Preface

In the past decade, the greatly increased availability of millions of words of text in machine-readable form has been associated with a resurgence of empirical methods in natural-language research. In particular, statistical methods of linguistic analysis have enjoyed a new-found popularity, with resultant benefits in machine translation, information retrieval, lexicography, message understanding, computer-assisted language learning, man-machine interfaces, and other fields. Research projects in computational text analysis continue to burgeon in innumerable research institutions throughout the world, leading to the development of practical applications in all the world's major languages.

Analysis of speech and dialogue, however, is (or has been up to now) different in kind. For the analysis of lexis and syntax, it is necessary to investigate how words interact. Given the typical size of the vocabulary of a natural language, very large quantities of text are required to distinguish significant co-occurences of words from chance, and it is not surprising that research has focused on the analysis of written text, which is now widely available in machine-readable form. Analysis of speech, on the other hand, has in the past tended to focus on phonetic and prosodic patterning, for which the chief empirical requirement is a very detailed level of transcription rather than a very large quantity of text. The lexical and syntactic behaviour of spoken language has been somewhat neglected.

It is now widely acknowledged that lexis and syntax of unscripted conversation may well be very different in kind from that of written language. But the details are still not well understood, in any language. This is due in no small part to the absence of large corpora of unscripted conversation. Such corpora are now beginning to be collected, and other advances in the analysis of speech continue to be made. The spotlight of research is swivelling gradually round to text, speech, and dialogue.

The TSD'98 Workshop is devoted to these areas of NLP in particular. One of its strong points lies in the effort to intertwine these, at present still separate branches of NLP research.

Considerable benefits are now being gained from the corpora of spoken text and dialogue that are now being collected, and the application of the methodology of computational and corpus linguistics. In fact, corpora have brought about relevant methodological changes not only in text exploration but also in linguistics and linguistic engineering in general. Techniques developed in corpus linguistics are becoming widely used in speech research as well. The relevant results in speech synthesis and recognition could not have been obtained without corpora of both spoken and written texts.

Speech research has recently become the centre of attention for many researchers and research institutions, and their number is constantly growing. There are good reasons to expect that many important practical applications will be developed or significantly improved in the near future, with the application of empirical methods of research. The contributions in TSD'98 certainly reflect this tendency.

In the third part of the TSD'98 Proceedings one can observe a synthesis of two formerly separate research areas — the various approaches to the dialogue issues

demonstrate a tendency to unification and indicate that significant progress in the development of dialogue systems is being made. As John Sinclair has recently observed, man-computer communication currently tends to be uni-directional: the computer is not expected to answer back! However, if information technology is to develop further and serve human needs better, man-machine communication has to undergo an essential transformation: it must become truly interactive two-way communication. It is not difficult to see that current dialogue research and dialogue systems are tending in this direction.

We hope that the work of the TSD'98 Workshop, and the publication of its proceedings, will prove helpful to many different kinds of researchers working in NLP research. One of the major overall aims of TSD'98 can be summarized as improving man-machine communication, pushing it to higher levels and making it more "friendly" for potential users.

Special thanks must go to the editors, who have laboured long and hard to prepare these Proceedings in advance of the Workshop — Ivan Kopeček, Karel Pala, Václav Matoušek and especially Petr Sojka whose TEXpertise was indispensable to give the Proceedings their present shape. We would also like to express our thanks to Eva Žáčková, Pavel Rychlý, Pavel Smrž, Aleš Horák and Marek Veber, whose dedicated help with proofreading contributed substantially to the final form of the Proceedings.

Patrick Hanks

Chair, Programme Committee, TSD'98

Organization

TSD'98 is organized by the Faculty of Informatics, Masaryk University in cooperation with University of West Bohemia in Plzeň. Workshop WWW page is `http://www.fi.muni.cz/tsd98/`.

Programme Committee

Baudoin Geneviève (France), Ferencz Attila (Romania), Hanks Patrick (Great Britain, chair), Hermansky Hynek (USA), Jelinek Frederick (USA), Krishnamurthy Ramesh (Great Britain), Kopeček Ivan (Czech Republic), Matoušek Václav (Czech Republic), Müller Johannes (Germany), Noeth Elmar (Germany), Pala Karel (Czech Republic), Pavesić Nikola (Slovenia), Rubio Antonio (Spain), Schukat-Talamazzini E. Günter (Germany), Skrelin Pavel (Russia).

Organizing Committee

Bártek Luděk, Batůšek Robert, Komárková Dana *(secretary)*, Kopeček Ivan *(chair)*, Matoušek Václav, Pala Karel, Smrž Pavel, Sojka Petr, Staudek Jan, Žáčková Eva *(principal contact)*, Žižka Jan.

Sponsoring Institutions

Computer Press