Preface

There have been a number of workshops in recent years on collocations, terminology and named entity recognition. However, multiword expressions (MWEs) that encompass all of these subtypes remain a real challenge for natural language processing (NLP) despite several decades of research effort. The aim of this SIGLEX workshop is to bring together NLP researchers working on all areas of MWEs. The objectives are to summarise what has been achieved in the area, to establish common themes between different approaches and to discuss future trends, with particular emphasis on addressing the problems that MWEs pose for real-world NLP applications. We welcomed submissions on all aspects of analysis, acquisition and treatment of these 'words with spaces' which often require special semantic interpretation and may have peculiar syntactic behaviour.

We received 30 submissions (8 from Asia, 14 from Europe and 8 from the Americas), and accepted 13 of them for presentation. Each submission was reviewed by three members of the program committee, who not only judged each submission but also gave detailed comments to the authors. The overall quality of submissions was high, making the final selection very difficult. The papers in these proceedings are those which were finally selected for presentation. Many of the papers deal with MWEs in general, rather than aiming at specific subtypes, with several papers on detection and extraction using a variety of methods: one using semantics and the rest relying on statistics and syntax. There are papers dealing with particular subtypes of MWEs, of which three papers specifically target phrasal verbs. Some papers address compositionality and semantic interpretation and a couple look at MWEs in the context of real applications, namely Question Answering and Machine Translation. Apart from English, other languages are also investigated, with proposals for both French and Japanese.

We would like to thank all the authors who submitted papers. We also thank all the members of the program committee for their time and effort in ensuring that the papers were fairly assessed.

The workshop was supported by

- Research Collaboration between NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation and CSLI, Stanford University
- UK EPSRC project GR/N36493 "Robust Accurate Statistical Parsing (RASP)"
- EU FW5 project IST-2001-34460 "MEANING"

Finally, we wish to thank the organizers of the main conference, in particular the conference workshop co-chairs, Lori Levin, Takenobu Tokunaga and Alessandro Lenci.

Francis Bond, Anna Korhonen, Diana McCarthy and Aline Villavicencio

May 2003