

Briefly Noted

Memory-Based Parsing

Sandra Kübler

(University of Tübingen)

Amsterdam: John Benjamins (Natural language processing series, edited by Ruslan Mitkov, volume 7), 2004, viii+294 pp; hardbound, ISBN 90-272-4991-1, \$138.00, and 1-58811-590-9, €115.00

This monograph, based on Sandra Kübler's Ph.D. thesis, introduces the reader to research at the intersection of data-driven parsing and memory-based learning. Compared to other approaches to parsing, including knowledge-based methods, memory-based parsing takes the provocative standpoint that new structures can be parsed by analogical reasoning over stored structures, rather than by abstracted rules. All that the approach needs is parsed example sentences stored in memory and a similarity function to find candidate nearest-neighbor sentences that can act as the basis for the analogical-reasoning step.

Kübler begins with a walk past the known approaches in the field to partial and full memory-based parsing. As for partial parsing (constituent chunking and basic grammatical-relation assignment such as subject-verb relations), the work of Daelemans, Veenstra, Buchholz, Tjong Kim Sang, and others using memory-based learning is summarized, as well as Krymolowski, Argamon, and Dagan's work on memory-based sequence learning. Kübler then proceeds to review Streiter, whose memory-based parser is an example of a more holistic, sentence-oriented approach which, in contrast to the aforementioned approaches, needs a more complex similarity metric to compute the distance between a complete new sentence and stored parsed sentences. A separate chapter is devoted to data-oriented parsing, which uses probabilistic machinery and extensive back-off from larger to smaller substructures, instead of a single similarity function; in particular, two nonprobabilistic variants by Bod and De Pauw are close cousins of the other memory-based approaches.

The heart of the book is the TüSBL (Tübingen Similarity-Based Learning) memory-based parser, which implements a similarity-based approach that, analogous to Streiter's approach, attempts to fully parse complete sen-

tences by analogy, as rapidly as possible. Kübler's solution is original. While a naive approach based on a similarity between full sentences would be able to correctly parse only a few sentences very close to sentences in memory, Kübler introduces at least two smart generalization enhancements. One is that the TüSBL parser has more than one similarity metric. When a new sentence is parsed, it is first analyzed at the levels of part-of-speech tags and base constituents. If no reliable nearest neighbors matching on the word level can be found in memory, the other levels act as back-offs on which to measure similarity. The second enhancement is that the search for nearest neighbors is extended by allowing them to have a word or constituent too many or too few or to be longer but contain a good matching subtree. TüSBL's similarity metric, or rather its case-based reasoning function, is actually aware of the internal structure of the nearest-neighbor trees and the partial syntactic structure of new sentences.

TüSBL is put to the test on the NEGRA-formatted TüBa-DS treebank of spontaneous speech in specific domains (hotel reservations, business appointments, and travel scheduling), gathered in the context of the VERBMOBIL project. Kübler makes the credible point that data of this type offer a more interesting challenge to parsing methods than nonspontaneous, professionally authored texts in a (similarly) closed domain, such as the *Wall Street Journal* Penn Treebank. An excellent point is made on the limitations of the standard PARSEVAL evaluation's focus on syntactic chunking and labeling; arguably, the correctness of the parser in assigning functional labels to grammatical relations is at least as interesting as an evaluation metric. From the reported results we learn that TüSBL does a good job; it attains a PARSEVAL *F*-score of about 85 on the spontaneous speech corpus. We also learn that the back-off part of TüSBL performs as well as the whole TüSBL system in PARSEVAL terms. This underlines the point that parsing based on matching on smaller, local structures, as is done by most other memory-based methods, performs at least on par with more holistic methods. However, TüSBL's holistic memory-based core is more reliable in assigning correct functional tags to correctly identified grammatical relations.

In sum, the book offers a comprehensive and well-illustrated overview of the area of memory-based parsing, makes all the right methodological points, and describes a system that performs a complex task in a refreshingly simple and smart way.—*Antal van den Bosch, Tilburg University*

Computational Linguistics and Beyond

**Chu-Ren Huang and
Winfried Lenders (editors)**

(Institute of Linguistics, Academia Sinica,
and University of Bonn)

Taipei: Institute of Linguistics, Academia Sinica (Language and linguistics monograph series B, Frontiers in linguistics I), 2004, viii+207 pp; paperbound, ISBN 957-01-7610-5

This volume consists of the plenary lectures and two workshop panels from the 19th International Conference on Computational Linguistics (Taipei, 2002), complementing the already-published conference proceedings.

The contents of the volume are as follows:

- “Computational Linguistics and Beyond: An Introduction,” by Chu-Ren Huang and Winfried Lenders
- “FrameNet and Representing the Link between Semantic and Syntactic Relations,” by Charles J. Fillmore, Josef Ruppenhofer, and Collin F. Baker
- “Computational Studies of Language Evolution,” by William S.-Y. Yang, Jinyun Ke, and James W. Minett
- “New Chances for Deep Linguistic Processing,” by Hans Uszkoreit
- “The Roles of Natural Language and XML in the Semantic Web,” by Graham Wilcock, Paul Buitelaar, Antonio Pareja-Lora, Barrett Bryant, Jimmy Lin, and Nancy Ide
- “Chinese Language Processing at the Dawn of the 21st Century,” by Benjamin K. T’sou

Anaphora Processing: Linguistic, Cognitive, and Computational Modelling

**António Branco, Tony McEnery,
and Ruslan Mitkov (editors)**

(Universidade de Lisboa, Lancaster University, and University of Wolverhampton)

Amsterdam and Philadelphia: John Benjamins (Current studies in linguistic theory), 2005, x+449 pp; hardbound, ISBN 90-272-4777-3 and 1-58811-621-2, €120.00, \$144.00

This volume contains revised and extended versions of 20 papers chosen from among those presented at the 2002 Discourse Anaphora and Anaphora Resolution Colloquium in Lisbon.

The contents of the volume are as follows:

- “A Sequenced Model of Anaphora and Ellipsis Resolution,” by Shalom Lappin
- “How to Deal with Wicked Anaphora,” by Dan Cristea and Oana-Diana Postolache
- “A Machine Learning Approach to Preference Strategies for Anaphora Resolution,” by Roland Stuckardt
- “Decomposing Discourse,” by Joel Tetreault
- “A Lightweight Approach to Coreference Resolution for Named Entities in Text,” by Marin Dimitrov, Kalina Bontcheva, Hamish Cunningham, and Diana Maynard
- “A Unified Treatment of Spanish *se*,” by Randy Sharp
- “Binding and Beyond: Issues in Backward Anaphora,” by Eric Reuland and Sergey Avrutin
- “Modelling Referential Choice in Discourse: A Cognitive Calculative Approach and a Neural Network Approach,” by André Grüning and Andrej A. Kibrik
- “Degrees of Indirectness: Two Types of Implicit Referents and Their Retrieval via Unaccounted Pronouns,” by Francis Cornish
- “Pronominal Interpretation and the Syntax-Discourse Interface: Real-Time Comprehension and Neurological Properties,” by Maria Mercedes Piñango and Petra Burkhardt
- “Top-Down and Bottom-Up Effects on the Interpretation of Weak Object Pronouns in Greek,” by Stavroula-Thaleia Kousta
- “Different Forms Have Different Referential Properties: Implications for the Notion of ‘Salience,’” by Elsi Kaiser

- “Referential Accessibility and Anaphor Resolution: The Case of the French Hybrid Demonstrative Pronoun *celui-ci/celle-ci*,” by Marion Fossard and François Rigalleau
- “The Predicate-Argument Structure of Discourse Connectives: A Corpus-Based Study,” by Cassandre Creswell, Katherin Forbes, Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, and Bonnie Webber
- “Combining Centering-Based Models of Saliency and Information Structure for Resolving Intersentential Pronominal Anaphora,” by Costanza Navarretta
- “Pronouns without NP Antecedents: How Do We Know When a Pronoun Is Referential?” by Jeanette Gundel, Nancy Hedberg, and Ron Zacharski
- “Syntactic Form and Discourse Accessibility,” by Gregory Ward and Andrew Kehler
- “Coreference and Anaphoric Relations of Demonstrative Noun Phrases in Multilingual Corpus,” by Renata Vieira, Susanne Salmon-Alt, and Caroline Gasperin
- “Anaphoric Demonstratives: Dealing with the Hard Cases,” by Marco Rocha
- “Focus, Activation, and *This*-Noun Phrases: An Empirical Study,” by Massimo Poesio and Natalia N. Modjeska

