Learnability in Optimality Theory

Bruce Tesar and Paul Smolensky
(Rutgers University and The Johns Hopkins University)

Cambridge, MA: The MIT Press, 2000,
vii+140 pp; hardbound, ISBN
0-262-20126-7, $25.00

Reviewed by
Walter Daelemans
University of Antwerp

There are several ways in which the algorithmic acquisition of language knowledge and behavior can be studied. One important area of research is the computational modeling of human language acquisition using statistical, machine learning, or neural network methods. See Broeder and Murre (2000) for a recent collection of this type of research. And there is of course the use of statistical and machine learning methods in computational linguistics and language technology (Manning and Schütze 1999). The theoretical study of the learnability of formal grammars in close relationship with linguistic explanation in generative grammar ("the logical problem of language acquisition") is yet another relevant area of research.

The new book by Bruce Tesar and Paul Smolensky (T&S) is an example of the latter approach, but supported by computational modeling. In the book, an approach to learning in optimality theory (OT) is proposed. OT is an alternative to the principles and parameters (P&P) approach to Universal Grammar. OT claims that all languages have in common a set of constraints on well-formedness, and differ only in which constraints have priority in case of conflict. Priorities for each language are characterized in a language-specific ranking (dominance hierarchy). This hierarchy assigns harmony values to structural descriptions of language input (depending on which constraints are broken), and the analyses with maximal harmony are the well-formed ones. In OT, the problem of language learning is therefore transformed from a parameter-setting problem into a constraint-ranking problem. In P&P approaches to learning, either comprehensive structural detail has to be encoded in the learning principles (as in cue learning), or the search is completely uninformed (as in triggering learning). The central claim of T&S is that OT provides sufficient structure to allow efficient and grammatically informed learning, and therefore offers an optimal trade-off.

Chapter 1 introduces the problem of language learning and sketches its solution. The problem is that a learner cannot parse the language input until a grammar has been learned, but a grammar cannot be learned until the input can be parsed. The solution proposed for this unsupervised learning problem will be familiar to researchers in statistical natural language processing: a variant of expectation-maximization in which the model (grammar) and the input-output pairing (parse) are iteratively optimized. In Chapter 2, OT is explained as a series of general principles. These principles are illustrated by means of a phonological example (syllable parsing) and a syntactic example (distribution of clausal subjects). This is an excellent chapter for those wanting a concise and clear introduction to OT. Chapter 3 introduces constraint demotion (CD) as the grammar-learning part in the expectation-maximization set-up explained earlier. The basic mechanism is simple and elegant: given an initial constraint ordering, and pairs of well-formed structural descriptions (winners) and competing not-well-formed structural descriptions (losers), demote the constraints violated by a winner down in the hierarchy, making each winner more harmonic than its competing losers. But where do the losers come from? OT presupposes that Universal Grammar provides access to a function *Gen* that generates all possible candidate structural descriptions for a language input. This provides implicit negative evidence for the learner. The selection of specific losers among a potentially infinite number of possible losers is again based on a familiar notion for statistical NLP practitioners: error-driven learning. Selection of learning material is guided by failure of the parser. In Chapter 7, T&S formally show that the CD algorithm is correct, works both from an initially unordered constraint hierarchy and from arbitrary initial hierarchies, and that data complexity is favorable. CD requires $n(n-1)/2$ informative data pairs (where n is the number of constraints) for the initially unordered hierarchy case, and $n(n-1)$ for the arbitrary initial hierarchy case. Chapter 4 focuses on robust interpretive parsing (RIP, the parsing part in the expectation-maximization set-up) and its combination with constraint demotion (RIP/CD learning for short) in the context of stress grammar learning. Dynamic programming is used in RIP to guide the search for structural descriptions for inputs (explained in more detail in Chapter 8). In Chapter 4 we also find some simulation results, with success scores (in number of languages for which the correct stress system is learned) ranging from about 60% to 97% depending on the initial constraint hierarchy chosen. From the perspective of computational linguistics research methodology, this is a disappointing section. The short description leaves out essential details about data selection and preprocessing, implementation, and experimental set-up, and lacks a comparison with alternative approaches. We do learn more about cases where the RIP/CD algorithm can fail. Chapters 5 and 6 deal with a number of interesting issues and extensions (e.g., the subset principle, learning of underlying forms, and lexicon learning), but they seem to me more speculative than the rest of the book.

This book is clearly groundbreaking for researchers in OT, formal learning theory, and generative grammar. But it is a remarkable piece of work also for more agnostic (computational) linguists interested in algorithms for language learning. I found it especially exciting to see how OT allows well-known statistical NLP tools like dynamic programming, expectation-maximization, and error-driven learning to be integrated and put to use in an elegant, efficient, and original solution to the language learning problem. The book certainly succeeds in conveying the importance of the T&S approach to learnability in OT and in general, but the approach needs more extensive simulation results to be completely convincing.

References

Broeder, Peter and Jaap Murre. 2000. *Cognitive Models of Language Acquisition*. Cambridge University Press.

Manning, Christopher D. and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, MA.

Walter Daelemans is a professor of computational linguistics and artificial intelligence at the University of Antwerp, and currently president of SIGNLL, ACL's special interest group on natural language learning. Daelemans's address is CNTS Language Technology Group, University of Antwerp, Universiteitsplein 1, Building A, B-2610 Antwerpen, Belgium; e-mail: daelem@uia.ac.be.