



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

CENTRUM FÜR INFORMATIONS- UND  
SPRACHVERARBEITUNG (CIS)



CIS · Oettingenstraße 67 · 80538 München

Dr. phil. Michaela Geierhos

Telefon +49 (0)89 2180-9719  
Telefax +49 (0)89 2180-9701

An das Studienbüro  
der Fakultät für  
Sprach- und Literaturwissenschaften  
der Ludwig-Maximilians-Universität  
München

micha@cis.uni-muenchen.de

www.cis.uni-muenchen.de

München, den 11.10.2012

## **Abschlussbericht zum studentischen Forschungsprojekt von Alena Mikhaylova im Sommersemester 2012 im Rahmen von Lehre@LMU**

Der vorliegende Abschlussbericht dokumentiert die Verwendung der finanziellen Mittel aus dem Programm Lehre@LMU im studentischen Forschungsprojekt von

**Frau Alena Mikhaylova**

mit dem Titel

### **„Automatische Alignierung verschiedener deutscher Übersetzungen von 'Verbrechen und Strafe' mit dem russischen Originalwerk von Fjodor M. Dostojewskij“**

Ziel ihres Forschungsvorhabens war es herauszufinden, wie bestimmte Konstruktionen im Russischen ins Deutsche übersetzt werden können bzw. in der Praxis tatsächlich übersetzt werden. Von besonderem Interesse waren dabei prädikative Konstruktionen des Russischen mit Vollverben, mit Idiomen und Stützverbgefügen sowie die jeweiligen Konstruktionen der entsprechenden Übersetzungen. Die Arbeit mit verschiedenen deutschen Übersetzungen ermöglichte es ihr, nicht nur zwischensprachliche Paraphrasierungen sondern auch Paraphrasierungen im Deutschen zu betrachten.

Folgende Schritte waren im ursprünglichen Arbeitsplan vorgesehen und wurden im Rahmen des Forschungsprojektes durchgeführt:

1. der Aufbau einer erweiterten Suche. Sie soll dem Benutzer die Möglichkeit geben, die Parameter für die Suchanfrage selbst anzugeben. D.h. wenn er beispielsweise nur den ersten Satz des ersten Teils in Kapitel 1 im ersten Absatz bei der Suche berücksichtigen möchte, soll ihm dieser Satz im Original und in allen seinen Übersetzungen parallel angezeigt werden.
2. die Erweiterung der existierenden lexikalischen Ressourcen (z.B. Personennamen, Phrasen=feste Wortverbindungen für die Übersetzung vom Russischen ins Deutsche)
3. die strukturelle Analyse der gefundenen Übersetzungen zur parallelen Alignierung der deutschen Varianten mit dem Originalsatz.

Im Folgenden werden die Arbeiten bei den oben genannten Punkten näher ausgeführt:

Das Personennamenlexikon (2) wurde um weitere Einträge ergänzt, was eine bessere Suche ermöglicht (1). Dabei wurde auch in Betracht gezogen, dass z.B. nur ein Vaternamen (Patronym) als Suchbegriff eingegeben werden kann (1). Bisher konnte er nur in einer Sprache gefunden werden, da dabei das Lexikon nicht verwendet wurde, sondern nur der Index der Suchplattform Apache Lucene. Die Erweiterung des Lexikons um Phrasen (2) stellte eine große Herausforderung dar, da die Analyse



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

CENTRUM FÜR INFORMATIONS- UND  
SPRACHVERARBEITUNG (CIS)



von Phrasen eine längere Bearbeitungszeit durch die Verwendung vieler anderer computerlinguistischen Werkzeuge verursacht. Da aber die Nutzer nicht lange auf die Antwort auf eine Suchanfrage warten möchten, wurden vorerst hier keine Phrasen behandelt. Zur besseren Visualisierung der Alignierungen werden die ausgegebenen Sätze nun in einem *Frame* je Text angezeigt (3). Indem diese Sätze beim Scrollen in einer Ebene angezeigt werden, kann der Benutzer die entsprechenden Sätze besser analysieren. Auf diese Weise können zwischensprachliche Paraphrasierungen und die Paraphrasierungen im Deutschen betrachtet werden.

Zu jedem Treffer wurde ein Link eingebaut, der es ermöglicht, zu einer Navigationsoberfläche zu gelangen. Dort kann der Benutzer im jeweiligen Text auf den Ebenen "Teil", "Kapitel", "Absatz" und "Satz" navigieren.

Auf diese Weise können alle vier Texte angepasst werden, sodass das Original und seine drei Übersetzungen verglichen werden können. So kann der Benutzer selbst einstellen, über welche



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

CENTRUM FÜR INFORMATIONS- UND  
SPRACHVERARBEITUNG (CIS)



Parameter er in den Suchergebnissen navigieren möchte: Konkret bedeutet dies, dass er nach der Angabe des Parameters (z.B. „Teil“) über zwei Buttons „up“ und „down“ innerhalb der Textteile vor und zurück navigieren kann. Analog gilt dies für „Kapitel“, „Absatz“ und „Satz“.

Zur Vernetzung mit anderen Nachwuchswissenschaftlern und gemeinsamen Erfahrungsaustausch besuchte Alena Mikhaylova zwei Workshops in Deutschland. Beim Workshop vom 13. bis 14. Juli 2012 an der Universität Stuttgart zum Thema „Prefix Verbs: The Impact of Preposition-like Elements on the Syntax and Semantics of Verbs“ (<http://www.ims.uni-stuttgart.de/veranstaltungen/p-workshop/program.html>). Die Vortragenden kamen aus verschiedenen Ländern (wie z.B. Norwegen, Brasilien, Spanien, Deutschland), was die internationale Forschungslandschaft gut abzeichnete. Die Präpositionen wurden in verschiedenen Sprachen behandelt, was den Vorteil hatte, Gemeinsamkeiten und Unterschiede hervorheben zu können. Beim Workshop vom 27. bis 28. September 2012 am Institut für Deutsche Sprache in Mannheim hatte Alena Mikhaylova die Gelegenheit einige Wissenschaftler aus dem Bereich der Computerlinguistik kennenzulernen und sich mit Ihnen auszutauschen. Der Workshop hat „Webkorpora in der Computerlinguistik und Sprachforschung“ (<http://hypermedia.ids-mannheim.de/gsci-ak/workshop12.html>) zum Thema. Dabei wird natürlich die Frage diskutiert, wie ein Korpus für die Sprachanalyse erstellt und welche Aspekte dabei beachtet werden sollen: wie die Daten aus dem Web aufbereitet werden, oder welche computerlinguistischen Werkzeuge dabei benutzt werden.

Alena Mikhalyova lernte viele Methoden kennen, die sie nun in ihrem Projekt einsetzen kann und bekam auch Anregungen, wie sie manche Implementierungsaspekte zukünftig anders realisieren könnte. Da der Workshop erst Ende September stattfand, konnten sie in der Zeit der Förderung durch Lehre@LMU noch nicht verwirklicht werden. Darüber hinaus hat sie noch Einblick in viele Arbeitskreise zu Fachthemen der Computerlinguistik kennengelernt (<http://www.gsci.org>, <http://www.computerlinguistik.org>), in denen die Vortragenden des Workshops tätig sind, und in denen sie sich über Spezialgebiete der Computerlinguistik austauschen.