Roland Hauser

COMPUTATION OF LANGUAGE

An Essay on Syntax, Semantics, and Pragmatics in Natural Man-Machine Communication

Foreword by Dana Scott

Springer Series
SYMBOLIC COMPUTATION – Artificial Intelligence

SPRINGER-VERLAG
BERLIN HEIDELBERG NEW YORK
LONDON PARIS TOKYO
Remark on this electronic copy

The original dvi-file from which the book was published does not exist any more. Therefore, a new dvi-file had to be derived from the .tex-files. As a consequence, the pagebreaks agree only roughly with the printed version. The page numbers of the author and subject index relate to the original publication. The graphic 4.1.6 on page 62 could not be reproduced in this electronic copy. Some typing errors have been corrected.

The book has been out of print for some time now. However, a few printed copies are currently still available at the Laboratory of Computational Linguistics at Friedrich Alexander University Erlangen Nürnberg. Readers interested in getting a copy for a nominal fee for shipping and handling should write to

Abteilung Computerlinguistik
Universität Erlangen Nürnberg
Bismarckstr. 12
D-91054 Erlangen

or send email to tommi@linguistik.uni-erlangen.de.


ISBN 3-540-5882-1 Springer-Verlag Berlin Heidelberg New York
ISBN 3-87-5882-1 Springer-Verlag Berlin Heidelberg New York

Library of Congress Cataloging-in-Publication Data
ISBN 3-387-5882-1 (U.S.)

©Springer-Verlag Berlin Heidelberg 1989
Printed in the United States of America
Foreword

The study of linguistics has been forever changed by the advent of the computer. Not only does the machine permit the processing of enormous quantities of text—thereby securing a better empirical foundation for conclusions—but also, since it is a modelling device, the machine allows the implementation of theories of grammar and other kinds of language processing. Models can have very unexpected properties—both good and bad—and it is only through extensive tests that the value of a model can be properly assessed. The computer revolution has been going on for many years, and its importance for linguistics was recognized early on, but the more recent spread of personal workstations has made it a reality that can no longer be ignored by anyone in the subject.

The present essay, in particular, could never have been written without the aid of the computer. I know personally from conversations and consultations with the author over many months how the book has changed. If he did not have at his command a powerful typesetting program, he would not have been able to see how his writing looked and exactly how it had to be revised and amplified. Even more significant for the evolution of the linguistic theory is the easy testing of examples made possible by the implementation of the parser and the computer-held lexicon. Indeed, the rule set and lexicon grew substantially after the successes of the early implementations created the desire to incorporate more linguistic phenomena.

As a general principle, the use of a computer not only forces the researcher to be more precise, but it also encourages him to be more ambitious. The use of the computer is an inspiration to thinking: once you pass a certain point of solidification of the implementation, you then find that whenever you get a new idea it is relatively easy to try it out. And we can share software and results with others, making independent verifications more feasible. I hope the fruits of the present work will be quickly tested in this way.

As the author explains in his preface, the book as a whole is divided into three parts, each of which has five chapters. This balance arose naturally as the shape of the essay emerged from the literary composition process. There are always many decisions to be made concerning the amount of detail that ought to be presented. Also, it is only after you start writing that you appreciate the strengths and weaknesses of your argument. Thus, over several months, the chapters split, moved, expanded, and contracted until the author was satisfied that the outline of the thesis he is putting forward was clear.
The emphasis in Part I is to give the general outlines of a theory of syntax, semantics, and pragmatics. As the author points out from the beginning, this theory did not arise from a preconceived theoretical standpoint: he was forced into it step by step by his efforts at implementation of a categorial grammar. After seeing what emerged from this work, he then realized that he could justify it on many other grounds. Once the main features of this justification are set out in Part I, the argument needs further underpinning, however. The mathematical mechanism of grammar construction is the topic of Part II of the book, where many results and comparisons with other approaches are presented.

I feel Roland Hausser is remarkably successful in showing that left-associative grammar is practical. I think he is also equally successful in showing in Parts I and II
Of course, the parsing of correct sentences is an excellent thing to do, but we also have to give much more thought to error analysis. It will be clear from the many examples in the book that the approach can be adapted to doing this, but more software design is needed to make this actual and useful. It should also be clear that the parsing of phrases as well as sentences can be done. For uses in information retrieval that we have in mind, it will be essential to do this flexibly. For example, there are well known constructions with discontinuous constituents in English where the underlying concept has to be located by means of parsing.

Also, as everyone who works in natural-language processing knows, the lexicon is never large enough, and something will have to be developed to allow the parser to “guess” at the category of words not contained in the lexicon. This could be very useful in doing empirical work in expanding the lexicon. In general, lexicon construction is very labor intensive, and machine aids are badly needed to make the task easier. In any case, I am much encouraged that the NEWCAT Parser is turning out to be a real tool. The further evaluation of the method will also be easier when we have additional languages to try out. There already exists a sizable LA-grammar for German, which was put aside during the writing of this book. I hope the author can return to this project soon.

I think it must be emphasized that this is an essay and not an exhaustive treatise. Much more could (and has!) been said on many of the topics discussed here. The present book, however, is an attempt to establish a concept by making a real effort to explain how it fits into a philosophical and psychological point of view. But the attempt will fail if it is drowned in too much detail. If the basic concept is accepted, the elaboration can come later.

The conclusions in this book may be controversial, but the author has provided—in a pleasant, non-dogmatic way—a new platform for a fresh and rational discussion. I hope linguists and philosophers will take the trouble to understand his theory and continue the debate.

Dana Scott  
Carnegie Mellon University  
February, 1989
Author’s Preface

My grandmother sometimes played dominoes\textsuperscript{1} with me. Later I was struck by the idea that language worked very much like the pieces of this game. Each piece you add must be compatible with the previous one, and it in turn determines what can be added next. This is the idea of \textit{possible continuations}.

Contemporary linguistic theory takes a different approach. It looks at language as a system of abstract sentence schemata: one gets a real sentence by replacing variables in the main schema with more elaborate sub-schemata, until finally the variables in the sub-schemata are replaced by actual words. This is the idea of \textit{possible substitutions}.

What is the difference between the two approaches? The substitution approach is based on partially \underline{ordered} replacements in abstract schemata: in a schema with
rethink the syntactic derivation process. An explicit statement of all possible continuations after each word of the sentence requires a more distributed packaging of the grammatical information than in a substitution system. But the basic intuitions of possible continuations are clear and firm: for any given “sentence start” one can easily determine which “next word” categories are grammatical, and which are not.

The resulting symbolic parallel distributed system of Left-Associative Grammar
of the formal system. Part II may be read independently of Part I, though Part II provides the formal definitions underlying the syntactic analysis outlined in Part I.

A brief conclusion is followed by three appendices describing different aspects of ECAT. ECAT is an LA-grammar of a sizeable fragment of English, which is implemented in Lisp as a parser.

The parsers for English (ECAT) and German (DCAT) were originally developed at CSLI-Stanford in INTERLISP-D on a Xerox 1109 machine. After translation into Common Lisp, ECAT was substantially expanded. The Common Lisp version of semantically interpreted ECAT presented in this book was developed on an HP-9000 work station, but has been successfully brought up on Micro-VAXes, TI-explorers, and SUN work stations.

The parsers for the natural and formal languages described in this book are all running at the Laboratory for Computational Linguistics of Carnegie Mellon University. The sample derivations contained in the following pages are unedited verbatim outputs of the parsers.

Acknowledgements

The research presented in this book was supported by a five year Heisenberg grant by the Deutsche Forschungsgemeinschaft, West Germany. During this time, I visited the Philosophy Department of Stanford University (1983/4), the Center for the Study of Language and Information at Stanford (1984/6), the Department of Computer Science at Carnegie Mellon University (1986/7), and the Laboratory for Computational Linguistics at Carnegie Mellon University (1987/8). I am grateful to Julius Moravcik, Stanley Peters, Betsy Macken, John Perry, Jaime Carbonell, Masaru Tomita, David Evans, and Dana Scott for sponsoring my stays at their research institutions.

Portions of Chapters 3, 4, 6, and 7 were previously published in Computers and Translation. Passages in Chapters 14 and 15 were lifted from articles in the Journal of Semantics, Conceptus and Language Research. I wish to acknowledge my indebtedness to the editors of these journals. Otherwise, to the best of my recollection, no portions of the work have been previously published.

Edward Gibson, Stanley Peters, Stuart Shieber, Helmut Schwichtenberg, Dana Scott, and Masaru Tomita offered vigorous resistance, as well as constructive ideas, to various preliminary views and formulations, thus contributing greatly to the outcome of this book. But they are in no way responsible for the result.

I wish to thank Jaime Carbonell, Brian McWhinney, Teruko Mitamura, Sergei Nirenburg, Carl Pollard, and David Turoetzky for helpful discussions and important suggestions. Crucial help with programming problems at various times was provided.
by Jaime Carbonell, Todd Kaufmann, Eric Nyberg, and Peter Shell. Johanna Seibt and Robert Carpenter provided detailed comments on the manuscript, which resulted in numerous corrections and improvements. I was fortunate in being able to discuss much of the material with the members of my seminars and lecture courses at the University of Munich.

The book was written between November 1, 1987 and September 30, 1988 at the Laboratory for Computational Linguistics, Carnegie Mellon University, Pittsburgh, Pennsylvania. The production of this book would have been virtually impossible without the \LaTeX software running on the workstations provided by CMU. I am grateful for access to the computing facilities of the Laboratory for Computational Linguistics and the Department of Computer Science.

I am indebted to Kathryn Gula, who created the many difficult pictures in \LaTeX. I am grateful to David Sours, who proofread and copyedited the text meticulously and with great skill. Also I would like to thank the editor at Springer-Verlag Heidelberg, Dr. Hans Wössner, for his support. The book was produced from a camera-ready copy supplied by the author.
Contents

Foreword by Dana Scott V

Author’s Preface IX

Acknowledgements XI

Introduction 3

I Natural Language and Formal Grammar 9

1 Goals and Results 11
   1.1 A Unified Perspective 11
   1.2 Assumptions Accepted and Assumptions Rejected 13
   1.3 Theory and Practice 16
   1.4 Summary of the Formal Results 18
   1.5 Applications and Future Work 21

2 Grammar and Interpretation 23
   2.1 Limitations of Constituent Structures 23
   2.2 Basic Communication 25
   2.3 Database Semantics and Pragmatic Interpretation 27


III Logic and Communication

11 Principles of Pragmatics
11.1 Peirce and the Theory of Signs .................. 267
11.2 Meaning, Use, and Grice’s Intentions .......... 271
11.3 The Principles of Reference ..................... 274
11.4 Successful Communication ...................... 279
11.5 Non-Literal Uses .................................. 282

12 Meaning, Truth, and Ontology .................. 289
12.1 Tarski’s Truth Definition ......................... 289
12.2 Model Theory and Natural Language ............ 292
12.3 A Simple Robot: the Color Reader ............... 295
12.4 Intensional and Extensional Meaning Analysis .... 298
12.5 Four Basic Approaches to Meaning ............... 301

13 Model Theory and Artificial Intelligence .......... 305
13.1 Reconstructing Truth in a Robot ................. 305
13.2 Autonomy from the Metalanguage ................. 307
13.3 Semantics and Reference ......................... 310
13.4 Why Meanings are in the Head .................... 311
13.5 The Speaker-Simulation Device .................. 313

14 Reference and Denotation ......................... 319
14.1 The Epimenides Paradox ......................... 319
14.2 Consistency and Truth .......................... 322
14.3 Reconstructing Epimenides Pragmatically ......... 323
14.4 Vagueness ....................................... 326
14.5 The Sorites Paradox ............................. 329

15 Surface Compositional Semantics ................. 333
15.1 Presuppositions .................................. 333
15.2 Restricted Quantification ......................... 336
15.3 Tautologies, Presupposition Failure, and Vagueness .... 338
15.4 Intensional Contexts ............................ 341
15.5 A Unified Semantic Account ..................... 344

Conclusion ........................................ 351

Appendices ........................................ 357
Introductory Remarks ............................. 357
A ECAT Category Segments and Categories
   A.1 Alphabetical List of Category Segments ................. 359
   A.2 List of ELEX categories .................................. 366
   A.3 Derived ECAT categories .................................. 371

B Sample Derivations .............................................. 373
   B.1 Yes/No Interrogative ..................................... 373
   B.2 Ambiguity in a Passive Sentence ......................... 374
   B.3 Recursion of Control ..................................... 377
   B.4 Unbounded Dependency ................................... 381
   B.5 Relative Clauses ......................................... 383
   B.6 Genitive Recursion ....................................... 384
   B.7 Idiom .......................................................... 386

C The Current Example Set of ECAT .............................. 389

References .......................................................... 403

Name Index .......................................................... 417