Lecture Notes in Computer Science

Edited by G. Goos and J. Hartmanis

377

M. Gross D. Perrin (Eds.)

Electronic Dictionaries and Automata in Computational Linguistics

LITP Spring School on Theoretical Computer Science
Saint-Pierre d'Oléron, France, May 25–29, 1987
Proceedings

Springer-Verlag
Berlin Heidelberg New York London Paris Tokyo Hong Kong

Editors

Maurice Gross
Université Paris 7, LADL
2 place Jussieu, F-75251 Paris Cedex 05, France

Dominique Perrin
Université Paris 7, LITP
2 place Jussieu, F-75251 Paris Cedex 05, France

CR Subject Classification (1989): I.7, I.2.7, F.2.2

ISBN 3-540-51465-1 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-51465-1 Springer-Verlag New York Berlin Heidelberg

Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.
2145/3140-543210 – Printed on acid-free paper

Foreword

This volume gathers the texts of lectures given at the 15th Spring School in Theoretical Computer Science. This meeting was devoted to the relations between Formal Systems (automata, codes, grammars) and Natural Language Processing. It was organized jointly by M. Borillo, M. Gross, M. Nivat and D. Perrin and took place at Saint-Pierre d'Oléron in May 1987.

The two fields of Formal Systems and Natural Language Processing have strong historical links as exemplified by the joint work of N. Chomsky and M.P. Schützenberger[1] and the studies of Z. S. Harris on formalization[2] . Subsequently, they developed independently. Linguists stopped using elementary mathematical models either to characterize complete grammars or even to describe limited linguistic phenomena. They deepened their linguistic studies and built more specific formal representations. On their side, computer scientists have found a natural domain of application of formal systems in Compilation Theory and in the study of algorithms on strings[3].

With recent developments in text processing, the need for high performances is again bringing together the two fields. In the same way, since libraries can now contain all publications in their computer form, the large size of their collections requires sophisticated processing, even on the most powerful computers.

One basic component of linguistic systems is the dictionary look-up procedure : words of current texts have to be matched with the words of a given dictionary. For example, spelling checkers are based on this component. Many algorithmic questions then arise. They can be shaped according to specific applications and to the amount of information available about the texts and in the dictionaries. A nonexhaustive list is the following :

- fast string searching, with or without loss of information,
- fast searching of families of strings,
- coding texts and dictionaries in order to compact them,
- the use of large dictionaries of finite automata without loops for parsing,
- parsing of sentences of ambiguous languages,
- error correction in texts,
- the statistics on the distribution of words in texts.

The papers presented at the Spring School ranged from purely theoretical studies to detailed grammatical descriptions. Although a continuum of interests characterized the School, only

---

[1] Chomsky, N., Schützenberger M.-P. 1963 : The Algebraic Theory of Context-Free Languages, in *Computer Programming and Formal Systems*, P. Braffort and L. Hirschberg eds., North Holland Publishing Co.

[2] Z.S. Harris 1968 *Mathematical Structure of Language*, New York : Wiley Interscience.

[3] Aho, A., Sethi, R., Ullman, J. 1986 *Compilers, Principles and Tools*, Addison Wesley.

those papers which bore directly on computational issues have been included in the present volume. The other papers will be published in a special issue of *Lingvisticae Investigationes* (Amsterdam-Philadelphia : J. Benjamins) to appear in 1989.

Financial support came from the two **Programme de Recherches Coordonnées** *Mathématiques et Informatique* and *Informatique Linguistique* of the Ministry of Research and Higher Education, and from the *FIRTECH Industries de la langue* of the Ministry of Education. We express our thanks the three laboratories of the *CNRS* which helped in the organization of the meeting (*LADL* and *LITP* and *LSI* of Toulouse). Special mention of Mrs. Colette Ravinet for her perfect handling of all material questions has to be made.

Paris, January 1989 M. Gross, D. Perrin

Contents