The Mathematics of Syntactic Structure: Trees and their Logics

Hans-Peter Kolb and Uwe Mönnich (editors)
(Eberhard-Karls-Universität Tübingen)

Berlin: Mouton de Gruyter (Studies in
Generative Grammar, edited by Jan
Koster and Henk van Riemsdijk,
volume 44), 1999, 347 pp; hardbound,
ISBN 3-11-016273-3, $127.75, DM 198.00

Reviewed by
Gerald Penn
Bell Laboratories, Lucent Technologies

Regular languages correspond exactly to those languages that can be recognized by a finite-state automaton. Add a stack to that automaton, and one obtains the context-free languages, and so on. Probably all of us learned at some point in our university studies about the Chomsky hierarchy of formal languages and the duality between the form of their rewriting systems and the automata or computational resources necessary to recognize them. What is perhaps less well known is that yet another way of characterizing formal languages is provided by mathematical logic, namely in terms of the kind and number of variables, quantifiers, and operators that a logical language requires in order to define another formal language. This mode of characterization, which is subsumed by an area of research called **descriptive complexity theory**, is at once more declarative than an automaton or rewriting system, more flexible in terms of the primitive relations or concepts that it can provide resort to, and less wedded to the tacit, not at all unproblematic, assumption that the right way to view any language, including natural language, is as a set of strings.

This book is a collection of recent papers that pursue the connection between logical and natural languages further, mainly with regard to monadic second-order logic (MSO). Their relevance is apparent in the light of three earlier results. The first is the considerable amount of previous work in formal language theory on "tree languages." Given the extent to which trees and concepts based on tree position are used to express linguistic regularities, it is natural to think of using trees rather than strings as the basic elements of a formal language. It is also known that the tree languages definable in MSO are precisely those whose yields are the context-free string languages. The second is Rogers's landmark thesis (Rogers 1994), which demonstrated that significant portions of Government-and-Binding Theory (GB) can be very transparently expressed in MSO. The third is a collection of evidence from Dutch and Swiss German (Huybregts 1976, 1984; Shieber 1985) for the non-context-freeness of at least some natural languages. Thus, while MSO may be a very elegant choice, it does not quite appear to be a sufficient choice.

While no clear answer to this dilemma presents itself here, several of the chapters in this volume do address this issue directly. Rogers shows that dropping the usual requirement of bounded branching on trees leads to the same expressive power as infinite context-free grammars that are finitely presentable as regular expressions of context-free rules. These can be used in certain accounts of coordination. Morawietz shows how MSO can be thought of as a constraint language and thereby incorporated into a constraint logic programming language, which can then generate any

computable language. Mönnich demonstrates how to use a kind of "macro grammar" to define tree languages in MSO for the non-context-free phenomena that one can find in natural languages by replacing the normal substitution process that rewriting systems assume in their notion of derivation by a higher-order variant in which operations such as composition can be explicitly represented as objects in the macro language.

Other chapters approach tree description from different perspectives. Kallmeyer presents a tree description language that allows for true underspecification of dominance relations rather than just their path lengths. Kracht generalizes trees to handle some of the thorny formal issues surrounding the use of "categories," complexes of possibly several tree nodes that play a significant role in Chomsky's Barriers theory. There are also two chapters (by Kolb and Cornell) that discuss the application of declarative approaches to tree description to Chomsky's much more derivation-based Minimalist Program. In overall content, the book offers a very pleasing balance between formal rigor and linguistic application and a very well-informed set of references.

To say the least, this is not a book for the mathematically faint of heart. It must be said, however, that the editors and authors have managed to organize the book so that it is remarkably self-contained on the formal side. The only notable exception is the paper by Palm on a first-order alternative for formulating linguistic principles, which relies too heavily on a report published elsewhere to address the numerous counterexamples and exceptions to the near-theorems it presents. Otherwise, the presumed familiarity with the Minimalist Program in the later chapters may actually pose a greater barrier to comprehension by likely readers who fall outside the book's intended audience, i.e., researchers and graduate students in linguistics with a strong background in logic. The book also suffers from a rather large number of grammatical and typographical errors, including several in formal definitions, which makes the reading difficult in places.

Another slight problem, which is really a more systemic one in mathematical linguistics, is that no direct attempt is made to address other trends in generative linguistics that have a bearing on the subject under formal consideration. The use of parallel, yet very different levels of syntactic representation, such as in Lexical-Functional Grammar (LFG), or of more than one kind of constituent, as in the Prague school of linguistics, leaves open the possibility of using MSO to formulate constraints on multiple representations that, when intersected or otherwise combined, may produce very elegant analyses of non-context-free phenomena. There is also an established precedent in computer science of using MSO on richer graph-based languages than tree languages (Courcelle 1988; Engelfriet 1991), which might make MSO a realistic option for formalizing principles over feature-structure-based theories of language in a more constrained fashion. A courteous attempt is, in fact, made to relate the importance of this research to schools of linguistic thought other than GB/Minimalism, e.g., to LFG or Head-driven Phrase Structure Grammar, but for the most part these are throw-away remarks or outright misrepresentations.

The Mathematics of Syntactic Structure is, nevertheless, a faithful representation of the state of the art in this area, and can be recommended as a useful primer for computational linguists on descriptive complexity theory and its applications to linguistic theory.

References

Courcelle, Bruno. 1988. Some applications of logic, of universal algebra, and of category theory to the theory of graph transformations. *EATCS Bulletin*, 36:161–218.

Engelfriet, Joost. 1991. A regular characterization of graph languages definable in monadic second-order logic. *Theoretical Computer Science*, 88(1):139–150.

Huybregts, M. A. C. 1976. Overlapping dependencies in Dutch. Technical Report, Utrecht Working Papers in Linguistics.

Huybregts, M. A. C. 1984. The weak adequacy of context-free phrase structure grammar. In Ger J. de Haan, Mieke Trommelen, and Wim Zonneveld, editors, *Van periferie naar kern*. Foris, pages 81–99.

Rogers, James. 1994. *Studies in the Logic of Trees with Applications to Grammar Formalisms*. Ph.D. thesis, University of Delaware.

Shieber, Stuart M. 1985. Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, 8(3):333–344.

Gerald Penn is a research scientist in the Language Modeling Research Department at Bell Laboratories, and has published papers on a variety of topics pertaining to the interface between mathematical linguistics and computer science. Penn's address is: Bell Laboratories 2D-551, Lucent Technologies, 600 Mountain Ave., Murray Hill, NJ 07974, USA; email: gpenn@research.bell-labs.com