

Foreword by the Managing Editors¹

Language Technology and the Information Society

The information age is characterized by a fast growing amount of information being made available either in the public domain or commercially. This information is acquiring an increasingly important function for various aspects of peoples' professional, social and private life, posing a number of challenges for the development of the Information Society.

In particular, the classical notion of universal access needs to be extended beyond the guarantee for physical access to the information channels, and adapted to cover the rights for all citizens to benefit from the opportunity to easily access and effectively process information.

Furthermore, with the globalization of the economy, business competitiveness rests on the ability to effectively communicate and manage information in an international context.

Obviously, languages, communication and information are closely related. Indeed, language is the prime vehicle in which information is encoded, by which it is accessed and through which it is disseminated.

Language technology offers people the opportunity to better communicate, provides them with the possibility of accessing information in a more natural way, supports more effective ways of exchanging information and control its growing mass.

There is also an increasing need to provide easy access to multilingual information systems and to offer the possibility to handle the information they carry in a meaningful way. Languages for which no adequate computer processing is being developed, risk gradually losing their place in the global Information Society, or even disappearing, together with the cultures they embody, to the detriment of one of humanity's great assets: its cultural diversity.

What Can Language Technology Offer?

Looking back, we see that some simple functions provided by language technology have been available for some time—for instance spelling and grammar checking. Good progress has been achieved and a growing number of applications are maturing every day, bringing real benefits to citizens and business. Language technology is coming of age and its deployment allows us to cope with increasingly difficult tasks.

Every day new applications with more advanced functionality are being deployed—for instance voice access to information systems. As is the case for other information technologies, the evolution towards more complex language processing systems is rapidly accelerating, and the transfer of this technology to the market is taking place at an increasing pace.

¹The ideas expressed herein are the authors' and do not reflect the policies of the European Commission and the Italian National Research Council.

More sophisticated applications will emerge over the next years and decades and find their way into our daily lives. The range of possibilities is almost unlimited. Which ones will be more successful will be determined by a number of factors, such as technological advances, market forces, and political will.

On the other hand, since sheer mass of information and high bandwidth networks are not sufficient to make information and communication systems meaningful and useful, the main issue is that of an effective use of new applications by people, which interact with information systems and communicate with each other.

Among the many issues to be addressed are difficult engineering problems and the challenge of accounting for the functioning of human languages—probably one of the most ambitious and difficult tasks.

Benefits that can be expected from deploying language technology are a more effective usability of systems (enabling the user) and enhanced capabilities for people (empowering the user). The economic and social impact will be in terms of efficiency and competitiveness for business, better educated citizens, and a more cohesive and sustainable society. A necessary precondition for all this, is that the enabling technology be available in a form ready to be integrated into applications.

The subject of the thirteen chapters of this Survey are the key language technologies required for the present applications and research issues that need to be addressed for future applications.

Aim and Structure of the Book

Given the achievements so far, the complexity of the problem, and the need to use and to integrate methods, knowledge and techniques provided by different disciplines, we felt that the time was ripe for a reasonably detailed map of the major results and open research issues in language technology. The Survey offers, as far as we know, the first comprehensive overview of the state of the art in spoken and written language technology in a single volume.

Our goal has been to present a clear overview of the key issues and their potential impact, to describe the current level of accomplishments in scientific and technical areas of language technology, and to assess the key research challenges and salient research opportunities within a five- to ten-year time frame, identifying the infrastructure needed to support this research. We have not tried to be encyclopedic; rather, we have striven to offer an assessment of the state of the art for the most important areas in language processing.

The organization of the Survey was inspired by three main principles:

- an accurate identification of the key work areas and sub-areas of each of the fields;
- a well-structured multi-layered organization of the work, to simplify the coordination between the many contributors and to provide a framework in which to carry out this international cooperation;

- a granularity and style that, given the variety of potential readers of the Survey, would make it accessible to non-specialist and at the same time to serve for specialists, as a reference for areas not directly of their own expertise.

Each of the thirteen chapters of the Survey consists of:

- an introductory overview providing the general framework for the area concerned, with the aim of facilitating the understanding and assessment of the technical contributions;
- a number of sections, each dealing with the state of the art, for a given sub-area, i.e., the major achievements, the methods and the techniques available, the unsolved problems, and the research challenges for the future.

For ease of reference, the reader may find it useful to refer to the analytical index given at the end of the book.

We hope the Survey will be a useful reference to both non-specialists and practitioners alike, and that the comments received from our readers will encourage us to edit updated and improved versions of this work.

Relevance of International Collaboration

This Survey is the result of international collaboration, which is especially important for the progress of language technology and the success of its applications, in particular those aiming at providing multilingual information or communication services. Multilingual applications require close coordination between the partners of different languages to ensure the interoperability of components and the availability of the necessary linguistic data—spoken and written corpora, lexica, terminologies, and grammars.

The major national and international funding agencies play a key role in organizing the international cooperation. They are currently sponsoring major research activities in language processing through programs that define the objectives and support the largest projects in the field. They have undertaken the definition of a concrete policy for international cooperation² that takes into account the specific needs and the strategic value of language technology.

Various initiatives have, in the past ten years, contributed to forming the cooperative framework in which this Survey has been organized. One such initiative was the workshop on 'Automating the Lexicon' held in Grosseto, Italy, in 1986, which involved North American and European specialists, and resulted in recommendations for an overall coordination in building reusable large scale resources.

Another one took place in Turin, Italy, in 1991, in the framework of international cooperation agreement between the NSF and the ESPRIT programme

²Several international cooperation agreements in science and technology are currently in force: more are being negotiated.

of the European Commission. The experts convened at that meeting called for cooperation in building reusable language resources, integration between spoken and written language technology—in particular the development of methods for combining rule-based and stochastic techniques—and an assessment of the state of the art.

A special event convening representatives of American, European and Japanese sponsoring agencies was organized at COLING 92 and has since become a permanent feature of this bi-annual conference. For this event, an overview³ of some of the major American, European and Japanese projects in the field was compiled.

The present Survey is the most recent in a series of cooperative initiatives in language technology.

Acknowledgements

We wish to express our gratitude to all those who, in their different capacities, have made this Survey possible, but first of all the authors who, on a voluntary basis, have accepted our invitation, and have agreed to share their expert knowledge to provide an overview for their area of expertise.

Our warmest gratitude goes to Oscar Garcia, who co-inspired the initiative and was an invaluable colleague and friend during this project. Without his scientific competence, management capability, and dedicated efforts, this Survey would not have been realized. His successor, Gary Strong, competently and enthusiastically continued his task.

Thanks also to the commitment and dedication of the editorial board consisting of Joseph Mariani, Hans Uszkoreit, Annie Zaenen and Victor Zue. Our deep-felt thanks to Ron Cole, who coordinated the board's activities and came to serve as the volume's editor-in-chief.

Mark Liberman, of the University of Pennsylvania and initially member of the editorial board, was instrumental in having the idea of this Survey approved, and his contribution to the design of the overall content and structure was essential. Unfortunately, other important tasks called him in the course of this project.

Invaluable support to this initiative has been provided by Y.T. Chien, the director of the Computer and Information Science and Engineering Directorate of the National Science Foundation, Vincente Parajon-Collada, the deputy-director general of Directorate General XIII of the European Commission, and Roberto Cencioni head of Language Engineering sector of the Telematics Application Programme.

Vince Weatherill, of Oregon Graduate Institute, dedicated an extraordinary amount of time, care and energy to the preparation and editing of the Survey.

³Synopses of American, European and Japanese Projects Presented at the International Projects Day at COLING 1992. In: *Linguistica Computazionale*, volume VIII, Giovanni Battista Varile and Antonio Zampolli, editors, Giardini, Pisa. ISSN 0392-6907 (out of print). This volume was the direct antecedent of and the inspiration for the present survey.

xx

Colin Brace carried out the final copyediting work within an extremely short time schedule.

The University of Pisa, Italy, the Oregon Graduate Institute, and the Institute of Computational Linguistics of the Italian National Research Council generously contributed financial and human resources.

Antonio Zampolli
University of Pisa

Giovanni Battista Varile
Directorate General XIII
of the European Commission
Luxembourg