Advanced Programming in PROLOG for Computational Linguistics and Artificial Intelligence

RODRIGUE SABIN MOMPELAT
Digital Preservation Dept., Copenhagen University Library

Prolog is a programming language particularly suitable for developing programs in research areas such as corpus linguistics, machine translation, concordance compilation, testing of grammars and many other areas where statistical counts and analyses are required. We will restrict our linguistic and computational techniques to the elaboration of frequency lists and indexes on the basis of the English and Danish languages.

Prolog is also suitable for solving problems concerned with the computerized implementation of intelligent human behaviour such as presentation of knowledge, teaching, learning, association, planning, explaining, acquisition of language etc. We will examine powerful computational tools, the automata, which can analyse, recognize and generate languages, and whose representation by network can apply to many contexts.

Moreover, Prolog has intrinsic qualities that make its success. It enables to construct databases and their corresponding management system. For that matter, we will develop not only different kinds of procedures of manipulation of information in the databases, but also a module-based database management system which will have to be reliable, robust and user-friendly.

Our approach of Prolog endeavours to make the user capable of using and building systems such as natural language interfaces, terminology and dictionary databases or text-processing programs of great machine-readable text corpora.

ISBN 978 3 89586 359 2 (Hardbound). **LINCOM Studies in Computational Linguistics 01**. 246pp. USD 151.10 / EUR 102.80 / GBP 74.00. 2007.

Mathematical Foundations of Linguistics

H. MARK HUBEY
Montclair State University

Only a few decades ago, only mathematicians, physicists and engineers took calculus courses, and calculus was tailored for them using examples from physics. This made it difficult for students from the life sciences including biology, economics, and psychology to learn mathematics. Recently books using examples from the life sciences and economics have become more popular for such students. Such a math book does not exist for linguists. Even the computational linguistics books (Formal Language Theory) are written for mathematicians and computer scientists.

This book is for linguists. It is intended to teach the required math for a student to be a scientific linguist and to make linguistics a science on par with economics, and computer science.

There are many concepts that are central to the sciences. Most students never see these in one place and if they do, they have to wait until graduate school to obtain them in the often-dreaded "quantitative" courses. As a result sometimes it takes years or even decades before learners are able to integrate what they have learned into a whole, if ever. We have little time and much to do.

In addition to all of these problems we are now awash in data and information. It is now that the general public should be made aware of the solution to all of these problems. The answer is obviously "knowledge compression". Knowledge is structured information; it is a system not merely a collection of interesting facts.

What this book does, and what all other math books do is teach people the tools with which they can structure and thus compress information and knowledge around them. It has also been said that mathematics is the science of patterns; it is exactly by finding such patterns that we compress knowledge. We can say that mathematics is the science of knowledge compression or information compression.

This book provides the basic tools for mathematics (even including a short and intuitive explanation of differential and integral calculus). The broad areas of linguistics, probability theory, speech synthesis, speech recognition, computational linguistics (formal languages and machines), historical linguistics require mathematics of counting/combinatorics, Bayesian theory, correlation-regression analysis, stochastic processes, differential equations, vectors/tensors. These in turn are based on set theory, logic, measurement theory, graph theory, algebra, Boolean algebra, harmonic analysis etc.

The mathematical fields introduced here are all common ideas from one which one can branch off into more advanced study in any of these fields thus this book brings together ideas from many disparate fields of mathematics which would not normally be put together into a single course. This is what makes this a book especially written for linguists.

ISBN 3 89586 641 5. **LINCOM Handbooks in Linguistics 10**. 260pp. USD 88.80 / EUR 60.40 / GBP 43.50. 1999.

Mathematical and Computational Linguistics

H. MARK HUBEY
Montclair State University

As Lass (1980) has remarked, "system" is something talked about constantly in linguistics but never beyond paying just lip-service to the concept. This book shows how linguistics constitutes a "system". Linguists (except those who study Formal Language Theory) are confronted with a dilemma. What they study is partially based on physics and is in many respects mathematical; yet the mathematics books are divorced from linguistics and linguistics books are divorced from mathematics and physics. There are no books that teach mathematics for linguists or linguistics with mathematics. This book goes a long way toward accomplishing the integration of mathematics, physics and linguistics into a whole, in other words "a system", just like those that are studied by others in the quantitative disciplines such as physics, engineering, computer science or economics.

The methods of mathematics which are used in the books to elucidate system concepts and others that are needed in linguistics includes boolean algebra, differential equations, and fuzzy logic.

Furthermore it also explains in an intuitive manner,those concepts are not only from mathematics but also from the underlying physics and engineering up to and including acoustic theory of speech, speech recognition, and even nonlinearity/catastrophe theory and quantality of phonemic systems.

All the mathematics needed to form the mathematical foundations of linguistics is illustrated with examples from linguistics and thus may be thought of as "theories", those that should replace the standard literary linguistics tradition in the same way that literary economics is no longer the de facto standard. Physical/acoustic theory of speech is blended naturally into the phonological and phonetic standard, and the standard works are used as springboards to the development of vector space concepts that are necessary for comprehension of new works in speech synthesis and speech recognition. It is rather easy then to show how seemingly nonrelated topics such as sonority scales, child language development, and various linguistics processes such as assimilation, metathesis, fortition/lenition can be seen to be a part of the greater whole. Historical processes are also treated in terms of sound change and also in terms of the most basic ideas which are needed for a thorough understanding of the problems such as multiple scale phenomena, distance and similarity, probability theory, and stochastic processes. A book of this length cannot possibly discuss all of the mathematics necessary in detail, however, there is sufficient material to motivate the topics, and furthermore to point in the direction of further study.

ISBN 3 89586 639 3. **LINCOM Handbooks in Linguistics 09**. 450pp. USD 140.80 / EUR 95.80 / GBP 69.00. 1999.

Lexical-semantic information in Head-driven Phrase Structure Grammar and Natural Language Processing

MARTIN HOELTER
University of Bochum

Lexical-semantic information traditionally has certainly not been in the focus of formal grammar theories or computational linguistics. With the recent emergence of complex Natural Language Processing (NLP) systems, however, an immediate practical need for semantically richer but formalized language-related information has arisen. The question then is how e.g. selectional restriction information which normally has been dealt with by separate models and theories can be integrated in a formal grammar framework also adequate for NLP tasks.

The suitable grammar theory employed is Head-driven Phrase Structure Grammar (HPSG), whose complex feature structure models and attribute-value matrices as major means of representation have made it the linguistic framework of choice in state-of-the-art NLP. Cobuild dictionaries on the other hand provide the ideal lexicographic framework for a data base: their unique definition style using simple English sentences only is shown to be both of high theoretical and practical relevance.

"Lexical-semantic information in Head-driven Phrase Structure Grammar and Natural Language Processing" proposes a concept of information transfer between natural language dictionaries, formal grammar, and language engineering systems by mapping dictionary definitions to HPSG lexical entries. The aim is to show how retrieval of lexical-semantic information from a Cobuild dictionary can be organized and what the theoretical assumptions on lexical-semantic information in a syntactically and semantically integrative HPSG model are.

ISBN 3 89586 583 4. **LINCOM Studies in Theoretical Linguistics 10**. 200pp. USD 105.30 / EUR 81.00 / GBP 56.70. 1999.