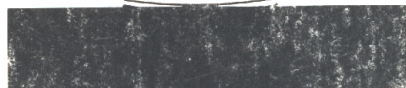
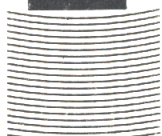




WALKER · KARLGREN · KAY



Natural Language

in

INFORMATION



SCIENCE





**NATURAL LANGUAGE
IN
INFORMATION SCIENCE**

Perspectives and Directions for Research

Edited by: Donald E. Walker
SRI International

Hans Karlgren
KVAL Institute for Information Science

Martin Kay
Xerox Palo Alto Research Center

Results of a
WORKSHOP ON LINGUISTICS AND INFORMATION SCIENCE
Biskops-Arnö, Sweden, 3-5 May 1976

Organized by the Committee on Linguistics in Documentation
of the International Federation for Documentation and KVAL

FID Publication 551

Skriptor, Stockholm, Sweden

© Skriptor AB, Stockholm, 1977.

ISBN 91 7282 131-0

Printed in U.S.A. by B and P Printing, Fairfax, Virginia, 1977.

TABLE OF CONTENTS

Introduction Donald E. Walker, <i>SRI International</i>	3
The Organization and Conduct of the Workshop Donald E. Walker, <i>SRI International</i>	7
Perspective Paper: Information Science F.W. Lancaster, <i>University of Illinois</i>	19
Perspective Paper: Library Science Derek Austin, <i>The British Library</i>	45
Perspective Paper: Quantitative Linguistics Wolf Moskovich, <i>Hebrew University of Jerusalem</i>	57
Perspective Paper: Computational Linguistics Naomi Sager, <i>New York University</i>	75
Perspective Paper: Linguistics Petr Sgall, <i>Charles University</i>	101
Perspective Paper: Complex Semantic Information Processing Teun A. van Dijk, <i>University of Amsterdam</i>	127
Perspective Paper: Terminology J. Goetschalckx, <i>Commission of the European Communities</i>	165
Challenge Paper: Homeosemy -- On the Linguistics of Information Retrieval Hans Karlgren, <i>KVAL Institute for Information Science</i>	167
Linguistics and Information Science: After Five Years Karen Sparck Jones, <i>Cambridge University</i> Martin Kay, <i>Xerox Palo Alto Research Center</i>	183
The Workshop Participants	193

INTRODUCTION

In 1970, the Committee on Linguistics in Documentation of the International Federation for Documentation commissioned the preparation of a survey of the linguistic components of document analysis, description, and retrieval. Karen Sparck Jones and Martin Kay, the Committee members from the United Kingdom and the United States, respectively, accepted the responsibility for carrying out this study. The survey included examinations of the relation that linguistic theory bears to the structure of knowledge; the relevance of phonology, morphology, syntax, and semantics to the organization, storage, and transmission of information; and how computational linguistics has been and how it might be applied to information handling. It also considered topics like automatic morphological and syntactic analysis, machine translation and machine-aided translation, automatic indexing and abstracting, and the machine preparation of indexes and thesauruses. The results were published under the title *Linguistics and Information Science**. On the basis of their analysis, Sparck Jones and Kay concluded that the impact of linguistics on information science had been small. They recommended strongly that efforts be directed toward determining what could be done to provide more effective collaboration and communication between these disciplines. Toward this goal and responsive to a suggestion by Hans Karlgren, the Committee member from Sweden, who had prompted the initial study, the Committee decided at its meeting in 1975 to organize a Workshop that would bring together specialists in the relevant areas to work toward a comprehensive research plan. The Workshop was scheduled for 3-5 May 1976; Sweden was selected as the site. Karlgren volunteered to handle local arrangements and secured as a conference facility the Biskops-Arnö estate, which is located near Stockholm and run by an association for Intra-Scandinavian Cultural Cooperation.

The people invited to participate represented a broad range of interests and backgrounds. In addition to linguists and information scientists, the list included computational linguists, quantitative linguists, specialists in library science, terminologists, translators, and research workers in artificial intelligence. There were thirty participants at the workshop, from thirteen countries: Belgium, Denmark, France, the German Federal Republic, Hungary, Israel, Italy, Luxembourg, the Netherlands, Norway, Sweden, the United Kingdom, and the United States. An additional ten, who had accepted the invitation but were prevented for various reasons from attending at the last minute, would have added Bulgaria, Canada,

*Academic Press, New York, 1973.

Introduction

Czechoslovakia, India, and the USSR to that list. Several of this latter group contributed significantly to the Workshop, in spite of their absence, through the papers they prepared for it. The names and institutional affiliations of the participants are provided at the end of the book.

In preparation for the Workshop, the participants were asked to read a number of documents. For background, they were supplied with *Linguistics and Information Science* and two reviews of automated language processing that had been prepared for the *Annual Review of Information Science and Technology* (Walker, 1973, from Volume 8, and Damerau, 1976, from Volume 11.* In addition, they received copies of perspective papers written by Derek Austin, J. Goetschalckz, Wilfred Lancaster, Wolf Moskovich, Naomi Sager, Petr Sgall, and Teun van Dijk to address the issues raised by the Sparck Jones and Kay book from the vantage points, respectively, of library science, terminology and translation, information science, quantitative linguistics, computational linguistics, linguistics, and the approach to complex semantic information processing found in recent research on text grammars. Finally, they were confronted with a *challenge paper* prepared by Hans Karlgren to delineate some major issues for the Workshop. The challenge paper and the perspective papers constitute the major portion of this volume.**

The sessions at the Workshop consisted of focused discussions organized around the issues raised by the source documents. Broadly characterized, the plan for the participants was on the first of the three days to establish the state of our current knowledge; on the second day to consider the kinds of capabilities that would be desirable in the future; and on the third day to propose research that would provide those capabilities. The discussions reflected the heterogeneity of disciplines represented in the participants, the variety of professional roles they occupied, and the differences in the time frames within which they believed practical benefits should be achieved. Respecting these dimensions of variability, there were *clusters of consensus* rather than one uniform view shared by all. Consequently, a comprehensive plan for research did not emerge from the Workshop, although there were steps toward such a plan that both require significant work for further development and show considerable promise for future activities in linguistics and information science.

At the end of the Workshop, Sparck Jones and Kay, the protagonists whose initial study motivated and guided its deliberations, decided that they should update their book. The resulting evaluation of "Linguistics and Information Science: After Five Years" provides an effective concluding chapter.

The Workshop, in its perspectives, in its challenges, and in its views on relevant research, did capture the breadth and richness of the relationship between linguistics and information science. Correspondingly, we believe this volume, which embodies the substance of that Workshop, will present to its readers, who certainly vary in interests and backgrounds as much as the participants, both ideas and the stimulus to sharpen, perfect, and implement them across the full range of relevant research. There is clearly much to do, but it is equally clear that the benefits to information science can be substantial.

*The Committee expresses its appreciation to Dr. Fred Damerau, the author, to Dr. Martha Williams, the editor for Volume 11, and to the American Society for Information Science, its publisher, for permission to distribute this review to the Workshop participants in advance of publication.

**Some of the authors of the perspective papers included as attachments papers published elsewhere that provided additional material. These attachments are not included here, but references are provided to them where appropriate in the text.

Introduction

As Chairman of the Committee on Linguistics in Documentation of the International Federation for Documentation, I wish to express my appreciation to the participants in the Workshop and to my colleagues on the Committee, particularly to Hans Karlgren and his KVAL Institute for Information Science for physical, financial, and spiritual support for the Workshop and to Martin Kay for his help in preparation of the manuscript. We are indebted to the International Federation for Documentation, the U. S. National Academy of Sciences, and the U. S. National Science Foundation for direct financial assistance. We also thank the other organizations and agencies that supported the attendance of various participants at the Workshop.

Donald E. Walker

SRI International
April 1977

References

Damerau, F.J. "Automated Language Processing." In Williams, M.E., ed., *Annual Review of Information Science and Technology*, Volume 11. Washington, D.C., American Society for Information Science, 1976. Pp. 107-161.

Walker, D.E. "Automated Language Processing." In Cuadra, C.A., and Luke, A.W., eds., *Annual Review of Information Science and Technology*, Volume 8. Washington, D.C., American Society for Information Science, 1973. Pp. 69-119.