

Catalysis 92 in Warsaw

In June 1992, the University of Warsaw managed to arrange a successful workshop on a topic, which has recently been much talked about, but which had not materialized so far. Computers have become an commonplace in everyday text production, but their usefulness in research is also obvious. The Institute of Oriental Studies of the University of Warsaw took a brave step in inviting scholars and students to share experiences in the field of computational linguistics, with particular reference to Non-European languages. The writing systems themselves of some of those languages, such as Arabic, Chinese, Japanese, Amharic, Tigrinya etc., pose specific problems, and Catalysis had groups to deal with such problems.

In fact the activities of Catalysis had a double aim of sharing experiences and programs on one hand, and of training the use of some useful programs on the other. The original aim of making the Polish students, teachers and researchers acquainted with recent developments in computational linguistics was expanded into an international workshop, including participants from several countries, including developing countries of Africa and Asia. The participation of the representatives of developing countries was made possible by the grant of the Dutch government, and so the workshop had a really international character.

The training included introductory courses into text processing programmes of Non-Latin characters, such as Chinese, Japanese, Arabic, Amharic, and some Indian languages. Another field of interest was information management in general, and a number of tools for such tasks were introduced. Among the basic tools were programmes designed for lexicological work, tools for linguistic fieldwork, and more general systems of information management. Among such programs are the packages of The Summer School of Linguistics, such as Shoebox, Ample and Cecil, TeX, different versions of Beta, tools in the Unix environment, and a number of commercially available concordance programs. There was discussion on the need of data banks which could be shared by researchers of various countries. Such facilities are offered, for example, by the Unix environment, which facilitates the use of data through net, while at the same time the danger of misuse of the data (e.g. copying) is eliminated by sophisticated security systems.

A number of lectures were given on a wide variety of topics, such as computer-aided lexicography, comparative linguistic work, concordances, machine translation, automatic morphological analyzers, Icon programming, etc.

Linguistic data and other kinds of materials are available in computer form in increasing amounts. There is an imminent need of tools for multiple means of information management and retrieval of computerized data. It was stated that to a certain level such information management may be achieved by using general tools and uncoded source data. For more sophisticated information management, however, language-specific tools and/or coding, or tagging, of the source data is

necessary. Such tools, as for example morphological and syntactic parsers, are being developed for a number of languages. As it seems, the development goes towards comprehensive information management systems, where a set of tools is available for carrying out many kinds of automated processes. The desired result will be achieved with the aid of a set of independent programs, rather than with one stiff and clumsy program.

It seems that in near future cooperation between universities and research centres will increase substantially, and joint efforts of several universities and individual researchers will intensify, so as to create sizeable text corpora and other kinds of data banks. Computer linguistics is a research area where considerable advances may be expected in near future, and where the study of Non-European languages is likely to give a specific contribution to the study of linguistics. At the same time, it provides a set of tools which greatly alleviate tedious phases in our everyday work. One may only hope continuation to such meetings and workshops, which effectively contribute to the distribution of knowledge in this rapidly developing field.

Arvi Hurskainen