Book Reviews

New Developments in Parsing Technology

Harry Bunt, John Carroll, and
Giorgio Satta (eds.)
Tilburg University, University of Sussex,
and University of Padua

Dordrecht: Kluwer Academic Publishers, 2004,
xi+403 pp; paperback,
ISBN 1-4020-2294-8, $64.95

Reviewed by
Stefan Riezler
Palo Alto Research Center

New Developments in Parsing Technology is a collection of papers based on contributions to the International Workshop on Parsing Technology in the years 2000 and 2001. The publication format of a collection might raise the following questions: Is the whole of the collection more than the sum of its previously published parts by virtue of an inspired selection of the most seminal papers in the area? Or does the collection go beyond a mere reprint of revised versions of workshop papers by including insightful overview articles or other previously unpublished material? In case of *New Developments in Parsing Technology* the answers to these questions are *yes* concerning added value by the inclusion of a previously unpublished invited talk by Michael Collins, and *no* concerning exceeding the sum of its previously published parts.

Table 1 lists the table of contents of the book. The book starts out with an introductory chapter written by the editors. In this article, the editors motivate an interest in parsing technology by listing 12 application areas that make crucial use of parsing techniques. Given the limited pool of candidate papers from two workshops, unfortunately

Table 1
List of chapters in *New Developments in Parsing Technology*.

1 Developments in parsing technology: From theory to application (*Harry Bunt, John Carroll, and Giorgio Satta*)

2 Parameter estimation for statistical parsing models: Theory and practice of distribution-free methods (*Michael Collins*)

3 High precision extraction of grammatical relations (*John Carroll and Ted Briscoe*)

4 Automated extraction of TAGs from the Penn Treebank (*John Chen and K. Vijay Shanker*)

5 Computing the most probable parse for a discontinuous phrase-structure grammar (*Oliver Plaehn*)

6 A neural network parser that handles sparse data (*James Henderson*)

7 An efficient LR parser generator for tree-adjoining grammars (*Carlos A. Prolo*)

8 Relating tabular parsing algorithms for LIG and TAG (*Migual A. Alonso, Éric de la Clergerie, Víctor J. Díaz, and Manual Vilares*)

9 Improved left-corner parsing for large context-free grammars (*Robert C. Moore*)

10 On two classes of feature paths in large-scale unification grammars (*Liviu Ciortuz*)

11 A context-free superset approximation of unification-based grammars (*Bernd Kiefer and Hans-Ulrich Krieger*)

12 A recognizer for minimalist languages (*Henk Harkema*)

13 Range concatenation grammars (*Pierre Boullier*)

14 Grammar induction by MDL-based distributional classification (*Yikun Guo, Fuliang Weng, and Lide Wu*)

15 Optimal ambiguity packing in context-free parsers with interleaved unification (*Alon Lavie and Carolyn Penstein Rosé*)

16 Robust data oriented spoken language understanding (*Khalil Sima'an*)

17 SOUP: A parser for real-world spontaneous speech (*Marsal Gavaldà*)

18 Parsing and hypergraphs (*Dan Klein and Christopher D. Manning*)

19 Measure for measure: Towards increased component comparability and exchange (*Stephan Oepen and Ulrich Callmeier*)

the only application area that is addressed in the collection concerns the processing of spoken language (chapters 15–17). Topical clusters for the remaining papers can be induced loosely from the grammar frameworks they refer to. The large majority of papers can be tagged as extensions of context-free grammars, with some work on HPSG and TAG interspersed. Again, the limited candidate pool prevented a broader spectrum of grammar frameworks, thus excluding work in the areas of LFG, CCG, and other linguistically deep parsing frameworks. Statistical parsing techniques, which arguably are the focus of most recent developments in parsing technology, are addressed only in chapters 2–4. Again, this misrepresents the state-of-the-art in parsing research. Without diving into short summaries of the papers contained in the collection—the editors do an excellent job on this task—it seems unfortunate that it was not possible to extend the paper selection beyond the pool of contributions to two workshops, for example, by accompanying the selected papers by related work, or follow-up papers that apply or criticize the presented work.

However, genuinely "new developments" in the area of statistical parsing are presented in the invited paper by Michael Collins (chapter 2). This chapter truly lives up to promise of the book title and it also serves many purposes: Firstly, it is an excellent, self-contained introduction to large-margin methods for machine learning. Collins gently leads the reader from the well-known territory of statistical parameter estimation for PCFGs to generalization theory and algorithms for large-margin classifiers. Generalization theory asks the question of how well a learner classifies unseen data given only a limited amount of training data, instead of referring to the law of large numbers for guarantees in parameter estimation. In other words, it asks how much training data is needed for an estimator to converge to a point where it has minimal error on unseen data, i.e., where it is probably approximately correct. Collins manages to provide the novice reader with an intuitive explanation of the most important convergence

bounds in the framework of probably approximately correct (PAC) learning theory. Furthermore, the paper serves the advanced reader by showing how large-margin classifiers can be applied to multi-class classification problems such as parsing, together with previously unpublished proofs for bounds on the generalization error of large-margin–based parse selection methods. The algorithms presented in the paper include support vector machines, boosting, and the voted perceptron, all of which have been shown to provide significant improvements in parse selection in previously published experiments. The paper concludes with a discussion of the relation of large-margin methods to Markov random fields (MRFs). Collins points out that these methods are closely related, e.g., to boosting techniques, and refers to publications that show the similarity of both methods in a maximum-likelihood framework (Lebanon and Lafferty, 2001; Collins, Schapire, and Singer, 2002). The missing link of a PAC interpretation of regularized MRFs was presented in the same year as this collection in a paper by Andrew Ng. Ng (2004) presents impressive generalization bounds for l_1-regularized logistic regression, showing that the sample complexity of such learners grows only logarithmically in the number of irrelevant features instead of linearly as for learners without feature selection techniques. This result fills a gap in the literature and shows that both "parametric" and "distribution-free" methods can be justified in a framework of maximum-likelihood estimation as well as in a PAC learning setting. Unfortunately, the three-year gap between the workshops and the publication of the collection prevented a reference to this result in Collins's article. On the same note it should be mentioned that polynomial bounds for multi-class classifiers outperforming Collins's Theorems 8 and 9 that require an exponential number of constraints were already presented in 2003 by Taskar, Guestrin, and Koller (2003).

Overall, the recommendation on whether to read *New Developments in Parsing Technology* must be as follows: Collins's invited contribution is so outstanding that it alone

makes it worthwhile to get hold of a copy of the book. Each of the selected workshop papers is a worthwhile read in itself, however, beyond the fact that the papers appeared in two consecutive years of IWPT, there is no added value in having them brought together in this particular collection. Considering that most workshop papers are available online as well as in the workshop proceedings, it is really Collins's chapter alone that justifies a purchase of the book.

References

Collins, Michael, Robert E. Schapire, and Yoram Singer. 2002. Logistic regression, AdaBoost and Bregman distances. *Machine Learning*, 48(1–3):253–285.

Lebanon, Guy and John Lafferty. 2001. Boosting and maximum likelihood for exponential models. In *Advances in Neural Information Processing 14 (NIPS'01)*, Vancouver.

Ng, Andrew Y. 2004. Feature selection, l_1 vs. l_2 regularization, and rotational invariance. In *Proceedings of the 21st International Conference on Machine Learning (ICLM'04)*, Banff, Canada.

Taskar, Ben, Carlos Guestrin, and Daphne Koller. 2003. Max-margin Markov networks. In *Advances in Neural Information Processing Systems 17 (NIPS'03)*, Vancouver, Canada.

Stefan Riezler is a researcher at the Palo Alto Research Center since 2001. He received his PhD in Computational Linguistics in 1998 from the University of Tübingen, Germany. His research focus is on machine learning techniques for statistical parsing, and in applications of deep statistical parsing to various tasks, including lexicon induction, sentence condensation, and recently, machine translation. Riezler's address is: Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304; e-mail: riezler@parc.com