Briefly Noted

An Introduction to Language Processing with Perl and Prolog

Pierre M. Nugues
(Lund University)

Springer (Cognitive technologies series, edited by Dov Gabbay and Jörg Siekmann), 2006, xx+513 pp; hardbound, ISBN 978-3-540-25031-9, $109.00, €90.90

This comprehensive NLP textbook is strongly algorithm-oriented and designed for talented computer programmers who might or might not be linguists. The book occupies a market niche in between that of Jurafsky and Martin (2008) and my own humble effort (Covington 1994); it resembles the latter in approach and the former in scope. Perhaps more than either of those, Nugues's book is also useful to working professionals as a handbook of techniques and algorithms.

Everything is here—everything, that is, except speech synthesis and recognition; phonetics receives only a four-page summary. Those wanting to start an NLP course by covering phonetics in some depth should consider Coleman (2005) as well as Jurafsky and Martin (2008).

After a brief overview, Nugues covers corpus linguistics, markup languages, text statistics, morphology, part-of-speech tagging (two ways), parsing (several ways), semantics, and discourse. "Neat" and "scruffy" approaches are deftly interleaved and compared. Unification-based grammar, event semantics, and tools such as WordNet and the Penn Treebank are covered in some detail. The syntax section includes dependency grammar and even the very recent work of Nivre (2006), as well as partial parsing and statistical approaches. Many important algorithms are presented ready to run, or nearly so, as Prolog or Perl code. If, for example, you want to build a Cocke–Kasami–Younger parser, this is the place to look for directions.

Explanations are lucid and to-the-point. Here is an example. Nugues is discussing the fact that, if you sample a corpus for *n*-grams, some will not occur in your sample at all, but it would be a mistake to consider the unseen ones to be infinitely rare (frequency 0). Thus the counts need to be adjusted:

> Good-Turing estimation . . . reestimates the counts of the *n*-grams observed in the corpus by discounting them, and it shifts the probability mass it has shaved to the unseen bigrams.

The formula follows, but the reader approaches it equipped with the unforgettable image of someone shaving excess material off the peaks and bulges of a histogram and spreading it into the valleys.

This is an ambitious textbook, and, in my opinion, too much for one semester; it must be used selectively. If nothing is skipped, then besides getting a thorough course in natural language processing (except phonetics), the student is expected to learn both Perl and Prolog along the way, aided by a 50-page Prolog handbook in Appendix A. My experience is that Prolog is very hard to learn on the fly; in fact, extensive experience with other languages may be a disadvantage because Prolog is so different. Nonetheless, all the information needed to learn Prolog is here. Perl is treated more casually because it lends itself much more easily to incremental learning.

Other minor quibbles are possible. On page 31, *automata* is used as singular. On page 69, the treatment of RTF, TEX, and LATEX is so short that the student will come away largely unaware of what each of them is actually designed for. This is not pernicious, however, as the student will realize that he or she is inadequately informed.

Nonetheless, this is an unusually useful and well-written book, and I plan to recommend it to my students as well as using it myself as a handbook.—*Michael A. Covington, The University of Georgia*

References

Coleman, John. 2005. *Introducing Speech and Language Processing*. Cambridge, UK: Cambridge University Press.

Covington, Michael A. 1994. *Natural Language Processing for Prolog Programmers.* Englewood Cliffs, NJ: Prentice Hall.

Jurafsky, Daniel, and James H. Martin. 2008. *Speech and Language Processing*, 2nd edition. Upper Saddle River, NJ: Prentice Hall.

Nivre, Joakim. 2006. *Inductive Dependency Parsing.* Dordrecht, The Netherlands: Springer.

Language and the Learning Curve: A New Theory of Syntactic Development

Anat Ninio
(Hebrew University of Jerusalem)

Oxford University Press, 2006, xiv+206 pp; hardbound, ISBN 0-19-929981-1/ 978-0-19-929981-2, £55.00; paperbound, ISBN 0-19-929982-X / 978-0-19-929982-9, £24.95

Ninio proposes a provocative theory of early syntactic development. According to her theory, children do not create any abstract representation of language in the form of rules or schemas, nor do they develop any systematic *linking rules* between syntax and semantics. Instead, they learn a lexicalist syntax. Syntactic development consists of learning, for each individual word, the potential predicate–argument relations and their appropriate syntactic realizations. Semantic valency (the potential semantic relations of a word with other items) and syntactic valency (the ways to express those relations in sentences) are learned for each word separately. The syntactic structure of a sentence is projected from the lexical valency of words by recursively applying a single binary operation, Merge, which combines the Head (the predicate word) to its Dependent (the argument).

Ninio argues that the elimination of the abstract rules from the process of syntactic development is supported by evidence from child language research. Examination of the learning curves (i.e., performance vs. experience) for different syntactic patterns shows no sudden change in children's productivity with a particular syntactic pattern, or in their ability to generalize that pattern to novel items. At the same time, learning a syntactic pattern speeds up with experience. The learning curves have an accelerating nonlinear shape, which suggests that the acquired item-specific syntactic forms facilitate the acquisition of the same syntactic form for new items. Ninio suggests that, even though the knowledge of syntax is item-specific, lexical items are not isolated. Instead, they are interconnected via *transfer*, the ability to extend what has been learned in one context to new contexts by analogy. The notion of transfer is supposed to explain children's nonlinear learning curve, as well as their ability to generalize their lexical knowledge to novel items. However, the details of this central mechanism are left unspecified: it is not clear how and under which constraints the transfer between two items can take place. Moreover, one expects that the interplay between the item-specific knowledge and the transfer of learning be used to explain some well-studied stages of language learning in children, including imitation, (over)generalization, and recovery from making overgeneralization errors. (This so-called U-shaped learning curve is quite distinct from the nonlinear learning curve that gives the book its title.)

Another radical proposal by Ninio is that semantic similarity plays no role in early syntactic development. More specifically, transfer of learning, and therefore generalization, is based solely on a similarity of form. The developmental evidence provided for this claim is intriguing. For example, in the priming experiments where training on a large set of verbs that appear in transitive sentences helps to elicit transitive usages from children on novel nonce verbs, the semantic similarity between the training and the testing verbs does not affect the rate of generalization. However, there are many studies showing that both children and adults use some kind of mapping between form and meaning in comprehension tasks. For example, preferential-looking studies show that infants choose one novel action over another based on the form of the sentence introducing the action (e.g., Naigles 1990; Fisher 2002). Similarly, adult subjects predict certain semantic properties for the arguments of a novel verb based on the form of the sentence the verb appears in (e.g., Gleitman 1990; Kako 2006). In the absence of any abstract rules, these findings can be explained only through a generalization mechanism that takes into account both syntactic and semantic similarity.

Despite the vagueness of some of the suggested mechanisms and occasional insufficient analysis, Ninio's book presents a thought-provoking account of syntactic development that can be of especial interest to the computational linguistics community. Its proposed view of syntax could be validated through computational modeling. If

it turns out that the proposed theory is indeed compatible with empirical data, the underlying ideas could be directly applied to the representation and use of syntactic knowledge in different computational linguistic applications.—*Afra Alishahi, University of Toronto*

References

Fisher, C. 2002. Structural limits on verb mapping: The role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science*, 5(1):55–64.

Gleitman, L. 1990. The structural sources of verb meanings. *Language Acquisition*, 1:135–176.

Kako, E. 2006. Thematic role properties of subjects and objects. *Cognition*, 101:1–42.

Naigles, L. 1990. Children use syntax to learn verb meanings. *Journal of Child Language*, 17:357–374.

Advances in Natural Multimodal Dialogue Systems

Jan van Kuppevelt, Laila Dybkjær, and Niels Ole Bernsen (editors)
(University of Southern Denmark)

Springer (Text, speech, and language technology series, edited by Nancy Ide and Jean Véronis, volume 30),
2005, ix+373 pp; hardbound,
ISBN 978-1-4020-3032-4, $179.00

"The chapters in this book jointly contribute to what we shall call the field of natural and multimodal interactive systems engineering. This is not yet a well-established field of research and commercial development but, rather, an emerging one in all aspects. It brings together, in a process that, arguably, was bound to happen, contributors from many different, and often far more established, fields of research and industrial development. To mention but a few, these include speech technology, computer graphics, and computer vision. The field's rapid expansion seems driven by a shared vision of the potential of new interactive modalities of information representation and exchange for radically transforming the world of computer systems, networks, devices, applications, and so on, from the GUI (graphical user interface) paradigm into something which will enable a far deeper and much more intuitive and natural integration of computer systems into people's work and lives.

"Jointly, the chapters present a broad and detailed picture of where natural and multimodal interactive systems engineering stands today. The book is based on selected presentations made at the International Workshop on Natural, Intelligent, and Effective Interaction in Multimodal Dialogue Systems held in Copenhagen, Denmark, in 2002 and sponsored by the European CLASS project. CLASS was initiated on the request of the European Commission with the purpose of supporting and stimulating collaboration among Human Language Technology (HLT) projects as well as between HLT projects and relevant projects outside Europe. The purpose of the workshop was to bring together researchers from academia and industry to discuss innovative approaches and challenges in natural and multimodal interactive systems engineering."—*From the editors' preface*

The contents of the volume are as follows:

"The psychology and technology of talking heads: Applications in language learning" by Dominic W. Massaro
"Effective interaction with talking animated agents in dialogue systems" by Björn Granström and David House
"Controlling the gaze of conversational agents" by Dirk Heylen, Ivo van Es, Anton Nijholt, and Betsy van Dijk
"MIND: A context-based multimodal interpretation framework in conversational systems" by Joyce Y. Chai, Shimei Pan, and Michelle X. Zhou
"A general purpose architecture for intelligent tutoring systems" by Brady Clark, Oliver Lemon, Alexander Gruenstein, Elizabeth Owen Bratt, John Fry, Stanley Peters, Heather Pon-Barry, Karl Schultz, Zack Thomsen-Gray, and Pucktada Treeratpituk
"MIAMM—A multimodal dialogue system using haptics" by Norbert Reithinger, Dirk Fedeler, Ashwani Kumar, Christoph Lauer, Elsa Pecourt, and Laurent Romary
"Adaptive human–computer dialogue" by Sorin Dusan and James Flanagan
"Machine learning approaches to human dialogue modelling" by Yorick Wilks, Nick Webb, Andrea Setzer, Mark Hepple, and Roberta Catizone

Evaluation of Text and Speech Systems

Laila Dybkjær, Holmer Hemsen, and Wolfgang Minker (editors)
(University of Southern Denmark and University of Ulm)

Springer (Text, speech, and language technology series, edited by Nancy Ide and Jean Véronis, volume 37), 2007, xxiii+288 pp; hardbound, ISBN 978-1-4020-5815-4, $149.00

"This book has its point of departure in courses held at the Tenth European Language and Speech Network (ELSNET) Summer School on Language and Speech Communication which took place at NISLab in Odense, Denmark, in July 2002. The topic of the summer school was 'Evaluation and Assessment of Text and Speech Systems.'

"Nine (groups of) lecturers contributed to the summer school with courses on evaluation of a range of important aspects of text and speech systems, including speaker recognition, speech synthesis, talking animated interface agents, part-of-speech tagging and parsing technologies, machine translation, question-answering and information retrieval systems, spoken dialogue systems, language resources, and methods and formats for the representation and annotation of language resources. Eight of these (groups of) lecturers agreed to contribute a chapter to the present book. Since we wanted to keep all the aspects covered by the summer school, an additional author was invited to address the area of speaker recognition and to add speech recognition, which we felt was important to include in the book. Although the point of departure for the book was the ELSNET summer school held in 2002, the decision to make a book was made considerably later. Thus the work on the chapters was only initiated in 2004. First drafts were submitted and reviewed in 2005 and final versions were ready in 2006."—*From the editors' preface*

The contents of the volume are as follows:
"Speech and speaker recognition evaluation" by Sadaoki Furui
"Evaluation of speech synthesis" by Nick Campbell
"Modelling and evaluating verbal and non-verbal communication in talking animated interface agents" by Björn Granström and David House
"Evaluating part-of-speech tagging and parsing" by Patrick Paroubek
"General principles of user-oriented evaluation" by Margaret King
"An overview of evaluation methods in TREC ad hoc information retrieval and TREC question answering" by Simone Teufel
"Spoken dialogue systems evaluation" by Niels Ole Bernsen, Laila Dybkjær, and Wolfgang Minker
"Linguistic resources, development, and evaluation of text and speech systems" by Christopher Cieri
"Towards international standards for language resources" by Nancy Ide and Laurent Romary

Text Entry Systems: Mobility, Accessibility, Universality

I. Scott MacKenzie and Kumiko Tanaka-Ishii (editors)
(York University, Toronto and University of Tokyo)

San Francisco: Morgan Kaufmann Publishers, 2007, x+332 pp; paperbound, ISBN 978-0-12-373591-1, $49.95, £28.99, €41.95

"Advances in technology have helped create a connected world, and this is none more apparent than in the area of text entry. The growing popularity and success of textual communication has inspired a variety of text entry systems, modalities, and users. There is a text entry system that will meet the needs of all kinds of users—a teen sending an IM with her phone, a businessman checking e-mail on his blackberry, or a visually impaired woman using a Braille keyboard on her PDA. The capabilities and modalities of text entry are widespread and evolving.

"*Text Entry Systems* is a guidebook on the details and foundations of text entry methods, the effectiveness of current modes of text entry, advances in technology, and the creation of new systems. Authoritative researchers from the fields of HCI, handwriting recognition, speech recognition, computational linguistics, natural language processing, universal access, industrial design, cognitive science, and image processing all provide their expertise. They address the wide reach of this technology, including design for various languages and accommodations for those with special physical conditions, using specific examples and offering solutions."—*From the publisher's announcement*

Publications Received

Books listed below that are marked with a † have been selected for review in a future issue, and reviewers have been assigned to each.

Authors and publishers who wish their publications to be considered for review in *Computational Linguistics* should send a copy to the book review editor, Graeme Hirst, Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 3G4. All relevant books received will be listed, but not all can be reviewed. Technical reports (other than dissertations) will not be listed or reviewed. Authors should be aware that some publishers will not send books for review (even when instructed to do so); authors wishing to enquire as to whether their book has been received for review may contact the book review editor.

Readers who wish to be considered as book reviewers for the journal should contact the book review editor, outlining their qualifications, by e-mail at *gh@cs.toronto.edu* or at the address above.

Words and Intelligence II: Essays in Honor of Yorick Wilks
Khurshid Ahmad, Christopher Brewster, and Mark Stevenson (editors)
(Trinity College, Dublin and University of Sheffield)
Springer (Text, Speech and Language Technology series, edited by Nancy Ide and Jean Véronis, volume 36), 2007, xiv+279 pp; hardbound, ISBN 978-1-4020-5832-5, $139.00

The Categorization of Spatial Entities in Language and Cognition
Michel Aurnague, Maya Hickmann, and Laure Vieu
(CNRS–Université de Toulouse–Le Mirail, CNRS–Université de Paris VIII, and CNRS–Université Paul Sabatier, Toulouse III) John Benjamins Publishing (Human Cognitive Processing series, edited by Marcelo Dascal et al., volume 20), 2007, viii+371 pp; hardbound, ISBN 978-90-272-2374-6, $144.00, €120.00

Semi-Supervised Learning
Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien
(Max Planck Institute for Biological Cybernetics, Tübingen)
The MIT Press, 2007, xiii+506 pp; hardbound, ISBN 978-0-262-03358-9, $50.00, £32.95

Chomsky's Universal Grammar: An Introduction
V.J. Cook and Mark Newson
(University of Newcastle upon Tyne and Eötvös University)
Blackwell Publishing, 2007, vii+326 pp; hardbound, ISBN 978-1-4051-1186-7, $89.95; paperbound, ISBN 978-1-4051-1187-4, $39.95

Corpus Linguistics 25 Years on
Ronberta Facchinetti (editor)
(University of Verona)
Rodopi (Language and Computers: Studies in practical linguistics, edited by Christian Mair et al., volume 62), 2007, v+385 pp; hardbound, ISBN 978-90-420-2195-2, €80.00

†Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues
Trude Heift and Mathias Schulze
(Simon Fraser University and University of Waterloo)
Routledge (Routledge series in computer-assisted language learning, edited by Carol Chappelle), 2007, xviii+283 pp; hardbound, ISBN 978-0-415-36191-0, $115.00

Word Frequency and Lexical Diffusion
Betty S. Phillips
(Indiana State University)
Palgrave Macmillan (Palgrave studies in language history and language change, edited by Charles Jones), 2007, xiv+252 pp; hardbound, ISBN 978-1-4039-3232-7, $80.00

Prosodic Orientation in English Conversation
Beatrice Szczepek Reed
(University of Nottingham)
Palgrave Macmillan, 2006, xiv+331 pp; hardbound, ISBN 978-0-230-00872-4, $80.00

Language, Discourse, and Social Psychology
Ann Weatherall, Bernadette M. Watson, and Cindy Gallois
(Victoria University of Wellington and University of Queensland)
Palgrave Macmillan (Palgrave advances in linguistics, edited by Christopher N. Candlin), 2007, xvii+309 pp; hardbound, ISBN 978-1-4039-9594-0, $90.00; paperbound, ISBN 978-1-4039-9595-7, $34.95

Contrastive Linguistics: History, philosophy, and methodology
Pan Wenguo and Tham Wai Mun
(East China Normal University and Nanyang Technological University)
Continuum, 2007, xii+287 pp; hardbound, ISBN 978-0-8264-8634-9, £85.00, $150.00

On the Syntactic Composition of Manner and Motion
Maria Luisa Zubizarreta and Eunjeong Oh
(University of Southern California and Korea University)
The MIT Press, 2007, xi+228 pp; paperbound, ISBN 978-0-262-74029-6, $32.00, £19.95