FUNDAMENTAL STUDIES IN COMPUTER SCIENCE 9

computational models of natural language processing

b.g. bara
and
g. guida
editors

north-holland

BIBLIOTHEQUE DU CERIST

BIBLIOTHEQUE DU CERIST

To our wives
Marcella and Antonia

COMPUTATIONAL MODELS OF NATURAL LANGUAGE PROCESSING

BIBLIOTHEQUE DU CERIST

Fundamental Studies in Computer Science

Advisory Board:
J. Feldman, R. Karp, L. Nolin, M. O. Rabin, J. C. Shepherdson, A. van der Sluis, and P. Wegner

VOLUME 9

BIBLIOTHEQUE DU CERIST

NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD

C
1865

Computational Models of Natural Language Processing

BIBLIOTHEQUE DU CERIST

Edited by

BRUNO G. BARA
Università di Milano, Italy

and

GIOVANNI GUIDA
Politecnico di Milano, Italy

1984

NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD

© ELSEVIER SCIENCE PUBLISHERS B.V., 1984

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN: 0 444 875980

Publishers:
ELSEVIER SCIENCE PUBLISHERS B.V. (NORTH-HOLLAND)
P.O. Box 1991
1000 BZ Amsterdam
The Netherlands

Sole distributors for the U.S.A. and Canada:
ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
52 Vanderbilt Avenue
New York, N.Y. 10017
U.S.A.

BIBLIOTHEQUE DU CERIST

4586

Library of Congress Cataloging in Publication Data

Main entry under title:

Computational models of natural language processing.

(Fundamental studies in computer science ; v. 9)
Includes index.
1. Linguistics--Data processing--Addresses, essays, lectures. I. Bara, Bruno G., 1949- . II. Guida, Giovanni, 1951- . III. Series.
P98.C6123 1984 410'.28'54 84-13689
ISBN 0-444-87598-0

PRINTED IN THE NETHERLANDS

BIBLIOTHEQUE DU CERIST

PREFACE

This book is the result of a work started about two years ago, with the aim of assembling a collection of papers which could contribute to a unitary understanding of the multidisciplinary area of Natural Language Processing. A unifying point of view of the different paradigms of artificial intelligence, linguistics, and cognitive science may be identified, in our opinion, in the fundamental issue of developing computational models of language competence. We claim that progress in the design of high-performance natural language systems greatly depends on the advances in the study of theoretical models of language comprehension and production. The main concern of the volume is therefore the illustration of models for natural language processing, and the discussion of their role in the development of computational studies of language.

The volume contains 14 papers, including an introductory paper by the editors. All papers are original, unpublished contributions, and present significant research results, often supported by experimental activity. The volume is intended for scholars acquainted with natural language research, but not necessarily specialists in all the disciplines included in the area of natural language processing.
A brief presentation of the content of each paper follows.

In the introductory paper (Competence and performance in the design of natural language systems) Bruno G. Bara and Giovanni Guida divide the models used in the study of natural language processing into three classes: cognitive, pragmatic, and linguistic. They sketch a brief history of computational linguistics, using paradigmatic examples to discuss the roles of competence and performance in the design and evaluation of natural language systems.

Gabriella Airenti, Bruno G. Bara, and Marco Colombetti analyze in their paper (Planning and understanding speech acts by interpersonal games) communication as a mean for an agent to gain co-operation for the execution of a plan which requires many agents to be successful. Therefore, starting from the intentions of a first agent, a pragmatic model is developed to explain how it is possible to obtain that a second agent intends to perform the role that the first agent assigned to him in the original action plan. Games are introduced as mutually shared knowledge structures, used for planning and understanding speech acts. Examples of different kinds of games are discussed in detail, accounting for failure and failure recovery in communication.

The paper of Amedeo Cappelli, Giacomo Ferrari, Lorenzo Moretti, and Irina Prodanof (A framework for integrating syntax and semantics) deals with the problem of integration of syntactic and semantic processing in the parsing of natural language. Assuming that both syntax and semantics have relevant specific roles in language comprehension, the authors present a hypothesis of integration, which consists in defining a general processor which can manipulate representations of different levels, through the activation of complex functions. This model is supported by several linguistic motivations, and it appears intuitively close to the human process of understanding.

Cristiano Castelfranchi, Domenico Parisi, and Oliviero Stock describe in their contribution (Knowledge representation and natural language: Extending the expressive power of proposition nodes) a knowledge representation schema focused on proposition nodes in semantic nets. A number of developments and improvements to the classical approach of Schubert (1976) is proposed, with the aim of fully exploiting the expressive power of complex proposition nodes in representing in a subtle and detailed way the meaning of natural language utterances.

Garrison W. Cottrell and Steven L. Small present in their paper (Viewing parsing as word sense discrimination: A connectionist approach) the first results that connectionism is obtaining in the study of natural language. This fairly new research paradigm utilizes a parallel architecture, significantly different from the sequential computation model of a usual computer and closer to that of a human brain. The kernel of this approach is represented by an active semantic network scheme, which entails a large number of interconnected computing units that communicate through weighted levels of excitation and inhibition.

Danilo Fum, Giovanni Guida, and Carlo Tasso present in their paper (A propositional language for text representation) a representation language which is proved to support a detailed and transparent formalization of the meaning of a natural language text. Several aspects of meaning representation, including quantification, reference, and time, are discussed in detail. A sketch for the implementation of a parser, able to generate a propositional representation from a natural language input text, is also described.

The paper by Eva Hajičová and Petr Sgall (From topic and focus of a sentence to linking in a text) analyzes the role of the topic-focus articulation in determining the structure of a text. A connection is established between topic-focus articulation and coreference. This kind of linking in a text is based on an account of the degree of salience of the elements belonging to a common knowledge base, shared by the speaker and the hearer.

The paper by Eduard H. Hovy and Roger C. Schank (Language generation by computer) is focused on the problem of language production. The research paradigm is the well known conceptual dependency theory, based on semantic primitives. A major point in this contribution is represented by the very refined process of language generation, which takes explicitely into account social, psychological, and pragmatic aspects of conversation to produce text tailored to the various types of hearers.

Jiang Xinsong, Li Yingtan, and Chen Yu present in their contribution

BIBLIOTHEQUE DU CERIST

(Understanding the Chinese language) a comprehensive survey of Chinese language processing. The paper discloses to Western researchers several of the peculiarities of Chinese language, and discusses the basic issues of the problem of Chinese processing by computer. The major research projects presently ongoing in China and the results so far achieved are presented as well.

The paper by Philip N. Johnson-Laird (Semantic primitives or meaning postulates: Mental models or propositional representations?) faces the problem of the deep structure of language. The procedural semantics paradigm he proposes, allows the meaning of words to be decomposed into more primitive notions: namely, mental models. The introduction of mental models resolves the controversy over whether semantic primitives are equivalent to sets of meaning postulates, and the controversy over whether images are equivalent to sets of propositions.

The paper by Wendy G. Lehnert (Narrative complexity based on summarization algorithms) deals with the information processing requirements involved in the memory representation of narrative structures. The author introduces the notion of algorithmically equivalent plot unit graphs, which characterize narratives that can be processed by the same summarization retrieval algorithm. The notion of algorithmic equivalence creates a partition - weaker than that produced by the usual relation of isomorphism - of the set of plot unit graphs into classes. Three of these classes, which appear to produce reasonable summaries for a majority of the narratives, are discussed in detail.

Kathleen R. McKeown presents in her paper (Using focus to constrain language generation) a mechanism for text generation, which utilizes the focus of attention to guide the ability to make reasoned choices both from a large number of possible things to say, and from a large number of expressive possibilities. A computational model of focus of attention is developed, and a procedure is proposed to select words and syntactic structures that best express the system's intent, thus ensuring that the resulting text is coherent.

Alexandre S. Narin'yani discusses in his paper (Towards an integral model of language competence) a comprehensive model of language competence which includes in an integrated scheme syntax, semantics, and pragmatics. This general paradigm comprehends a hierarchy of models at different levels: physical world, social environment, verbal interaction, speech acts, coherent text, sentences. The model of speech interaction, anaphora as a part of the coherent text model, and a fragment of the sentence model are considered and discussed in detail.

Brian Phillips presents in his paper (An object-oriented parser) an analyzer of natural language expressions devoted to deal in an effective way with structural ambiguity. The focus of his contribution is on how syntactic analysis and semantic interpretation can co-operate to resolve ambiguities as soon as possible, in order to prevent an explosive growth of alternatives. A left-corner parsing algorithm with a reachability matrix for top-down filtering of rule selection is adopted. The parser is governed by a breadth-first control paradigm, which includes a delay mechanism used to queue constituents until all alternative analyses of a phrase have been completed.

Bruno G. Bara
Giovanni Guida

BIBLIOTHEQUE DU CERIST

BIBLIOTHEQUE DU CERIST

CONTENTS

PREFACE vii

COMPETENCE AND PERFORMANCE
IN THE DESIGN OF NATURAL LANGUAGE SYSTEMS
Bruno G. Bara, Giovanni Guida 1

PLANNING AND UNDERSTANDING SPEECH ACTS BY INTERPERSONAL GAMES
Gabriella Airenti, Bruno G. Bara, Marco Colombetti 9

A FRAMEWORK FOR INTEGRATING SYNTAX AND SEMANTICS
Amedeo Cappelli, Giacomo Ferrari, Lorenzo Moretti,
Irina Prodanof 33

KNOWLEDGE REPRESENTATION AND NATURAL LANGUAGE:
EXTENDING THE EXPRESSIVE POWER OF PROPOSITION NODES
Cristiano Castelfranchi, Domenico Parisi, Oliviero Stock 59

VIEWING PARSING AS WORD SENSE DISCRIMINATION:
A CONNECTIONIST APPROACH
Garrison W. Cottrell, Steven L. Small 91

A PROPOSITIONAL LANGUAGE FOR TEXT REPRESENTATION
Danilo Fum, Giovanni Guida, Carlo Tasso 121

FROM TOPIC AND FOCUS OF A SENTENCE TO LINKING IN A TEXT
Eva Hajičová, Petr Sgall 151

LANGUAGE GENERATION BY COMPUTER
Eduard H. Hovy, Roger C. Schank 165

BIBLIOTHEQUE DU CERIST

UNDERSTANDING THE CHINESE LANGUAGE
Jiang Xinsong, Li Yingtan, Chen Yu 197

SEMANTIC PRIMITIVES OR MEANING POSTULATES:
MENTAL MODELS OR PROPOSITIONAL REPRESENTATIONS?
Philip N. Johnson-Laird 227

NARRATIVE COMPLEXITY BASED ON SUMMARIZATION ALGORITHMS
Wendy G. Lehnert 247

USING FOCUS TO CONSTRAIN LANGUAGE GENERATION
Kathleen R. McKeown 261

TOWARDS AN INTEGRAL MODEL OF LANGUAGE COMPETENCE
Alexandre S. Narin'yani 275

AN OBJECT-ORIENTED PARSER
Brian Phillips 297

AUTHOR INDEX 323

LIST OF CONTRIBUTORS 325

BIBLIOTHEQUE DU CERIST