

# Merkblatt zur Abgabe/Archivierung von Sprachdaten zu Studentenprojekten

F. Schiel - 2014-05-09

## Einleitung

Im Rahmen von studentischen Projekten (BA, MA, PhD Dissertationen) werden häufig interessante Sprachdaten erhoben, bearbeitet, annotiert und analysiert.

Damit diese Daten nach Abschluss des Projekts nicht verloren gehen, und für fortführende Projekte/Veröffentlichungen etc. zur Verfügung stehen, wird jeder Student verpflichtet, diese Daten nach Abgabe seiner Arbeit in möglichst übersichtlicher Form am BAS abzuliefern.

Ausgenommen sind Arbeiten, die auf bereits im BAS (oder anderswo) veröffentlichten Daten basieren.

Dieses Merkblatt gibt allgemeine Hinweise, wie zu verfahren ist.

## Vorgaben zu den abgegebenen Daten

Bei der Erstellung sollten Sie allgemeine Regeln zur Erhebung von Sprachdaten beachten (s. z.B. <http://www.bas.uni-muenchen.de/forschung/BITS/TP1/Cookbook/>). Lassen Sie sich von den Spezialisten des BAS beraten (Draxler, Jansch, Schiel).

Um den Arbeitsaufwand für den Studenten möglichst gering zu halten, müssen keine standardisierten Metadatensätze abgegeben werden. Trotzdem sollten vorhandene Informationen zu Sprechern/Aufnahmen in Form von Tabellen dokumentiert sein (s. Muster unten).

Die Dokumentation der Daten sollte Englisch sein (kann aber auch Deutsch sein).

Sprecher sollten ein Formular ausfüllen und unterschreiben, in welchem sie ihre Zustimmung zur Nutzung und (anonymisierten) Weitergabe ihrer Sprachdaten geben; eine Vorlage zu so einem Formular befindet sich in:

[http://www.phonetik.uni-muenchen.de/Bas/BasTemplateInformedConsent\\_de.pdf](http://www.phonetik.uni-muenchen.de/Bas/BasTemplateInformedConsent_de.pdf)

Diese Formulare müssen mit abgegeben werden.

Die Struktur des abgegebenen Sprachkorpus sollte einer Standard-Struktur folgen, wie sie in folgendem Muster-Korpus vorgegeben ist:

/vdata/BAS/IPS-PROJECTS/MusterCorpus :

annot	: annotation files (e.g. TextGrid)
doc	: documentation, at least one README file
data	: speech data in final version as analysed (organized per speaker/recording etc.)
table	: computer readable metadata tables, mostly information per recording session (SESSION.TBL), per recording (CONTENT.TBL) and per speaker (SPEAKER.TBL)
(rawdata)	: raw recording data before processing (if applicable)

## Vorgaben zur Ausarbeitung

In der Ausarbeitung des Studentenprojekts (BAS, MA, PhD Dissertation etc.) sollte im Anhang kurz der Entstehungsablauf der abgegebenen Daten beschrieben werden, geg.falls zusammen mit Ausdrucken der verwendeten Skripte. Es muss klar sein, welche der abgegebenen Daten in die Studie eingeflossen sind, warum bestimmte Daten ausgeschlossen wurden etc.

## Ansprechpartner

Bei Fragen und zur konkreten Abgabe der Sprachdaten wenden Sie sich bitte an Florian Schiel ([schiel@bas.uni-muenchen.de](mailto:schiel@bas.uni-muenchen.de)) oder Christoph Draxler ([draxler@bas.uni-muenchen.de](mailto:draxler@bas.uni-muenchen.de)).

Abgabetermin ist spätestens der Termin der mündlichen Prüfung.