Book Reviews

The Mathematics of Language

Marcus Kracht
(University of California, Los Angeles)

Berlin: Mouton de Gruyter (Studies in generative grammar, volume 63), 2003, xvi+589 pp; hardbound, ISBN 3-11-017620-3, $127.00, €98.00

Reviewed by
Hans-Jörg Tiede
Illinois Wesleyan University

Mathematical linguistics is concerned with the study of mathematical properties of natural languages and linguistic theories. Since the mathematical properties of interest to mathematical linguists are usually from theoretical computer science (complexity classes, language hierarchies, formal learnability), mathematical linguistics can be considered to be an area of theoretical computational linguistics. However, since statistical methods are rarely used in mathematical linguistics, its relationship to current practices in computational linguistics is somewhat limited.

While the introduction of logic in linguistic research originally came from semantics, this line of work did not really use sophisticated metaresults. One of the main developments in mathematical linguistics in the last decade has been the introduction of sophisticated logical methods to the study of natural language syntax, for instance, the use of cut elimination and interpolation theorems in categorial grammar or Büchi's theorem relating finite automata and monadic second-order logic in model-theoretic syntax. The book under review is written by one of the main contributors to the logical turn in mathematical linguistics, and so it is not surprising that this is its main focus.

The contents of the book are as follows: Chapter 1, "Fundamental Structures," gives a concise introduction to the mathematical background needed for the rest of the book. It does not give an introduction to logic at this point, which is introduced as needed in the rest of the book. Chapter 2, "Context Free Languages," starts with the regular languages and then discusses normal forms, parsing, and ambiguity. It concludes with a proof of Parikh's theorem, which states that every context-free language is semilinear, and a discussion of non-context-free phenomena in natural languages. Chapter 3, "Categorial Grammar and Formal Semantics," contains, in addition to an introduction to the λ-calculus and combinators, an introduction to the Lambek calculus, culminating in a complete proof of Pentus's theorem, which states that Lambek grammars are context-free. Chapter 4, "Semantics", introduces algebraic tools for the study of natural language semantics. Chapter 5, "PTIME Languages," discusses extensions of context-free grammars, including tree-adjoining grammars, indexed grammars, and literal movement grammars. It includes a discussion of the class of mildly context-sensitive languages, which are widely held to be fairly good approximations to the complexity of natural languages. The last chapter, chapter 6, "The Model Theory of Linguistic Structure", is an introduction to us-

ing the approach of descriptive complexity theory to define conditions on strings and trees, as well as phonological representations. The chapter concludes with applications of the logical tools to grammar formalisms, including GPSG, HPSG, and GB.

As can be seen from the contents, the main focus of this book is on syntax, specifically the application of formal language theory and logic to natural languages. At this point, the classical introduction to this area is still Partee, ter Meulen, and Wall (1990). Kracht's book is significantly more advanced; in fact, a good working knowledge of Partee et al. is a prerequisite to Kracht's book. The *Handbook of Logic and Language* (van Benthem and ter Meulen 1997), which contains advanced material in this area, is much more focused on semantics than Kracht's book. Compared to introductory (e.g., Hopcroft and Ullman 1979) or advanced (e.g., Martín-Vide, Mitrana, and Păun 2004) books on formal language theory, Kracht's book emphasizes those aspects of formal language theory that are relevant to the study of natural languages, whereas the former do not. Thus, Kracht's book gives a uniform introduction, which currently does not exist at this level, to an important area of mathematical linguistics. Its main use will be in advanced graduate courses and for researchers interested in learning about mathematical linguistics.

The book stems from lecture notes that the author produced for a number of classes in this area; however, on the continuum that ranges from textbooks to research monographs, this book is located somewhere in the middle. For instance, from the point of view of a textbook, the importance that monadic second-order logic currently plays in mathematical linguistics would have warranted spending a larger part of chapter 6 on it rather than on quantified modal logic, which is a somewhat idiosyncratic choice. Such choices occur at a few other places in the book; however, they are balanced by the almost encyclopedic overview of formal grammars and important results about them that cannot currently be found in one volume. Another strength of this book is that it introduces logical tools incrementally together with the application for which they are needed. This reinforces the central role that logic plays in mathematical linguistics and makes it possible to read the book as an introduction to applied logic.

There are some minor inconsistencies. For instance, indexed grammars are introduced in the chapter on PTIME languages, and it is claimed that languages generated by indexed grammars can be parsed in PTIME, even though the recognition problem for indexed grammars is known to be NP-complete (Rounds 1973).

The chapter on semantics is a little too brief. Given the length of the book, it is understandable that no more space could be dedicated to it; however, it might have been more productive to replace that material with a more detailed description of formal models of GB, since this is an area in which Kracht has obtained many important results. Computational linguists will probably feel that parsing is not discussed in enough detail. While chart parsing is discussed for context-free grammars, a large number of papers on parsing of mildly context-sensitive grammars can be found in conference proceedings, but there is no uniform introduction to this area at this time.

While the book contains many exercises of different degrees of difficulty, I couldn't find any open research problems. This is unfortunate given that most readers of this book will likely be researchers and Ph.D. students.

Overall this book is an excellent introduction to advanced topics in mathematical linguistics that, given its advanced nature, requires a significant amount of mathematical maturity.

References

Hopcroft, John E. and Jeffrey D. Ullman. 1979. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley.

Martín-Vide, Carlos, Victor Mitrana, and Gheorghe Păun, editors. 2004. *Formal Languages and Applications*. Springer.

Partee, Barbara, Alice ter Meulen, and Robert E. Wall. 1990. *Mathematical Methods in Linguistics*. Kluwer.

Rounds, William C. 1973. Complexity of recognition in intermediate-level languages. In *14th Annual IEEE Symposium on Switching and Automata Theory (University of Iowa, Iowa City, Iowa, 1973)*, pages 145–158. IEEE Computer Society, Northridge, CA.

van Benthem, Johan and Alice ter Meulen, editors. 1997. *Handbook of Logic and Language*. MIT Press.

Hans-Jörg Tiede is an assistant professor of computer science at Illinois Wesleyan University. His research interest is the application of logic to mathematical and computational linguistics. Tiede's address is Department of Mathematics and Computer Science, Illinois Wesleyan University, P.O. Box 2900, Bloomington, IL 61702-2900; e-mail: htiede@iwu.edu.