook-Gruppe verlinkt wird. Die Ergebnisse der Übersetzungen werden in einer XML- oder Datenbankstruktur gespeichert.

Erwartete Vorkenntnisse

Zur Bearbeitung des Themas sind Kenntnisse im Bereich Markup-Sprachen (XML, HTML) und Datenbanken (z.B. SQL) sowie Grundkenntnisse im Bereich empirische Methoden / Statistik erforderlich (obligatorisch). Idealerweise bringt der/die Bearbeiter/in grundlegende sprach- oder kulturwissenschaftliche Kenntnisse mit (optional).

Weiterführende Quellen

- Facebook Graph-API: http://developers.facebook.com/docs/reference/api/
- Du bist ein echter Chamer wenn…: www.facebook.com/groups/ein.echter.chamer/
- Du bist ein echter Straubinger wenn…: http://www.facebook.com/groups/339783279437389/?fref=ts
- Voyant: http://voyant-tools.org/
- TAPoR-Tools: http://taporware.ualberta.ca/~taporware/

From:
https://wiki.mi.ur.de/ - **MI Wiki**

Permanent link:
https://wiki.mi.ur.de/arbeiten/bavarian_facebook

Last update: **12.12.2019 13:58**