

COMPUTERLINGUISTIK ANDERSWO

EINHARD KÖHLER, UNIVERSITÄT TRIER
FB II, LINGUISTISCHE DATENVERARBEITUNG

Im September 1991 unternahm der Autor dieses Beitrags eine Vortrags- und Informationsreise nach Leningrad (jetzt St. Petersburg), Smolensk und Moskau 1, in deren Verlauf er auch Gelegenheit hatte, eine Reihe von Forschungseinrichtungen zu besuchen. Obwohl seit vielen Jahren enge Kontakte und ein ständiger wissenschaftlicher Austausch zwischen deutschen und osteuropäischen Wissenschaftlern - wenigstens im Bereich der quantitativen Linguistik und besonders zwischen den Herausgebern der Reihe QUANTITATIVE LINGUISTICS und ihrem Gegenstück KVANTITATIVNAJA LINGVISTIKA I AVTOMATIÖESKIJ ANALIZ TEKSTA (Tartu)-bestanden hatte, ergab der Besuch Informationen über eine überraschende Vielfalt von Aktivitäten und eine Vielzahl von Instituten, Gruppen und Zentren, deren Arbeit bei uns nicht oder nur wenig bekannt sind.

Daraus resultierte die Idee, im LDVForum eine Rubrik einzurichten, in der aufgrund von Informationsreisen über die Situation von LDV und LDV-bezogenen Forschungen berichtet wird.

Die Forschungsgruppe Statistika reči in St. Petersburg

Die Gruppe statistika reči ("Sprachstatistik") ist eine der bedeutendsten linguistischen Forschungseinrichtungen auf dem Gebiet der ehemaligen Sowjetunion. Sie wurde im Jahre 1956 von R. G. Piotrowski und L. Novak gegründet. Heute gehören mehr als 200 Wissenschaftler zu ihr, von denen viele auch in Hochschulen und Forschungseinrichtungen anderer Länder (USA, Israel, Großbritannien, Frankreich u. a.) tätig sind. Die Gruppe besteht aus einer Reihe regionaler Untergruppen, die unter der Leitung von Prof. Piotrowski von St. Petersburg aus koordiniert werden, darunter die Gruppe St. Petersburg (Leitung Prof. Piotrowski, Prof. P. Alekseev, Dr. L. Belaeva), die Weißrussische Gruppe (Leitung Prof. A. Zubov, Dr.

I. Sovpel), die Moldauische Gruppe (Leitung Prof. W. Czyzakowski, Doz. V. Goncareno), die Kasachische Gruppe (Leitung Prof. K. Bektaev) und einige kleinere Gruppen (Wolga, Irkutsk, Usbekistan, Georgien).

Im Rahmen dieser Einrichtung finden Forschungen vor allem auf dem Gebiet der quantitativen Linguistik unter verschiedenen Anwendungsgesichtspunkten statt; unter den aktuellen Arbeitsschwerpunkten sind besonders die Projekte zur maschinellen Übersetzung hervorzuheben. Es handelt sich um pragmatisch orientierte Ansätze mit ausgeprägtem Lexikonanteil und einer statistischprobabilistischen Komponente. Die z. Zt. behandelten Sprachpaare sind Russisch ↔ Englisch, Russisch ↔ Französisch und Russisch ↔ Deutsch, wobei die Entwicklung des letzten noch in den Anfängen steckt. In diesem Zusammenhang beklagt R. Piotrowski den Mangel an ausgebildeten Fachkräften mit guten Deutsch-Kenntnissen; auch aus diesem Grund besteht auf russischer Seite ein Interesse an Kooperation. Entsprechende Kontakte und Vorbereitungen zur Aufnahme einer offiziellen Zusammenarbeit mit der Universität St. Petersburg hat das Fach Linguistische Datenverarbeitung der Universität Trier bereits eingeleitet.

Zur Zeit konzentrieren sich die theoretischen und praktischen Entwicklungsarbeiten auf den "polyfunktionalen linguistischen Automaten", der das Resultat mehrjähriger interdisziplinärer KI-orientierter Forschung unter Beteiligung von Mathematikern, Linguisten, Psychologen und anderen darstellt.²

Ein weiterer Schwerpunkt der Gruppe besteht weiterhin in den Untersuchungen zur Sprachstatistik und quantitativer Textanalyse sowie das Studium von pathologischer Sprache mit Anwendungen in der Computerlinguistik und anderen Fächern wie Medizin, Recht, Technik, Pädagogik und Wirtschaftswissenschaften.

² Ein Bericht über diesen Aspekt der St. Petersburger Arbeiten und eine Darstellung des Konzepts des linguistischen Automaten von W. Czyzakowski und R. G. Piotrowski ist in Vorbereitung und wird in einer der nächsten Ausgaben von GLOTTOMETRIKA (in der Reihe QUANTITATIVE LINGUISTICS) erscheinen.

¹ Auch an dieser Stelle sei der Deutschen Forschungsgemeinschaft für die gewährte Unterstützung gedankt.