sibility to select only a subset of the universal inventory of categories. Concerning the categories of topic and comment, Öhl claims that these notions are not primitives but should be derived from more general principles of PERSPECTIVATION. The article contains a careful overview of the relevant research and formulates justified criticism of approaches presented in the typological and/or generative framework.

In summary, this volume addresses a number of theoretically relevant questions about the architecture of grammar, has a clear methodological orientation, and provides new empirical data. The articles included represent solid work by respected scholars with several excellent articles. No doubt, this well-edited volume is an important contribution to research on various aspects of information structure.

REFERENCES

BÜRING, DANIEL. 2003. On D-trees, beans and B-accents. *Linguistics and Philosophy* 26.511–45.
CINQUE, GUGLIELMO. 1998. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.
É. KISS, KATALIN. 1998. Identificational vs. information focus. *Language* 74.2.245–73.
LI, CHARLES N., and SANDRA A. THOMPSON. 1976. Subject and topic: A new typology of language. *Subject and topic*, ed. by Charles N. Li, 457–89. New York: Academic Press.
VALLDUVÍ, ENRIC, and ELISABET ENGDAHL. 1996. The linguistic realization of information packaging. *Linguistics* 34.459–519.

Lund University
Center for Languages and Literature
S-223 62 Lund, Helgonabacken 12
Sweden
[valeria.molnar@tyska.lu.se]

The handbook of computational linguistics and natural language processing. Ed. by ALEXANDER CLARK, CHRIS FOX, and SHALOM LAPPIN. (Blackwell handbooks in linguistics.) Malden, MA: Wiley-Blackwell, 2010. Pp. xxiii, 775. ISBN 9781405155816. $199.95 (Hb).

Reviewed by GRAEME HIRST, *University of Toronto*

The research field that brings together computers and language cannot decide whether it is a branch of linguistics. Those within the field who think it is, or should be, a branch think of it as COMPUTATIONAL LINGUISTICS (CL); those who eschew linguistics think of it as NATURAL LANGUAGE PROCESSING (NLP). The dichotomy often reflects individual research goals. Those who pursue computational models of language in order to better understand the nature of language itself are obviously on the linguistic side. Those whose goal is to develop computer systems that carry out some useful task with linguistic input or output tend to be on the other side. The processing in natural language processing is based largely on statistical and machine-learning methods, so in this view the fact that the data are natural language and that the processing might use some information about syntax and semantics along the way does not make NLP a branch of linguistics any more than the use of arithmetic makes bookkeeping a branch of mathematics.

The editors of this new handbook, Clark, Fox, and Lappin (CFL), fall on the linguistic side of the divide. Their own research backgrounds are in linguistics and computational modeling of language. And their handbook is a member of the large and well-known series 'Blackwell handbooks in linguistics'. Nonetheless, they work both sides of the street, titling it *The handbook of computational linguistics and natural language processing* (hereafter *HCLNLP*). This contrasts with the one-sided titles of the recently published *Handbook of natural language processing* (*HNLP*) (Indurkhya & Damerau 2010), which is firmly in the no-stinkin'-linguists-here tradition,

and the earlier *Oxford handbook of computational linguistics* (*OHCL*) (Mitkov 2003), which despite its title has strong coverage of NLP as well as CL.[1] But title notwithstanding, it is clear that CFL's sympathies and interests are on the CL side. Their handbook has a strong emphasis on theory and methods for computational models of language, and far less on practical applications or NLP. Even the five-chapter section entitled 'Applications' is not so much about actual applications as about methods and techniques that underlie applications. A reader who wants to learn about NLP applications such as sentiment analysis or biomedical text mining will need to turn instead to the *HNLP*, which has a chapter on each of them; and a reader interested in text summarization or computer-assisted language learning must turn to the *OHCL*. By contrast, the more theoretically or linguistically oriented reader who wants to understand computational models of language learning and grammar induction will find them only in the *HCLNLP*, well explained in a chapter by ALEXANDER CLARK and SHALOM LAPPIN. These three handbooks are thus complementary; despite the implication of comprehensiveness in their titles and their overlap in some core topics, the editorial choices in each have resulted in much less redundancy between the books than might have been expected, and the *HCLNLP* offers the most detailed coverage of the theoretical and linguistic side.

The chapter authors of *HCLNLP* are well-chosen experts on their topics, including MARTHA PALMER and NIANWEN XUE on the linguistic annotation of electronic text, MATTHEW W. CROCKER on computational psycholinguistics, RALPH GRISHMAN on information extraction from text, and EHUD REITER on natural language generation. All told, the volume contains twenty-two chapters in 741 pages (including references), an average of about thirty-three pages each—similar to *HNLP*'s twenty-six chapters in 666 (more densely packed) pages and contrasting with *OHCL*'s thirty-eight rather shorter chapters in 716 pages (similar in size to those of *HCLNLP*). Many chapters of *HCLNLP* give longer and more detailed treatments than their counterparts in the other handbooks. Some topics, especially some machine-learning methods, get a very deep and theoretical treatment; entire chapters are devoted to maximum-entropy methods (ROBERT MALOUF), decision trees (HELMUT SCHMID), and memory-based learning (WALTER DAELEMANS and ANTAL VAN DEN BOSCH), all of which rate at most a few pages in the other handbooks.

But while some topics get a long, in-depth treatment, many others get short shrift, and the inclusion of essential topics that do not warrant—or, at least, do not get—long chapters of their own is a bit hit or miss. For example, the important CL topic of part-of-speech tagging has a fourteen-page chapter in *OHCL* and a thirty-two-page chapter in *HNLP*, but in *HCLNLP* it has only two pages in the context of decision-tree methods and a brief mention in the context of maximum-entropy methods; the topic as a whole gets no coherent treatment at all. Similarly, word-sense disambiguation, which has chapter-length treatment in both *OHCL* and *HNLP*, is mentioned only in the context of a two-and-a-half-page case study in the chapter on methods for evaluating NLP systems (PHILIP RESNIK and JIMMY LIN), and nothing is said about actual methods for disambiguation. Perhaps most surprisingly, in view of the extensive and deep coverage given to some other machine-learning methods, there is no treatment at all of support-vector machines, which is surely the single most important method for text classification in NLP; indeed, there is no unified treatment of the fundamental task of text classification that underlies many contemporary applications of NLP, and the term has few index entries even though the concept runs through, in particular, the chapters on memory-based learning and decision trees.

'Language' in the *HCLNLP* almost invariably means written language. In a reflection of the state of CL and NLP themselves, spoken language is limited to chapters on speech recognition (STEVE RENALS and THOMAS HAIN) and computational models of dialogue (JONATHAN GINZBURG and RAQUEL FERNÁNDEZ).

---

[1] Another recent publication with a similar title is *Handbook of natural language processing and machine translation* (Olive, Christianson, & McCary 2011). Despite its title, however, this book is not actually a reference work like those discussed in this review; rather, it is 'a snapshot of the results of the first three years of groundbreaking progress' (Olive et al. 2011:xiv) in the large US government-funded Global Autonomous Language Exploitation program (commonly known as GALE), the goal of which 'is to empower the warfighter through the use of language technologies' (p. 845). Why a progress report on a research program would be labeled as a handbook when it obviously is not one is not explained.

The chapters vary widely in their style and format, but most are clear and well written; Resnik and Lin's chapter in particular stands out. Some chapters assume considerable mathematical sophistication, and a familiarity with probability theory is essential for many. While most chapters are general introductions to or surveys of their topic, ANDY WAY's chapter on machine translation is, in addition, rather too much an advertisement for the work of his own research group. A few chapters include sections with suggestions for further reading or other relevant resources, as one might expect in any handbook, but most do not. And some chapters, but not all, try to tie their topic into the larger context by making cross-references to other chapters where appropriate.

The handbook does not contain any material other than an introduction, the chapters themselves, a consolidated list of references not keyed to the chapters, and indices. That is, there is no glossary (*OHCL* includes a useful thirty-four-page glossary); no lists of relevant organizations, conferences, journals, and websites; and no guide to standard software and corpora except as provided in some of the individual chapters.

The production standards of the handbook are generally high, with only a few lapses. In SHULY WINTNER's chapter on formal language theory, italic and slant fonts appear in free variation, although the glyphs for some characters are quite different in the two fonts; and throughout the book, URLs appear randomly with or without a leading 'http://', even when in close proximity (e.g. p. 598). The table of contents lists only chapter names, not subsection headings, preventing the reader from getting a quick overview of a chapter. Way's chapter seems hastily written or concatenated from preexisting documents, so that, for example, some points made on page 531 (second paragraph of section 1) are repeated on page 553 (first paragraph of section 3.3) as if they had not been mentioned before; this should have been noticed in editing.

This volume is a valuable addition to contemporary reference works on CL and NLP—both handbooks and large, advanced textbooks such as that of Jurafsky and Martin (2009). Because of its greater emphasis on theoretical linguistic aspects of computational linguistics, its coverage complements that of other recent works.

REFERENCES

INDURKHYA, NITIN, and FRED J. DAMERAU (eds.) 2010. *Handbook of natural language processing*. 2nd edn. Boca Raton, FL: CRC Press.

JURAFSKY, DANIEL, and JAMES H. MARTIN. 2009. *Speech and language processing*. 2nd edn. Pearson Prentice-Hall.

MITKOV, RUSLAN (ed.) 2003. *The Oxford handbook of computational linguistics*. Oxford: Oxford University Press.

OLIVE, JOSEPH; CAITLIN CHRISTIANSON; and JOHN MCCARY (eds.) 2011. *Handbook of natural language processing and machine translation*. New York: Springer.

Department of Computer Science
10 King's College Road, room 3302
University of Toronto
Toronto, Ontario, Canada M5S 3G4
[gh@cs.toronto.edu]

Basic linguistic theory. By R. M. W. DIXON. Oxford: Oxford University Press, 2010. Vol. 1: Methodology. Pp. xvi, 381. ISBN 9780199571062. $55. Vol. 2: Grammatical topics. Pp. xvii, 489. ISBN 9780199571086. $49.95.

Reviewed by CAROL GENETTI, *University of California, Santa Barbara*

Volumes 1 and 2 of *Basic linguistic theory* are the first of a three-volume set scheduled to be completed in 2011. The front matter describes the work as a 'new and fundamental characterization of the nature of human language and a comprehensive guide to their description and analysis'. This is an accurate portrayal. These books are monumental and destined to become classics,