

SlaVaComp

COMPUtergestützte Untersuchung von VARIabilität im KirchenSLAvischen

Irina Podtergera, Markus Klein, Achim Rabus,
Juliane Besters-Dilger, Gerhard Schneider

Projektlaufzeit: 01.10.2012 – 30.09.2015

Albert-Ludwigs-Universität Freiburg



UNI
FREIBURG



Bundesministerium
für Bildung
und Forschung

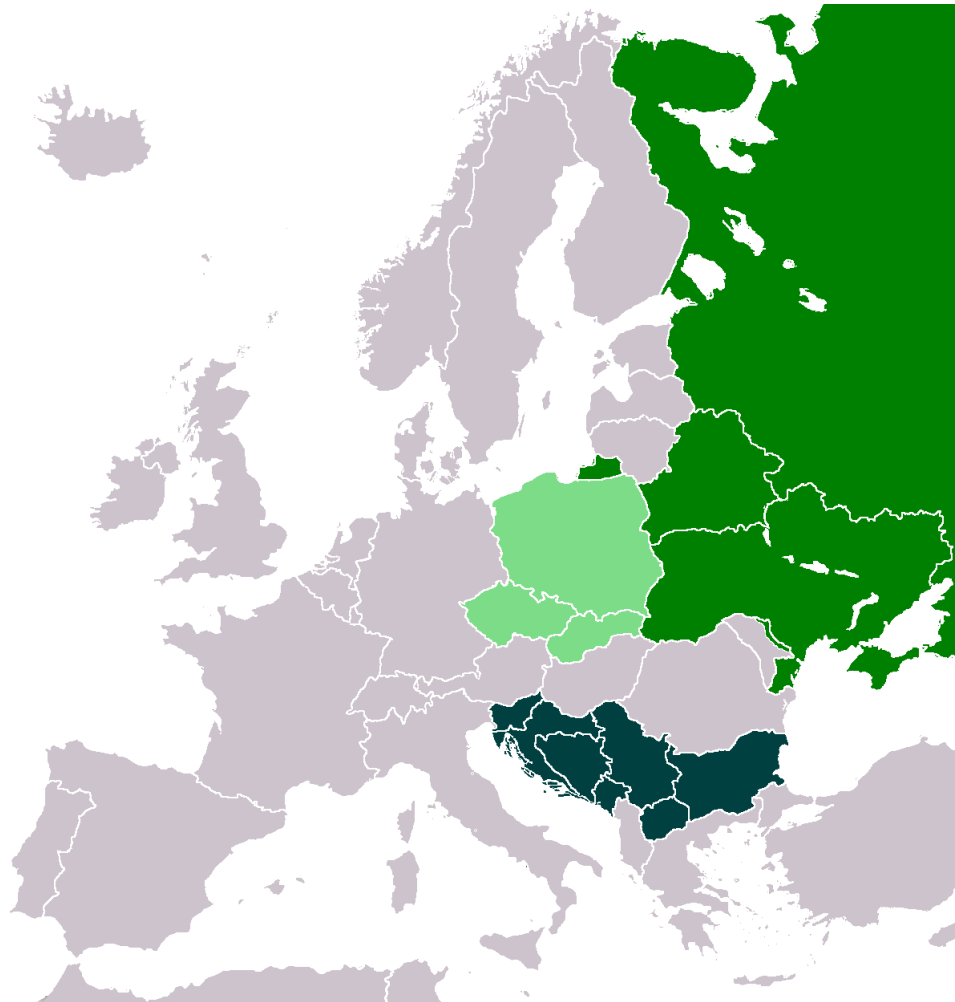
Slavisches Seminar
Universität Freiburg



Rechenzentrum der
Universität Freiburg

- Nachzeichnung der Entwicklung der kirchenslavischen Sprache
 - Lexikalische Variabilität bei Übersetzungen aus dem Griechischen: Zeit und Ort der Entstehung?
 - Graphische Variabilität in Zeit und Raum
 - Entwicklung kirchenslavischer Texte im Laufe ihrer handschriftlichen Überlieferung?
- Wörterbuch für die Zuordnung weiterer bisher nicht einzuordnender Übersetzungen und kirchenslavischer Originaltexte
- Werkzeuge slavistischer historischer Korpuslinguistik

Spachraum des Kichenslavischen





17 digitalisierte Quellen
aus dem 10.–16. Jh.
aus dem ganzen slavisch-
orthodoxen Raum:

- Texte
- ein- und
zweisprachige
Glossare
- Formenindizes

- Recodierung der Texte nach Unicode 5.1
- Zusammenführung griechisch-slavischer Glossare in ein „Metaglossar“
 - Variantenkonkordanz im Metaglossar
 - Scheidung graphischer versus lexikalischer Varianten
 - Festlegung einer graphischen Normalform
- Annotation und Tokenisierung
- Normalisierung der graphischen Varianten
- Lemmatisierung der Wortformen
- Indexierung: Erstellung eines umfassenden Glossars mit zeitlicher und räumlicher Einordnung der Fundstellen



Warum TextGrid?



- Framework: GUI, Editoren, Rechteverwaltung., Grid, ...
 - Handling mehrerer Textschichten nach Normalisierung bzw. Lemmatisierung
 - Nutzung linguistischer Algorithmen
 - Entwicklungen in Konformität mit Standards
 - Chance der Nachnutzung der eigenen Entwicklungen durch die Community / durch andere Philologien
- neue Konfigurationen und Module sind Open Source

- Philologisch:
 - Sprachen: → griechisch und kirchenslavisch
 - Schriftsystem: → griechisch und frühkyrillisch
 - Textsorten: → Bibelkommentare, Kirchenväter, ...
 - Zeitraum: → 10.–16. Jahrhundert
- Technologisch:
 - Texte in nicht-lateinischem Schriftsystem
 - Delimiter-Definition für kirchenslavische Texte
 - Anpassung des Normalisierers
 - Entwicklung eines Lemmatisierers
 - Weiterentwicklung des Indexierers