

Dr. Irina Podtergera, Slavisches Seminar der Universität Freiburg
Susanne Mocken, Rechenzentrum der Universität Freiburg

Unter Rubrik „Beispiele für disziplinspezifische Anwendungen in der ganzen Breite der Geisteswissenschaften, sowohl in ihren objektbezogenen (Archäologie, Ur- und Frühgeschichte, Kunstgeschichte etc.) als auch in ihren textbezogenen Ausprägungen.“

SlaVaComp – Kirchenslavisch digital: wozu?*

Bei dem Vortrag handelt es sich um einen Erfahrungsbericht aus dem Projekt SlaVaComp, das derzeit als ein Kooperationsprojekt zwischen dem Slavischen Seminar und dem Rechenzentrum der Universität Freiburg durchgeführt wird. Ziel des Projekts ist es, aus mehreren unterschiedlich formatierten griechisch-kirchenslavischen bzw. kirchenslavisch-griechischen Glossaren ein zweisprachiges Metaglossar zu erstellen, das die lexikalische Variabilität des Kirchenslavischen in seiner regionalen und chronologischen Entwicklung erfasst.

Bereits in den ersten Monaten der Projektarbeit ergaben sich mehrere Probleme philologischer, linguistischer und informatischer Art, die im Vortrag zur Diskussion gestellt werden sollen. Aus der Sicht der Philologie wird dabei der Schwerpunkt auf die Frage gelegt, inwieweit uns die moderne Computertechnologie hilft, unser Wissen über die erste Schriftsprache der orthodoxen Slaven auf einem qualitativ neuen Niveau darzustellen und strittige Fragen zu beantworten sowie Desiderata der historischen Slavistik zu erfüllen. Aus der Sicht der Informatik soll auf die Lösungen eingegangen werden, mit denen linguistisch und technisch heterogene Ausgangsdaten in den Quelldateien in eine komplexe Datenbank mit einem mehrstufigen System von Auswahl- und Suchoptionen vereinigt werden.

Die lexikographische Erfassung des Kirchenslavischen gehört zu den Problemfeldern der slavischen Philologie. Wir verfügen bis heute nur über ein Wörterbuch, das dieses Idiom in seiner ältesten Entwicklungsstufe darstellt. Es handelt sich dabei um das sog. Prager Wörterbuch des Altkirchenslavischen. Es umfasst den Wortschatz der Kanontexte aus dem 10. – 11. Jh., d. h. der Texte, auf deren Grundlage das Altkirchenslavische als Sprache rekonstruiert wurde. Eine weitere Entwicklung dieses Idioms in unterschiedlichen Regionen der Slavia orthodoxa vom 11. bis 17. Jh. bleibt immer noch ohne lexikographische Erfassung, die modernen Anforderungen der Sprachgeschichtsforschung entsprechen würde. Dies führt zu Fehlinterpretationen in slavischen Nationalphilologien.

In den letzten zwanzig Jahren wurden von makedonischen, bulgarischen, serbischen und russischen Philologen mehrere Wörterverzeichnisse zu den Editionen slavischer mittelalterlicher Texte vorbereitet. Diese in Papierform vorhandenen Glossare ergänzen zwar wesentlich unsere Vorstellungen über die Entwicklung des Kirchenslavischen im Laufe der sieben Jahrhunderte. Jedoch erlauben sie nicht, diese Entwicklung als ein vollständiges Bild darzustellen. Letzteres ist erst dann möglich, wenn alle Glossare zu einem Metaglossar zusammengeführt wurden. Gerade dies erlaubt das digitale Format. Mehr noch: Ein digitales Metaglossar ermöglicht uns, die regionale und funktionale Heterogenität des Kirchenslavischen, darunter auch in den Kanontexten, auf denen das Prager Wörterbuch basiert, of-

* Das Projekt „SlaVaComp – COMPutergestützte Untersuchung von VARIabilität im KirchenSLAvischen“ wird vom BMBF gefördert (FKZ: 01UG1251, Laufzeit: 15.01.2013–15.01.2016). Dr. Irina Podtergera ist Stipendiatin des Margarete von Wrangell-Habilitationsprogramms für Frauen (gefördert durch das Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg).

fensichtlich zu machen. Dies widerlegt die immer noch gängige Meinung über das Kirchenslavische als ein in sich geschlossenes System.

Im Vortrag soll illustriert werden, wie ein digitales Wörterbuch des Kirchenslavischen traditionelle Vorstellungen über diese Sprache ändert und von welcher Bedeutung dies für die weitere Erforschung der slavischen Sprachgeschichte ist. Daneben sollen linguistische Probleme zur Sprache kommen, die in den Altphilologien häufig unbeachtet bleiben, was eine negative Auswirkung auf die Darstellung des sprachlichen Stoffs und seine darauf folgende wissenschaftliche Interpretation hat. Es existiert beispielsweise kein Wörterbuch des Mittelhellenischen. Die Wörterbücher, die in der klassischen Philologie benutzt werden, sind in erster Linie Wörterbücher des klassischen Griechischen. Übersetzungen ins Kirchenslavische wurden jedoch aus dem christlichen Griechischen gemacht. Es gibt zwar Lexika, die das patristische oder das Bibelhellenische fixieren (Lampe, Bauer, Muraoka). Es gibt jedoch keinen Standard, nach dem das Mittelhellenisch in den Wörterbüchern dargestellt werden soll. Auch das „Lexikon zur byzantinischen Gräzität“, bei dem es sich im Grunde genommen um eine Ergänzung zum Liddell/Scott-Wörterbuch des klassischen Griechischen handelt, erfasst nicht das Mittelhellenische als Ganzes. Dieses Problem ist aus der Sicht der Paläoslavistik alles andere als trivial, weil sich mittelalterliche slavische Schreiber bei Übersetzungen aus dem Griechischen des Wortgebrauchs ihrer griechischen Quellen bedienten. Das von uns erstellte Metaglossar macht diese Diskrepanz offensichtlich. Dies soll im Rahmen des Vortrags an konkreten Beispielen erörtert werden. Ein besonderer Akzent wird dabei auf das Problem der Lemmatisierung des Mittelhellenischen für die Bedürfnisse der Altslavistik gelegt werden.

Einen weiteren Schwerpunkt des Vortrags bildet die Frage, wie philologische und historisch-sprachwissenschaftliche Probleme mit Hilfe der Informatik behoben werden können.

Auf technischer Ebene ergeben sich auf dem Weg von den einzelnen Glossaren zum Metaglossar mehrere Schwierigkeiten. Zunächst müssen die Glossare normalisiert werden. Obwohl alle Glossare, die in der Datenbank enthalten sein werden, bereits in digitaler Form (Microsoft-Word-Dateien) vorliegen, bedürfen sie einiger Nachbearbeitung. Fast alle Dateien sind zu einem Zeitpunkt entstanden, bevor die darin dargestellten Buchstaben Teil des Unicode-Standards waren. Da es noch keinen Font gab, der alle diese Buchstaben hätte abbilden können, wurden verschiedene Schriftarten für die Abbildung verwendet. Eine einfache Portierung in das neuere .docx-Format unter Verwendung einer einzigen Unicode-Schriftart wird dadurch unmöglich. Diese Schwierigkeit wurde gelöst durch die Programmierung einer kleinen Anwendung, die alle in einem Dokument enthaltenen Zeichen von Nicht-Unicode-Fonts in Unicode-Zeichen umwandelt und in einem Unicode-Font, d. h. Roman Cyrillic, darstellt. Der nächste Schritt sieht die Transformation der Worddateien in TEI-konformes XML vor. Auch das ist nicht trivial, weil die Glossare sehr heterogene Strukturen aufweisen, was bei der Strukturierung der XML-Dateien berücksichtigt und entsprechend umgesetzt werden muss.

Ein weiteres größeres Problem stellt der Umgang mit der graphischen Variabilität dar. In der XML-Datei wird jedem Lemma ein Hyperlemma, d. h. eine standardisierte Form zugewiesen. Während dieser Vorgang für das Griechische relativ einfach ist, weil es orthographische Standards gibt, nach denen man automatisiert Hyperlemmata erzeugen kann (z.B. bei der Erzeugung der 1. Person von Verben), muss dies fürs Kirchenslavische, für das die Hyperlemmata dem Standard des Prager Wörterbuchs folgen, (noch) weitgehend manuell erledigt werden. Derzeit wird an einer Möglichkeit gearbeitet, diese Normen automatisiert zu generieren, was jedoch sehr zeitaufwendig ist, da aufgrund der Formenvielfalt für fast jede graphische Realisierung eine eigene Transformationsregel formuliert werden muss.

Sobald alle Glossare fertig in XML-Dateien umgewandelt und die Probleme bezüglich graphischer Variabilität behoben sein werden, kann mit der Erschließung der Datenbank begonnen werden. Dazu ist ein Webservice mit einer je nach Anspruch des Benutzers mehr oder weniger ausführlichen Such- und

Filterfunktion in Vorbereitung. Mit Hilfe dieses Webservices kann das Kirchenslavische schließlich erstmals in seiner ganzen regionalen und funktionalen Heterogenität erfasst und beschrieben werden.

Somit wird deutlich, von welcher Relevanz das digitale Wörterbuchformat mit seiner dynamischen und flexiblen Struktur, die im traditionellen Papierformat ausgeschlossen ist, für die Erschließung der Idiome mit stark ausgebildeter graphischer, morphologischer u. a. Variabilität ist. Darüber hinaus ist es offensichtlich, dass die Lösung einer solchen Aufgabe nur im Rahmen einer interdisziplinären Kooperationsarbeit zwischen Philologen und Informatikern möglich ist.