Putting Linguistics into Speech Recognition: The Regulus Grammar Compiler

Manny Rayner, Beth Ann Hockey, and Pierette Bouillon
(NASA Ames Research Center and University of Geneva)

Stanford, CA: CSLI Publications (CSLI studies in computational linguistics, edited by Ann Copestake), 2006, xiv+305 pp; hardbound, ISBN 1-57586-525-4, $65.00; paperbound, ISBN 1-57586-526-2, $25.00

Reviewed by
Brian Roark
Oregon Health & Science University

This book provides a detailed overview of Regulus, an open-source toolkit for building and compiling grammars for spoken dialog systems, which has been used for a number of applications including Clarissa, a spoken language system in use on the International Space Station. The focus is on *controlled-language user-initiative* systems, in which the user drives the dialog, but is required to use constrained language. Such a spoken language interface allows for a constrained range of commands—for example, "open the pod-bay doors"—to be issued in circumstances where, ergonomically, other interface options are infeasible. The emphasis of the approach, given the kind of application that is the focus of the work, is thus less on robustness of speech recognition and more on depth of semantic processing and quality of the dialog management. It is an interesting book, one which succeeds in motivating key problems and presenting general approaches to solving them, with enough in the way of explicit details to allow even a complete novice in spoken language processing to implement simple dialog systems.

The book is split into two parts. The first half is a very detailed tutorial on using Regulus to build and compile grammars. Regulus, although itself open-source, makes use of SICStus Prolog for the dialog processing and the Nuance Toolkit for speech recognition. Grammars in Regulus are specified with features requiring unification, but are compiled into context-free grammars for use with the Nuance speech recognition system. An abundance of implementation details and guidance for the reader (or grammar writer) is provided in this part of the book for building grammars as well as a dialog manager. The presentation includes example implementations handling such phenomena as ellipsis and corrections. In addition, details for building spoken language translation systems are presented within the same range of constrained language applications. The tutorial format is terrifically explicit, which will make this volume appropriate for undergraduate courses looking to provide students with hands-on exercises in building spoken dialog systems.

One issue with the premise of an open-source toolkit that relies upon other software (SICStus Prolog, Nuance) for key parts of the application is that one is required to obtain and use that software. In the book, the authors note that an individual license of SICStus is available for a relatively small fee, and that Nuance has a program to license their Toolkit for research purposes. Unfortunately, since the writing of the book, corporate changes at Nuance have made obtaining such a research license more challenging, and this reviewer was only able to do so after several weeks of e-mail persistence. It would be very beneficial to future readers if the authors would scout out the current state of

the process of obtaining such a license and provide a detailed map of it on some easily accessible Web site. As it stands, the barriers to building the full spoken dialog systems detailed in the book are higher than they should be.

The second half of the book goes under the hood to examine, more generally, the issues involved in making the authors' approach work. The chapter on compilation of feature grammars into context-free grammars looks in depth at alternatives to exhaustive expansion, including efficient filtering techniques and grammar transformation. There is also a chapter on developing an English feature grammar specifically for integration with a speech recognition system. A third chapter presents the authors' approach for adapting the general English feature grammar to a particular domain, which they term grammar specialization. In effect, they induce a new feature grammar by transforming—flattening, to a greater or lesser extent—an automatically produced treebank of domain-specific text and extracting new rules from the treebank, as well as estimating production probabilities from the corpus. A subsequent chapter investigates the impact of various parameterizations on this grammar specialization approach—for example, the degree of flattening—thus establishing, at least for the applications presented, some best practices. A final chapter presents a comparison of their grammar-based approach with a class-based n-gram language model.

This half of the book will be more enjoyable for readers of this journal, who are presumably interested in more general questions of computation and language than the step-by-step tutorial format of the first half of the book. The details of the approach are interesting, particularly the insights about how to build linguistically rich grammars that can be effectively compiled into high-utility context-free grammars for speech recognition.

The primary shortcoming of this presentation lies in perpetuating the false dichotomy between "grammar-based" and "data-driven" approaches to language modeling for speech recognition, which motivates the final chapter of the book. In fact, the authors' approach is both grammar-based and data-driven, given the corpus-based grammar specialization and PCFG estimation, which the authors themselves demonstrate to be indispensable. Robust grammar-based language modeling is a topic that has received a fair bit of attention over the past decade (Chelba and Jelinek 2000; Charniak 2001; Roark 2001; Wang, Stolcke, and Harper 2004, among others), and while this line of research has not focused on the use of manually built, narrow-domain feature grammars, there is enough similarity between the approach described in this book and the cited papers that the papers would seem to be better comparison points than the class-based language models that are chosen to represent robust approaches in the comparison. Beyond language modeling, methods for PCFG induction from treebanks have been a popular topic in the field over the past decade, and some understanding of the impact of flattening trees can be had in Johnson (1998), where the beneficial impact of various tree transformations for probabilistic grammars is presented. None of this work is discussed or cited, and the naive reader might be left with the impression that data-driven approaches have been demonstrated to underperform relative to knowledge-based approaches, when in fact the authors simply demonstrate that their hybrid grammar-based/data-driven approach outperforms class-based language models. Perhaps this is worth demonstrating, but the chapter couches the results within the context of a clash between paradigms, which simply does not ring true.

This one misstep, however, does not detract from the quality of the authors' system, nor from the interesting presentation of too-often-ignored aspects of spoken language systems engineering. The book and the toolkit it describes can serve a very useful pedagogical purpose by providing a foundation upon which students and researchers

can build spoken dialog applications. For those interested in constrained spoken dialog systems, there is much of interest in this book.

References

Charniak, Eugene. 2001. Immediate-head parsing for language models. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pages 116–123, Toulouse.

Chelba, Ciprian and Frederick Jelinek. 2000. Structured language modeling. *Computer Speech and Language,* 14(4):283–332.

Johnson, Mark. 1998. PCFG models of linguistic tree representations. *Computational Linguistics,* 24(4):617–636.

Roark, Brian. 2001. Probabilistic top-down parsing and language modeling. *Computational Linguistics,* 27(2):249–276.

Wang, Wen,Andreas Stolcke, and Mary P. Harper. 2004. The use of a linguistically motivated language model in conversational speech recognition. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages 261–264, Montreal.

Brian Roark is an Assistant Professor in the Department of Computer Science & Electrical Engineering and the Center for Spoken Language Understanding at the OGI School of Science & Engineering of Oregon Health & Science University (OHSU). Before joining OHSU in 2004, he spent 3 years at AT&T Labs–Research. His research focuses on syntactic parsing of speech and text, and language modeling for speech recognition. With Richard Sproat, he is the author of *Computational Approaches to Syntax and Morphology* (Oxford University Press, forthcoming).