Book Reviews

The Case for Lexicase: An Outline of Lexicase Grammatical Theory

Stanley Starosta
(University of Hawaii at Manoa, Honolulu, HA)

London, England: Pinter, 1988, xii + 273 pp. (Open linguistics series)
ISBN 0-86187-639-3, £29.50 (hb)

Reviewed by
Norman Fraser
University College London

Dependency theory has been enjoying a minor renaissance recently (except in Eastern Europe, where it has always been the dominant tradition). Some of the main ideas of dependency grammar are finding their way into constituent-based models. The widespread use of the notion **head**, the recognition of grammatical relations as basic, and the trend towards maximizing the role of the lexicon and minimizing the number of rules in the grammar all exemplify the drift towards dependency. An increasing number of computational linguists are designing dependency-based systems for parsing (e.g., Hellwig 1986; Covington 1988), semantic interpretation (Danieli et al. 1987), and machine translation (Schubert 1987). However, very few formally explicit dependency-based general linguistic theories have been offered to compete with the more familiar constituency-based favorites. *The Case for Lexicase* by Stanley Starosta presents just such a theory. "Lexicase" is described as a "panlexicalist monostratal dependency variety of generative localistic case grammar".

Chapter 1 introduces the theory. Dependency structure is usefully presented in terms of a highly constrained version of X-bar theory in which all terminal nodes are words and every construction has at least one immediate lexical head (i.e., only single-bar phrases are possible). Lexical items are subcategorized only by their (dependent) sisters.

Chapter 2 considers the lexicon. The theory is "panlexicalist" in that grammatical rules can be viewed as generalizations about the lexicon. All information is stored in lexical entries that are three-part signs consisting of entries for sound, meaning, and distribution. The distribution entry is a set of binary or implicational features, consisting of an atomic word class feature, plus subcategorization, case, and role features where needed. Features are marked as implicational where they are expected but not absolutely necessary. This is claimed to facilitate the interpretation of metaphor but not enough detail is provided to assess the claim.

Chapter 3 is concerned with formalization, in particular with the properties of lexicase rules. Redundancy, subcategorization, and morphological rules supply predictable and default values to lexical entries. Derivation rules create new lexical entries.

Chapters 4 and 5 describe the lexicase case system, which is an outgrowth of Fillmore case grammar. A lexicase grammar distinguishes among case relations (AGENT, PATIENT, LOCUS, CORRESPONDENT, and MEANS), macroroles (Actor and Undergoer), and case forms (Nominative, Accusative, Ergative, etc). A lexical entry may include a feature in respect of each of the above together with any number of localistic features (source, goal, surface, etc).

Chapter 6 reviews the lexicase analyses of a number of common constructions. Most interesting is the analysis of coordinate structures (which have always been problematic in dependency grammar) as exocentric, having as many heads as conjuncts. The union of the head feature sets produces a virtual matrix for the whole coordinate structure. And finally, Chapter 7 presents an agenda for future research.

Starosta's presentation is clear and carefully argued but this serves to highlight the inexplicable omission of any serious discussion of "movement" phenomena. Assurances that lexicase can cope with movement and references to working papers are not satisfactory. On the whole, however, this is a book that deserves to be taken seriously in the debate between dependency and constituency, not least because of the weight of field work that supports it. Lexicase has been developing since the early 1970s and has been used in studies of almost 50 different (mostly non-Indo-European) languages. It has also formed the theoretical basis for some computer systems (e.g. Starosta and Nomura 1986). Although most of the book's examples are drawn from English, it has much of value to offer in correcting the English bias built into so many leading linguistic theories.

REFERENCES

Covington, Michael A. 1988 *Parsing Variable Word-Order Languages with Unification-based Dependency Grammar.* ACMC Research Report 01–0022, Advanced Computational Methods Center, University of Georgia, Athens, GA.

Danieli, Morena; Ferrara, Franco; Gemello, Roberto; Rullent, Claudio 1987 Integrating Semantics and Flexible Syntax by Exploiting Isomorphism Between Grammatical and Semantical Relations. In *Proceedings of the 3rd Conference of the European Chapter of the Association for Computational Linguistics*; 278–283.

Hellwig, Peter 1986 Dependency Unification Grammar. In *Proceedings of the 11th International Conference on Computational Linguistics* (COLING-86); 195–198.

Schubert, Klaus 1987 *Metataxis: Contrastive Dependency Syntax for Machine Translation*. Foris, Dordrecht, Holland.

Starosta, Stanley and Nomura, Hirosato 1986 Lexicase Parsing: A Lexicon-Driven Approach to Syntactic Analysis. In *Proceedings of the 11th International Conference on Computational Linguistics* (COLING-86); 127–132.

Norman Fraser is a research associate in linguistics at University College London. He is currently developing parsing algorithms for dependency grammars. Fraser's address is: Department of Phonetics and Linguistics, University College London, Gower Street, London, WC1E 6BT, U.K.

COMPUTERS AND LANGUAGES: THEORY AND PRACTICE

Anton Nijholt
(Free University, Brussels, Belgium)

Amsterdam: North-Holland, 1988, xiii + 482 pp.
(Studies in Computer Science and Artificial Intelligence 4)
ISBN 0-444-70463-9, $89.50, Dfl 170.00 (hb)

Reviewed by
Richard S. Rosenberg
University of British Columbia

This book is the fourth in the series *Studies in Computer Science and Artificial Intelligence* and as such does depend somewhat on the reader having a background in computer science. However, there is so much stuff in this eclectic book that almost anyone lacking specialized knowledge but with interests in artificial intelligence, history, linguistics, computer science, or social issues will find something to savor.

Computers and Languages consists of 13 chapters divided into five parts, namely: "Introduction" (history of computers, introduction to computability and formal language theory, and an introduction to intelligent applications and AI); "Military Background" (impact of computers on military needs and space and military applications of AI); "Viewpoints on Language" (introduction to generative grammar and associated issues such as acquisition, competence, performance, psychological validity and parsing, BNF programming and computer languages, and formal languages and parsing methods); "From Language to Intelligence" (a survey of natural language understanding systems from BASEBALL on, including interfaces and expert system applications, a variety of approaches such as ATNs, case grammar, Schank's conceptual dependency, Winograd's SHRDLU, semantic networks and frames, and natural language applications: interfaces, machine translation, and military applications including speech processing); and "The Military-Industrial-Academic Complex (University Research and the Military)".

The author notes in the preface (p. *x*) that the book is not intended to be a textbook although he has used parts of it in courses on computational linguistics, computers and society, and formal approaches to languages. The weakest parts of the book are in linguistics-related areas. Although the book was published in 1988, transformational theory, as represented by Chomsky's standard theory circa 1970, is described but not much beyond. The extended standard theory is mentioned, two sentences cover government binding, and there is one sentence each for lexical-functional grammar and generalized phrase structure grammar. No mention is made of the very important recent work in logic grammars, especially unification grammars. When programming languages for natural language are discussed (pp. 314–316), Lisp is briefly introduced (a few sentences), followed by a description of Planner (two pages), a language only of interest to AI archaeologists. Nothing is said of Prolog and its growing importance.

Because so many diverse topics are covered, an accusation of superficiality cannot be entirely avoided. Nevertheless, there are redeeming features in this book. It does provide a useful introduction to the diverse aspects of natural language understanding, including both formal and applications-oriented perspectives. It is rich in history, in the people and places involved in the major contributions. However, what is unique and most admirable about this book is the author's concern about the role of government, especially the military, in academic research in AI, especially natural language understanding. One cautionary remark should be made, however, that the entire discussion, except for two paragraphs, is framed in the context of the U.S. military enterprise as if in no other countries do the military establishments influence the directions of research in computer science. It may be the case that this process is most accelerated in the U.S., but surely Western Europe, Japan, and the Soviet Union cannot be far behind.

Nijholt has performed a valuable service in reminding researchers how intimately they have become involved in military research. Witness the following quotation from *Jane's Defence Weekly*, 17 May 1986:

> The market has become so vast that there is plenty of room for competent companies now that AI is well on the way from academia to the battlefield. (p. 145)

The final chapter of the book, "University Research and the Military", presents a concise description of "collaboration" between academia and the military, including such areas as the cold war, the Vietnam War, and those currently favored ventures, Star Wars and the Strategic Computer Initiative. With respect to professional responsibility and war research, Nijholt quotes relevant sections of the ACM's Code of Professional Conduct, perhaps in hopes of reforming the recalcitrant mercenary researcher.

This book is a concrete example of the word "eclectic". Although generally well laid out, it does have a major drawback that seriously interferes with ease of use: There is no subject index, though there is a name index. In addition, references appear only at the end of each chapter, thus requiring the use of the name