

Computational Linguistics: Models, Resources, Applications

Igor A. Bolshakov and Alexander Gelbukh (editors)

(Instituto Politécnico Nacional)

Mexico City: Centro de Investigación en Computación, Instituto Politécnico Nacional, 2004, 186 pp; ISBN 970-36-0147-2, free download from www.gelbukh.com/clbook

Reviewed by

Anna Feldman

The Ohio State University

The main objective of Bolshakov and Gelbukh's new textbook in computational linguistics is to provide students in computer science with a foundation in the fundamentals of general linguistics necessary to develop applied software systems, and to enable students to make informed choices of proper models and data structures. Although Spanish-language applications are emphasized, the textbook contains examples from English, French, Portuguese, and Russian as well. This text is freely available in an electronic format. Supplementary materials, links to the relevant resources, and errata can be found on-line as well.

The authors approach linguistic problems using two theoretical frameworks: meaning–text theory (MTT) (Mel'čuk 1974), and head-driven phrase structure grammar (HPSG) (Pollard and Sag 1994). These choices reflect primarily practical considerations. On the one hand, MTT can be used to describe any language but is particularly well-suited for the description of free-word-order languages. On the other hand, HPSG is arguably the most advanced, widely used, and user-friendly formalism in natural-language description and processing.

The structure of the book is as follows. Chapter I gives an introduction to different areas of linguistics and provides an overview of the state of the art in Spanish natural language processing (NLP).

Chapter II is a brief survey of the history of linguistics from Ferdinand de Saussure to Leonard Bloomfield to Noam Chomsky, discussing context-free grammars, transformational grammars, and more up-to-date approaches to grammar. Covering concepts such as valence, constraints, and unification formalism, the authors introduce Fillmore's work (Fillmore 1968), generalized phrase structure grammar (Gazdar et al. 1985), and HPSG. The last several sections are devoted to MTT. MTT has influenced the design of new grammar formalisms such as Dependency Tree Grammars, and some aspects have been adopted into other theories (e.g., integrating lexical functions into Pustejovsky's Generative Lexicon; Wanner 1997).

Chapter III provides a sketch of computational linguistics (CL) applications. The included discussion serves to illustrate that all of the relevant applications require sophisticated linguistic knowledge. The general lesson of this chapter can be summarized thus: "Most language processing tasks can be considered as special cases of the general task of language understanding, one of the ultimate goals of CL and AI."

Chapter IV describes the MTT theory in more detail and makes interesting comparisons between MTT, HPSG, and Chomskian frameworks. Chapter V introduces the problem of language modeling in CL. The last section of the book includes various exercises, review questions, and multiple-choice problems.

This book is a nice addition to the short list of CL textbooks. It is especially helpful for students in Spanish CL, given its numerous Spanish-related examples and

applications. The current volume complements other textbooks available now by offering an overview of MTT, providing a survey of the current state of Spanish NLP research, introducing the field of CL from the linguistic point of view (explaining what specific linguistic knowledge is involved in each “CL product” rather than introducing the techniques used in CL/NLP to create that product), providing plenty of examples and exercises, and making this book freely available on-line.

For the future editions, I would suggest the inclusion of glosses of the Spanish examples for the benefit of potential readers who are not Spanish-speakers. Additionally, exercises, comprehension questions, relevant literature, scientific associations, on-line tools, and language resources should be listed at the end of each chapter instead of at the end of the book. This would facilitate the practice, and hence acquisition, of the relevant skill set, and allow readers to more easily test their comprehension of each chapter.

This text offers an introduction to the field of linguistics for students who have already some background in computer science but lack special training in linguistics. However, those using this textbook in the classroom should consider providing supplementary materials that discuss the fundamental algorithms and techniques for speech and language processing as well as certain linguistic areas, such as phonetics and pragmatics. These topics receive relatively little attention here.

References

- Fillmore, Charles. 1968. The case for case. In E. Bach and R. T. Harms, editors, *Universals in Linguistic Theory*. Holt, Rinehart & Winston, New York, pages 1–88.
- Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Basil Blackwell, Oxford.
- Mel'čuk, Igor. 1974. *Experience in Theories of Meaning ⇔ Text Linguistic Models*. Nauka, Moscow.
- Pollard, Carl and Ivan A. Sag. 1994. *Head-driven Phrase Structure Grammar*. CSLI Publications, Stanford, CA.
- Wanner, Leo. 1997. *Recent Trends in Meaning–Text Theory*. Benjamins, Amsterdam.

Anna Feldman is a doctoral student in Linguistics at the Ohio State University. Her research interest is in cross-language knowledge induction. Feldman's address is 221 Oxley Hall, 1712 Neil Avenue, Columbus, OH, 43210-1298; e-mail: afeldman@ling.osu.edu.