

Konzeption und Umsetzung eines webbasierten Informationssystems für kreolische Sprachaufnahmen

Thema:

Konzeption und Umsetzung eines webbasierten Informationssystems für kreolische Sprachaufnahmen

Art:

BA

BetreuerIn:

Manuel Burghardt

BearbeiterIn:

Julia Grötsch

ErstgutachterIn:

Christian Wolff

ZweitgutachterIn:

N.N.

Status:

abgeschlossen

Stichworte:

Digital Humanities, Sprachwissenschaft, Informationssysteme

angelegt:

2017-01-09

Antrittsvortrag:

2017-10-09

Textlizenz:

Unbekannt

Codelizenz:

Unbekannt

Hintergrund

Im Rahmen einer Kollaboration mit der Romanistik liegen Daten für unterschiedliche Kreolsprachen (<https://de.wikipedia.org/wiki/Kreolsprachen>) vor. Die Daten beinhalten einerseits Audiofiles (Interview-Aufnahmen) sowie auch Transkriptionen, Metadaten und Übersetzungen der Audiofiles.

Zielsetzung der Arbeit

Ziel der Arbeit ist die Erstellung einer Webplattform, in der Audiosignal, Transkription, Metadaten und Übersetzung aligniert dargestellt und durchsucht werden können (vgl. Cocoon, DGD).

Einfaches Mockup: <https://dl.dropboxusercontent.com/u/4194636/mockup-creole.png>

Konkrete Aufgaben

- Related work: Wie ist der Stand der Kunst zu Tools und Webseiten, welche gesprochene Sprache aligniert als Audiosignal und als textuelle Transkription darstellen / durchsuchbar machen / etc.; welche wissenschaftliche Literatur gibt es dazu (in den Digital Humanities / Computerlinguistik; einschlägige Tagungen und Konferenzen werden mit dem Dozenten abgestimmt).
- Überführung der Transkriptionen, Metadaten und Übersetzungen, welche aktuell als Worddatei vorliegen, in eine Datenbank oder eine XML-Struktur (einfache Informationsextraktion mit einer beliebigen Programmiersprache, idealerweise Python + regular expressions). Das Datenmodell sollte so flexibel sein, dass man später ggf. auch noch weitere linguistische Informationen zu den Transkriptionen speichern kann (werden in der BA ggf. exemplarisch für ein Beispiel mit angegeben)
- Umsetzung eines Web-Audioplayer, in dem Audio und Transkription parallel abgespielt werden können (erster Prototyp von Alexander Bazo liegt bereits vor unter <https://github.com/alexanderbazo/TranscriptPlayer.js>)
- Umsetzung einer grundlegenden Suchfunktion, die es erlaubt die Transkriptionen nach bestimmten Schlüsselwörtern zu durchsuchen, und gleichzeitig an die entsprechende Stelle im Audiofile springt.
- Benutzerfreundliche Darstellung weiterer Informationen auf der Webseite, bspw. Übersetzungen, Sprecher, etc. (user centered design-Ansatz, d.h. frühes Testen von Prototypen mit Usern aus der Romanistik, und sukzessive Optimierung der Interface). Beim user interface design sollten zunächst bestehende Video-Transkriptionstools (siehe Links weiter unten) analysiert werden, da man sich hier ggf. best practices für die Alignierung anschauen kann.

Erwartete Vorkenntnisse

- Webentwicklung: Datenbanken / JavaScript
- Hilfreich: Python / reguläre Ausdrücke (<http://programminghistorian.org/lessons/understanding-regular-expressions>)
- Kenntnisse in den Kreolsprachen selbst sind nicht erforderlich, grundlegende Französischkenntnisse können aber nicht schaden

Weiterführende Quellen

Literatur

- Jacquin, J. (2016). IMPACT: A tool for transcribing and commenting on oral data, for teaching, learning, and research. *Digital Scholarship in the Humanities*, 31(3), 493– 498. <http://doi.org/10.1093/llc/fqv012>

Bestehende Web-Audioplayer:

- Cocoon; Beispiel für Web-Audioplayse: http://cocoon.huma-num.fr/exist/crdo/display/crdo-CFPP2000_2

- Datenbank für Gesprochenes Deutsch (DGD); Beispiel für Web-Audioplayse:
http://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome

Video-Transkriptionstools:

- ANVIL, <http://www.anvil-software.org>; ELAN, <https://tla.mpi.nl/tools/tla-tools/elan>; CLAN, <http://chilides.psy.cmu.edu/clan>; EXMARaLDA, <http://exmaralda.org>; Praat, <http://www.fon.hum.uva.nl/praat>; Transana, <https://www.transana.com>

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

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

Last update: **01.10.2019 12:30**

