Book Reviews

Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing

Anne Abeillé and Owen Rambow (editors)
(Université de Paris VII and AT&T Labs–Research)

Stanford, CA: CSLI Publications
(distributed by the University of
Chicago Press), 2000, vii+478 pp;
hardbound, ISBN 1-57586-251-4, $64.95;
paperbound, ISBN 1-57586-252-2, $24.95

Reviewed by
Geoffrey K. Pullum
University of California, Santa Cruz

Take a finite set $\mathcal{T}$ of trees, and close it under the operation of *substitution*—replacing daughterless nonterminals by other trees in $\mathcal{T}$ whose root node label matches the nonterminal. The set of all terminal strings of trees in the resultant tree-set will be a context-free language (CFL), and for every CFL there will be such a finite set of trees that generates it under substitution. For an elegant formalization of CFLs in these terms, see Rogers (1999, pages 25–26).

Now assume an additional operation: besides substituting trees for daughterless nonterminals, you can also squeeze new material into the middle of a tree, substituting it for a nonterminal node that has daughters. More precisely, assume a finite set $\mathcal{A}$ of insertable trees that share a special form: each $A \in \mathcal{A}$ has a single node on its frontier, known as A's foot node, that has the same label as A's root. Squeezing some $A \in \mathcal{A}$ into a tree $T \in \mathcal{T}$ means replacing a node n in T so that n's mother becomes the mother of A's root, and everything dominated by n comes to be dominated by A's foot node.

Closing $\mathcal{T}$ under the operation of squeezing in new material from $\mathcal{A}$ in this manner (an operation called tree adjunction) yields a tree-set of which the set of all terminal strings will be a tree adjoining language (TAL), and for every TAL there will be an appropriate pair $\langle \mathcal{T}, \mathcal{A} \rangle$. Such a pair (called a tree adjunct grammar in Joshi, Levy, and Takahashi [1975]) is known today as a tree adjoining grammar (TAG).

The research program on TAGs that Aravind Joshi has led since 1975 is perhaps the most interesting and significant research program in formal language theory of the last 40 years. General linguists have clearly underrated it, though computational linguists have in general kept more closely in touch with it. The TALs are a mathematically natural class with closure and decision properties very similar to those of the CFLs, including a polynomial-time recognition problem. Several independent but equivalent characterizations of the class have been discovered: Vijay-shanker and Weir (1994) present a weak equivalence result for head grammars (Pollard 1984; Roach 1987), linear indexed grammars (Duske and Parchmann 1984; Gazdar 1988), and combinatory categorial grammars (Steedman 1986). Additionally, Vijayashanker (1988) gave a characterization in terms of embedded pushdown automata; and Rogers (1998) gives a new model-theoretic characterization in terms of linearized terminal strings of three-dimensional tree models of monadic second-order logic formulae.

Downloaded from http://direct.mit.edu/coli/article-pdf/28/1/77/1797711/0891201023173417792.pdf by guest on 09 December 2023

Moreover, the descriptive capabilities of TAGs make them as plausible a theory of syntax for natural languages as has ever emerged from formal language theory. The ways in which TAGs exceed the descriptive capacity of context-free grammars seems remarkably close to what one would want in a theory of syntax sculpted for human languages.

However, like relational grammar in the 1970s and optimality theory in the 1990s, the theory of TAGs has never had the book-length exposition it deserves: there has been no coherent and comprehensive published monograph by the original developers that gives an integrated account of the framework and convincing examples of its application to a well-known language.

Tree Adjoining Grammars, edited by Anne Abeillé and Owen Rambow, does not entirely fill that gap, though it goes some of the way. It is a refereed collection of papers that were presented in earlier versions at an international workshop on TAGs and related formalisms in Paris in 1994. The editors contribute a substantial introduction, and the 18 other chapters are grouped into three sections: (1) "Formalisms," (2) "Linguistic Analysis," and (3) "Processing."

I cannot here summarize or critique all the papers in this rich collection, but I will comment briefly on a few of the papers that I think would on their own justify the purchase of the book (especially at CSLI's attractively low price of under $25 for the paperback).

Chapter 1 is an impressive 68-page essay by Abeillé and Rambow. An expansion of this essay with fuller exemplification and more precision in formulation, perhaps coauthored with Joshi, might have made an independent monograph on TAGs that would have filled the gap referred to above. But instead this lengthy essay is just an extended introduction to an anthology. It is not without flaws. For example, its introduction of the crucial notion "lexicalized" is remarkably casual and does not really permit one to discern what the definition is. (My understanding is that in a lexicalized TAG, each of the trees in $\mathcal{T}$ must contain one and only one terminal symbol [word in the dictionary]. But in that case, the diagrams in this chapter never really give an example of a lexicalized TAG, which is odd, since the authors take this concept to be centrally important.) That said, however, the amount of work this chapter represents is substantial, and the standard of exposition is mostly high. It does a lot more than the usual summary-of-the-rest-of-the-book that is typical for an editorial introduction.

Chapter 2, by Roger Evans, Gerald Gazdar, and David Weir, has an oversubtle title that is ruined by capitalization of significant words on the chapter title page and in the running heads. The title should read as follows:

"Lexical Rules" are just lexical rules.

The thesis is that the theoretical machinery needed to state the sort of lexical rules that everybody assumes (like the one that assigns irregular plurals in *-i* to certain Latinate nouns ending in *-us*), properly understood, suffices to take over all the work of the special class of generative "Lexical Rules" that many frameworks recognize—the rules for covering phenomena like the active/passive relation between verb subcategorization frames. A full consideration of what is needed to state the former type of lexical generalization in the representation language DATR reveals that such mechanisms can also handle the (apparently) heavier stuff. Building on earlier work by the authors (Evans, Gazdar, and Weir 1995), this elegant paper shows how to state lexicalized TAGs compactly and nonredundantly using DATR, and how to express a variety of so-called Lexical Rules without positing any additional devices.

Downloaded from http://direct.mit.edu/coli/article-pdf/28/1/77/1797711/0891201023173417792.pdf by guest on 09 December 2023

Chapter 6, "Complexity of Scrambling: A New Twist to the Competence-Performance Distinction" by Aravind K. Joshi, Tilman Becker, and Owen Rambow, deserves to become known as a classic. It proposes that certain limitations on scrambling (reordering of clausal constituents) in German should not be treated as performance limits (in the way the limits on center-embedding in English usually are) because there is a way to make them follow from a grammar formalism (namely, TAGs), and this provides a better explanation than positing a performance restriction stemming from some unknown psychological or neurological basis. In other words, the paper advocates letting the syntactic theory decide: whether a theory is available that will draw a certain syntactic distinction should be a relevant factor in deciding when and whether performance limitations are to be invoked. The argument is clever, convincing, and quite surprising.

Chapter 12, "Implications of Binding for Lexicalized Grammars" by Mark Steedman, is a characteristically wide-ranging and interesting look at what binding phenomena mean for transformational grammar, generalized phrase structure grammar, head-driven phrase structure grammar, TAGs, and various types of categorial grammar (recall that Steedman's combinatory categorial grammar is weakly equivalent to TAGs).

Many other papers in the volume will repay study. Among the less technical are Robert Frank's paper speculating on children's progress toward increasing syntactic complexity during language acquisition (Chapter 3), which is moderately interesting but does not go much beyond the suggestion that comparing acquisition time and processing load for various constructions might be a good idea, and the paper by Beth Ann Hockey and Heather Mateyak on the semantic features that influence the sequencing of determiners in English (Chapter 9), which does not make much essential use of TAGs. But others are highly (even indigestibly) technical studies of various theoretical modifications of or alternatives to TAGs (e.g., Gisela Pitsch's comparison of TAGs with hyperedge replacement grammars, Chapter 7). Most of the examinations of parsing and implementation issues (Chapters 13–19) are fairly demanding.

This is a valuable book, and I am glad to have it. But it is my duty as reviewer to express a small grumble about it. This book is not a credit to the editorial profession. The bibliographies to the chapters (some 30 pages altogether) are not harmonized in style (e.g., with respect to capitalization) and are not collated at the end of the book (which makes for some wasteful duplication). The index is simply unacceptable: it will not enable scholars to find in this book the things they are looking for. And the text contains many misprints and formatting errors. I noted: "probelm" for "problem" (page 51); "explicitely" for "explicitly" (page 147); "*theXP*" for "*the XP*" (page 169); "*connectives*or" for "*connectives* or" and "*markers*can" for "*markers* can" (page 249); "*alors* ." for "*alors*." (page 253); "bindingpossibilities" for "binding possibilities." (page 283); "analysisas such" for "analysis as such" and several other such errors (page 284, bottom); "gapsare" for "gaps are" and several other such errors (page 298, bottom); "slowliness" for "slowness" (page 324); "asemantic graph ? an answer" for "a semantic graph? An answer" (page 324); "Gazdar, G. G." for "Gazdar, G." (page 471); and so on (this list is not exhaustive).

There are typos in most books, of course; accurate proofreading is an arduous job. But this book falls below what might reasonably be expected. To see an extraneous paragraph break caused when LaTeX hit a double line break in the source file (see page 77) suggests that some parts of the book were hardly even looked at in final form, let alone proofread with care, by either the editors or the publisher's staff.

If books are going to be produced via ready-to-run LaTeX files, and volume editors do not take their jobs seriously, it bodes ill for the future of books. LaTeX makes beautiful

Downloaded from http://direct.mit.edu/coli/article-pdf/28/1/77/1797711/0891201023173417792.pdf by guest on 09 December 2023

pages (Donald Knuth and Leslie Lamport did their jobs), but it can't spell, and it can't insert or delete word breaks or line breaks in the source file. The jobs of copy editors and compositors and proofreaders still have to be done (the production of this journal by The MIT Press still involves a copy editor, a proofreader, and an expert LaTeX wrangler over and above the editors). I am sorry to say that the editors of this generally interesting and useful book have let their readers down.

References

Duske, Jürgen and Rainer Parchmann. 1984. Linear indexed languages. *Theoretical Computer Science*, 32:47–60.

Evans, Roger, Gerald Gazdar, and David Weir. 1995. Encoding lexicalized tree adjoining grammars with a nonmonotonic inheritance hierarchy. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pages 77–84.

Gazdar, Gerald. 1988. Applicability of indexed grammars to natural languages. In Uwe Reyle and Christian Rohrer, editors, *Natural Language Parsing and Linguistic Theories*. D. Reidel, Dordrecht, pages 69–94.

Joshi, Aravind K., Leon S. Levy, and Masako Takahashi. 1975. Tree adjunct grammars. *Journal of Computing and System Sciences*, 19:136–163.

Pollard, Carl. 1984. *Generalized Context-Free Grammars, Head Grammars and Natural Language*. Ph.D. thesis, Stanford University.

Roach, Kelly. 1987. Formal properties of head grammars. In Alexis Manaster-Ramer, editor, *Mathematics of Language*. John Benjamins, Amsterdam, pages 293–348.

Rogers, James. 1998. A descriptive characterization of tree-adjoining languages. In *Proceedings of the 17th International Conference on Computational Linguistics (COLING'98) and the 36th Annual Meeting of the Association for Computational Linguistics (ACL'98)*, pages 117–121.

Rogers, James. 1999. The descriptive complexity of generalized local sets. In Hans-Peter Kolb and Uwe Mönnich, editors, *The Mathematics of Syntactic Structure: Trees and Their Logics*. (Studies in Generative Grammar, 44.) Mouton de Gruyter, Berlin, pages 21–40.

Steedman, Mark. 1986. Combinators and grammars. In Richard Oehrle, Emmon Bach, and Deirdre Wheeler, editors, *Categorial Grammars and Natural Language Structures*. Foris, Dordrecht, pages 417–442.

Vijayashanker, K. 1988. *A Study of Tree Adjoining Grammars*. Ph.D thesis, University of Pennsylvania.

Vijay-shanker, K. and David J. Weir. 1994. The equivalence of four extensions of context-free grammars. *Mathematical Systems Theory*, 27:511–546.

Geoffrey K. Pullum is professor of linguistics at the University of California, Santa Cruz, where his teaching ranges from a linguistics graduate course in mathematical foundations of linguistics to a computer science freshman course on the Unix operating system. He is coauthor with Rodney Huddleston of a forthcoming book entitled *The Cambridge Grammar of the English Language* (Cambridge University Press). Pullum's e-mail address is pullum@ling.ucsc.edu; URL: http://ling.ucsc.edu/~pullum.

Downloaded from http://direct.mit.edu/coli/article-pdf/28/1/77/1797711/0891201023173417792.pdf by guest on 09 December 2023