

Preface

The need for a comprehensive survey-type exposition on formal languages and related mainstream areas of computer science has been evident for some years. In the early 1970s, when the book *Formal Languages* by the second-mentioned editor appeared, it was still quite feasible to write a comprehensive book with that title and include also topics of current research interest. This would not be possible anymore. A standard-sized book on formal languages would either have to stay on a fairly low level or else be specialized and restricted to some narrow sector of the field.

The setup becomes drastically different in a collection of contributions, where the best authorities in the world join forces, each of them concentrating on their own areas of specialization. The present three-volume Handbook constitutes such a unique collection. In these three volumes we present the current state of the art in formal language theory. We were most satisfied with the enthusiastic response given to our request for contributions by specialists representing various subfields. The need for a Handbook of Formal Languages was in many answers expressed in different ways: as an easily accessible historical reference, a general source of information, an overall course-aid, and a compact collection of material for self-study. We are convinced that the final result will satisfy such various needs.

The theory of formal languages constitutes the stem or backbone of the field of science now generally known as theoretical computer science. In a very true sense its role has been the same as that of philosophy with respect to science in general: it has nourished and often initiated a number of more specialized fields. In this sense formal language theory has been the origin of many other fields. However, the historical development can be viewed also from a different angle. The origins of formal language theory, as we know it today, come from different parts of human knowledge. This also explains the wide and diverse applicability of the theory. Let us have a brief look at some of these origins. The topic is discussed in more detail in the introductory Chapter 1 of Volume 1.

The main source of the theory of formal languages, most clearly visible in Volume 1 of this Handbook, is *mathematics*. Particular areas of mathematics important in this respect are combinatorics and the algebra of semi-groups and monoids. An outstanding pioneer in this line of research was

Axel Thue. Already in 1906 he published a paper about avoidable and un-avoidable patterns in long and infinite words. Thue and Emil Post were the two originators of the formal notion of a rewriting system or a grammar. That their work remained largely unknown for decades was due to the difficult accessibility of their writings and, perhaps much more importantly, to the fact that the time was not yet ripe for mathematical ideas, where noncommutativity played an essential role in an otherwise very simple setup.

Mathematical origins of formal language theory come also from mathematical logic and, according to the present terminology, computability theory. Here the work of Alan Turing in the mid-1930s is of crucial importance. The general idea is to find models of computing. The power of a specific model can be described by the complexity of the language it generates or accepts. Trends and aspects of mathematical language theory are the subject matter of each chapter in Volume 1 of the Handbook. Such trends and aspects are present also in many chapters in Volumes 2 and 3.

Returning to the origins of formal language theory, we observe next that much of formal language theory has originated from *linguistics*. In particular, this concerns the study of grammars and the grammatical structure of a language, initiated by Noam Chomsky in the 1950s. While the basic hierarchy of grammars is thoroughly covered in Volume 1, many aspects pertinent to linguistics are discussed later, notably in Volume 2.

The *modeling* of certain objects or phenomena has initiated large and significant parts of formal language theory. A model can be expressed by or identified with a language. Specific tasks of modeling have given rise to specific kinds of languages. A very typical example of this are the L systems introduced by Aristid Lindenmayer in the late 1960s, intended as models in developmental biology. This and other types of modeling situations, ranging from molecular genetics and semiotics to artificial intelligence and artificial life, are presented in this Handbook. Words are one-dimensional, therefore linearity is a feature present in most of formal language theory. However, sometimes a linear model is not sufficient. This means that the language used does not consist of words (strings) but rather of trees, graphs, or some other nonlinear objects. In this way the possibilities for modeling will be greatly increased. Such extensions of formal language theory are considered in Volume 3: languages are built from nonlinear objects rather than strings.

We have now already described the contents of the different volumes of this Handbook in brief terms. Volume 1 is devoted to the mathematical aspects of the theory, whereas applications are more directly present in the other two volumes, of which Volume 3 also goes into nonlinearity. The division of topics is also reflected in the titles of the volumes. However, the borderlines between the volumes are by no means strict. From many points of view, for instance, the first chapters of Volumes 2 and 3 could have been included in Volume 1.

We now come to a very important editorial decision we have made. Each of the 33 individual chapters constitutes its own entity, where the subject matter is developed from the beginning. References to other chapters are only occasional and comparable with references to other existing literature. This style of writing was suggested to the authors of the individual chapters by us from the very beginning. Such an editorial policy has both advantages and disadvantages as regards the final result. A person who reads through the whole Handbook has to get used to the fact that notation and terminology are by no means uniform in different chapters; the same term may have different meanings, and several terms may mean the same thing. Moreover, the prerequisites, especially in regard to mathematical maturity, vary from chapter to chapter. On the positive side, for a person interested in studying only a specific area, the material is all presented in a compact form in one place. Moreover, it might be counterproductive to try to change, even for the purposes of a handbook, the terminology and notation already well-established within the research community of a specific subarea. In this connection we also want to emphasize the diversity of many of the subareas of the field. An interested reader will find several chapters in this Handbook having almost totally disjoint reference lists, although each of them contains more than 100 references.

We noticed that guaranteed timeliness of the production of the Handbook gave additional impetus and motivation to the authors. As an illustration of the timeliness, we only mention that detailed accounts about DNA computing appear here in a handbook form, less than two years after the first ideas about DNA computing were published.

Having discussed the reasons behind our most important editorial decision, let us still go back to formal languages in general. Obviously there cannot be any doubt about the mathematical strength of the theory – many chapters in Volume 1 alone suffice to show the strength. The theory still abounds with challenging problems for an interested student or researcher. Mathematical strength is also a necessary condition for applicability, which in the case of formal language theory has proved to be both broad and diverse. Some details of this were already mentioned above. As the whole Handbook abounds with illustrations of various applications, it would serve no purpose to try to classify them here according to their importance or frequency. The reader is invited to study from the Handbook older applications of context-free and contextual grammars to linguistics, of parsing techniques to compiler construction, of combinatorics of words to information theory, or of morphisms to developmental biology. Among the newer application areas the reader may be interested in computer graphics (application of L systems, picture languages, weighted automata), construction and verification of concurrent and distributed systems (traces, omega-languages, grammar systems), molecular biology (splicing systems, theory of deletion), pattern matching, or cryptology, just to mention a few of the topics discussed in the Handbook.

About Volume 2

Some brief guidelines about the contents of the present Volume 2 follow. Problems about complexity occur everywhere in language theory; Chapter 1 gives an overall account. Parsing techniques are essential in applications, both for natural and programming languages. They are dealt with in Chapter 2, while Chapters 3–6 study extensions and variations of classical language theory. While Chapter 3 continues the general theory of context-free languages, Chapters 5 and 6 are motivated by linguistics, and Chapter 4, motivated by artificial intelligence, is also applicable to distributed systems. DNA computing has been an important recent breakthrough – some language-theoretic aspects are presented in Chapter 7. Chapter 8 considers the string editing problem which in various settings models a variety of problems arising from DNA and protein sequences. Chapter 9 considers several methods of word matching that are based on the use of automata. Chapter 10 discusses the relationship between automata theory and symbolic dynamics (the latter area has originated in topology). By its very nature the whole of cryptology can be viewed as a part of language theory. Chapter 11 gives an account of language-theoretic techniques that have turned out to be especially useful in cryptology.

Acknowledgements

We would like to express our deep gratitude to all the authors of the Handbook. Contrary to what is usual in case of collective works of this kind, we hoped to be able to keep the time schedule – and succeeded because of the marvellous cooperation of the authors. Still more importantly, thanks are due because the authors were really devoted to their task and were willing to sacrifice much time and energy in order to achieve the remarkable end result, often exceeding our already high expectations.

The Advisory Board consisting of J. Berstel, C. Calude, K. Culik II, J. Engelfriet, H. Jürgensen, J. Karhumäki, W. Kuich, M. Nivat, G. Păun, A. Restivo, W. Thomas, and D. Wood was of great help to us at the initial stages. Not only was their advice invaluable for the planning of the Handbook but we also appreciate their encouragement and support.

We would also like to extend our thanks from the actual authors of the Handbook to all members of the scientific community who have supported, advised, and helped us in many ways during the various stages of work. It would be a hopeless and maybe also unrewarding task to try to single out any list of names, since so many people have been involved in one way or another.

We are grateful also to Springer-Verlag, in particular Dr. Hans Wössner, Ingeborg Mayer, J. Andrew Ross, and Gabriele Fischer, for their cooperation, excellent in every respect. Last but not least, thanks are due to Marloes Boon-van der Nat for her assistance in all editorial stages, and in particular, for keeping up contacts to the authors.