

Book Reviews

Spotting and Discovering Terms through Natural Language Processing

Christian Jacquemin
(University of Paris 11)

Cambridge, MA: The MIT Press, 2001,
viii+378 pp; hardbound, ISBN
0-262-10085-1, \$52.95, £36.50

Reviewed by
Sophia Ananiadou
University of Salford

Christian Jacquemin's book *Spotting and Discovering Terms through Natural Language Processing* is a much needed and most welcome addition to the field of computational terminology. The central issue of this book is the in-depth examination of term variation, that is, the morphological, syntactic, or semantic transformation of multiword terms, capturing the fact that the same concept can be expressed through term variants. From an application point of view, dealing with term variation would result in more efficient indexing engines and better term extraction tools, among other applications. The book also describes in depth the author's FASTR system, a natural language processor for term variation. There are plenty of examples throughout the book and a set of metarules in the appendix for the interested reader who would like to use FASTR.

Chapter 2 provides an excellent overview and a comparison of numerous algorithms for automatic term extraction systems as well as techniques for recognizing multiword terms. Six systems are examined in detail. The author also discusses the related area of automatic indexing, the main purpose of which is to assign content descriptors to documents. He describes 11 studies of phrase indexing (indexing through multiword units). A useful comparison of the different indexers is provided (p. 113). The aim of this extended discussion of concerns, methods, and systems is to introduce the main ideas behind FASTR, which is used for both term recognition and terminological enrichment. The identification of variants of terms is crucial in Jacquemin's term-spotting technique, and because such variation occurs in corpora frequently, "it makes sense to build an indexer on a variational mechanism" (p. 115). In addition, variants represent on average 28% of multiword term occurrences (p. 219).

Since term spotting uses various linguistic features, FASTR's grammar formalism is unification-based, inspired by PATR-II, in order to represent different types of features. Information is embedded in nontyped feature structures, which include two additional facilities: disjunction and negation. Term rules in FASTR are composed of a context-free skeleton that describes the constituent structure of terms and logical constraints (feature structures) that denote the information linked to the nodes of the context-free skeleton. The morphological model of FASTR is concatenative morphology enriched with feature structures and a list of suffixes. This applies to both inflectional and derivational morphology, although in FASTR there are two ways of describing derivational morphology: dynamically (similar to inflection) and statically,

where derivational links between words and their stems are explicitly stated in the single-word lexicon. To represent the syntax of multiword terms and to cope with term variability, FASTR's formalism uses the notions of *extended domain of locality* and *lexicalization* from lexicalized tree adjoining grammars (LTAGs) (Abeillé and Rambow 2000).

Jacquemin's approach relies on metarules that exploit syntagmatic and paradigmatic information. Syntagmatically, they describe structural mappings between, say, a multiword term and its syntactic or phrasal variants. Paradigmatically, they describe lexical relationships between the words of a term and the words of its variants that participate in the mappings. These lexical relationships may be morphological or semantic in nature or both. Morphological relationships cover words that belong to the same derivational family and express the fact that such words share a common root. Semantic relationships cover words that are linked (by synonymy or antonymy, for example); that is, they capture the fact that the words have something in common semantically. Taken together, the syntagmatic and paradigmatic elements of metarules provide complementary constraints that prevent the generation of undesirable mappings.

Chapter 4 is dedicated to the metagrammar of FASTR, which operates on top of the lexical and terminological data. Metarules dynamically transform rules written for controlled terms into new rules that, in turn, can be used to extract variants from texts. The concept of metarule is inspired by the work of Harris (Harris et al. 1989) and is further influenced by generalized phase structure grammars (Gazdar et al. 1985) and feature-based LTAGs (Srinivas et al. 1994) and by work on lexically based formalisms that seek to reduce the complexity and size of the grammar. The author contrasts his use of metarules with other formalisms (p. 157). The main features are that metarules in FASTR are more generic than those in other formalisms and, since they are dedicated to term and variant extraction, they are also much simpler. Of further interest in FASTR's metarule formalism is that it contains a compiler and an interpreter of regular expressions, and this allows the output of more complex structures (for example, coordinated structures).

Term variation is explicitly described for English through four types of elementary term variation: coordination, permutation, modification/substitution, and elision. Chapter 5 provides a practical description of how to build, tune, and evaluate metarules for syntactic term variants in English that can be produced from a controlled vocabulary (this vocabulary would typically be the result of the term extraction phase of FASTR). The chapter begins by describing variations of binary terms, for example, how to link a binary compound term such as *tooth root* with a noun phrase having the same meaning, *the root of a lower premolar tooth*, by means of permutation metarules. Section 5.2 describes how to extend variations from binary to *n*-ary terms, taking into account all structural ambiguities. A detailed description is given for an example of variation involving coordination, as this is one of the most difficult phenomena to tackle. Section 5.4 illustrates how the metagrammar can be further tuned experimentally, using occurrences extracted from the corpus through paradigmatic rules. These paradigmatic rules are refined and transformed into filtering metarules (i.e., metarules augmented with constraints), which are then used by FASTR for extracting term variants. Evaluation of the different variants extracted by FASTR shows high precision and recall. The author reports poor performance in the case of elliptic variants, however, which he notes is largely due to the noncontextuality of FASTR's parsing approach (p. 169).

The recognition of morphosyntactic variants (as opposed to the syntactic variants of Chapter 5) is dealt with in Chapter 7. FASTR's formalism is extended to include metarules for morphosyntactic variations; for example, *development of mouse embryos* is considered a morphosyntactic rather than a syntactic variant of *embryonic development*,

because the adjective *embryonic* is transformed into the noun *embryo* in the variant (p. 273). Single words contain morphological links between words and their root lemma (for English, derived from the CELEX database); these links are also expressed in the metagrammar of morphosyntactic variation. Variants of binary terms are presented one morphological transformation at a time: adjective to noun variation, noun to verb variation, and so on. The analysis of these variants is important, as terminology work has concentrated mostly on nouns, leaving unexplored non-nominal terminological occurrences, which are equally important, as they capture relations between terms.

Semantic variation, described briefly in Chapter 8, concludes the study of types of term variation. FASTR's formalism is further extended to account for this type of variation. For example, *benign mouse skin tumors* can be recognized as a semantic variant of *benign neoplasms* provided that there is a semantic link between the words *tumors* and *neoplasms* and that the insertion of the modifier *mouse skin* is accepted (p. 299). The semantic links used for extracting semantic variants in English are the synonymy links of WordNet 1.6. Morphosyntactic term variation can be seen as a special case of semantic variation, since legitimate morphological links can be established between semantically related words.

I found it more useful to read Chapter 6 after Chapters 5, 7, and 8. It describes how FASTR can be used for "incremental term enrichment," taking into account existing terms and trying to relate newly acquired terms to the previously existing terminologies. The idea is that, by deconstructing term variants, we can detect possible term associations and acquire new terms. "The variant *uterine and carotid artery* of *uterine artery* is the opportunity for discovering the term *carotid artery*. The context of acquisition shows that both terms can be coordinated, indicating that the meanings of *artery* in the original term *uterine artery* and in the candidate term *carotid artery* are similar: a blood vessel" (p. 241). The analysis of such variants produces a set of patterns, called *pattern extractors*. Each pattern extractor is then attached to a specific metarule. Starting from any variant detected by such a metarule, the extractor outputs the corresponding candidate term. This technique of term enrichment is also very useful for producing conceptual relations that can be used for automatic thesaurus construction.

In conclusion, this book is clearly written and explores an interesting and very applicable area of computational terminology: term variation. An in-depth analysis of current techniques in terminology, a detailed bibliography, and a plethora of examples demonstrating how to write rules for and use FASTR make this book an important acquisition not only for researchers interested in computational terminology but also for the wider natural language processing community. Anyone concerned with issues of automatic indexing, automatic thesaurus construction, rapid adaptation to new domains, and robust processing of not only domain-specific terminology in the classical sense but also the phrases in which variants of terms are found will learn much from this book.

References

- Abeillé, Anne and Owen Rambow, editors. 2000. *Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing*. CSLI, Stanford, CA.
- Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Harvard University Press.
- Harris, Zellig S., Michael Gottfried, Thomas Ryckman, Paul Mattick, Jr., Anne Daladier, T. N. Harris, and S. Harris. 1989. *The Form of Information in Science: Analysis of Immunology Sublanguage* (Boston Studies in the Philosophy of Science, volume 104). Kluwer Academic.
- Srinivas, B., Dania Egedi, Christine Doran, and Tilman Becker. 1994. Lexicalization and grammar development. In Harald Trost, editor, *Proceedings of KONVENS '94*, Vienna, pages 310–319.

Sophia Ananiadou is a Senior Lecturer in Computer Science at the University of Salford, U.K. She teaches natural language processing, and her main research interests are in computational terminology, ontologies, and information extraction. Currently, she is applying natural language processing techniques to bioinformatics. Ananiadou's address is: Computer Science, School of Sciences, University of Salford, Salford M5 4WT, U.K.; e-mail: S.Ananiadou@salford.ac.uk.

Automatic Summarization

Inderjeet Mani

(MITRE Corporation and Georgetown University)

Amsterdam: John Benjamins (Natural language processing series, edited by Ruslan Mitkov, volume 3), 2001, xi+285 pp; hardbound, ISBN 1-58811-059-1, \$82.00; paperbound, ISBN 1-58811-060-5, \$29.95

Reviewed by

Chris D. Paice

Lancaster University

Though the first attempts at automatic abstracting were made about 45 years ago, this area was until quite recently a rather obscure specialism. No doubt, this was due to the nonavailability of machine-readable texts: It seemed absurd to go to the trouble and expense of keypunching complete texts and then use a program to throw away 90% of them! Since about 1990, with the urgent need to manage the mass of information available on the World Wide Web, the field of automatic summarization has exploded into prominence.

Automatic Summarization is the first monograph on the subject, and as such it is sure to be widely read and cited. The author, Inderjeet Mani, is currently one of the field's most active researchers, and it is therefore no surprise that the book has an authoritative feel. In general, it provides good, balanced coverage of the field (except that cross-lingual summarization is not covered) and describes a great deal of the published research in the area, with a bibliography of well over 200 items. Mani succeeds in conveying a sense that this is a most challenging field for system designers. With a grim topicality, many of the illustrations used refer to the summarization of reports on terrorist incidents.

The book is structured sensibly into nine main chapters. Chapter 1, "Preliminaries," introduces a range of fundamental terms and ideas. A distinction is made between *extracts*, consisting of material (typically sentences) taken straight from the source, and *abstracts*, "at least some of whose material is not present in the input." Chapter 2 looks at professional (i.e., nonautomatic) summarizing.

Chapter 3 is about extraction, which involves assigning an importance score to each sentence in a text and outputting those with the highest score. Various importance clues may be used, including the presence of concentrations of content words, use of cue expressions such as *we see that* and *certainly*, and the location of the sentence in the text. A long section describes and discusses the work by Edmundson and by later workers pursuing similar approaches. Coming up to date, the chapter continues with a discussion of corpus-based approaches to sentence extraction. Chapter 4 is entitled "Revision" and discusses how to deal with problems of cohesion in extracted sentences that arise from the occurrence of anaphors and other awkward features. Chapter 5 discusses the potential use of discourse-level information in approaches such as cohesion graphs, text tiling, lexical chains, and rhetorical structure theory.

In Chapter 6, the author turns to abstraction, as opposed to sentence extraction. Typical approaches here have involved the use of sketchy scripts or frames tailored to a particular domain. Analysis of a text results in instantiation of the appropriate

frame, and when this process is complete, output can be generated, either as stylized “canned text” or by the use of proper text generation techniques. A major problem with this approach is its inflexibility: A text can be summarized only if the correct frame is available and can be identified.

Chapter 7 deals with multidocument summarization, which is now seen as a most important area (terrorism news reports, again). This is not simply a matter of merging a number of separate descriptions of the same issue or event: There is a need to highlight discrepancies between different reports and to deal with time-related issues (e.g., changing estimates of numbers of casualties). Another nontrivial problem is handling cross-document coreferences—for example, ensuring that mentions of *Johnson* in two different documents actually refer to the same person.

Chapter 8, “Multimedia Summarization,” is in my view the weakest in the book. Mani starts by acknowledging that this area is evolving rapidly, so that “no clear principles have emerged” (p. 209). But he then takes this as an excuse for a rather haphazard collection of topics (summarization of dialog, of video, and of diagrams, and multimedia briefing generation) dealt with in varying degrees of detail. The chapter is only 12 pages long (nine pages of actual text), whereas to begin to tackle this area, at least twice the space is surely required. To give a proper structure, the chapter should begin with a discussion of possibilities for the direct analysis of various nontext media, followed by an examination, with examples, of how different media (especially text) can support one another in building useful systems.

Chapter 9, on the evaluation of summarization systems, is fully three times as long as the previous chapter, and there is no space here to go into details. Though the exposition gets a bit tangled in places, the account is generally authoritative and thorough and provides a valuable overview of this problematic area.

All the main chapters from Chapter 3 onward finish with (1) a table summarizing the various approaches discussed, with their strengths and weaknesses, and (2) a glossary giving a brief definition of the various terms introduced in the chapter. These are both valuable features, though it seems structurally inconsistent that the former are included as ordinary numbered tables and the latter as separate “Review” sections.

Since this is the only textbook in its field, it will probably become required reading for new researchers, who will assume that it provides a full and balanced view. Its very patchy presentation of historical research therefore seems regrettable. Surely, the seminal 1958 paper by H. P. Luhn deserves a section of its own? Apart from its historical importance, this work provides a clear and clean introduction to the sentence extraction paradigm. Luhn’s paper is in fact mentioned twice in the book, but in neither case is it very prominent. Neither is there any mention of James E. Rush and his work on the ADAM system (Rush, Salvador, and Zamora 1971) (though the successor system for the Chemical Abstracts Service is discussed). Chapter 4, on sentence revision, is short and should certainly have included a summary of the pioneering efforts by Mathis, Rush, and Young (1973). So little research was carried out prior to the 1980s that it seems a pity to ignore so much of it! (By contrast, the ideas of Skoro-chod’ko [1972] on cohesion graphs are—quite properly—discussed at some length in Chapter 5.)

Although (Chapter 8 apart) I have no problems with the general structure and content of the book, there were various minor defects that for me rather spoiled the overall impression. First, although on the whole the book is readable enough, there are a few places where the description of processes or ideas becomes so convoluted or condensed that even repeated rereading failed to reveal the meaning. In some places, intelligibility would have been much improved by proper use of examples (p. 82 is a glaring example). I also felt that some of the formulae could have been justified and

explained a bit more fully. The diagrams vary between clear and pertinent to obscure or curious (two figures in Chapter 7 seem to show a succession of aquaria with various documents suspended inside them).

A book like this is certainly going to be used very much for reference purposes and thus the reader should be able to find her way around it. Here, in fact, is my biggest gripe. In a short space of time, I was able to compile a list of 10 phrasal terms that were discussed in the text but did not appear either in the index or in the relevant Review section. Moreover, cross-referencing within the text is minimal. In some cases, there are references to equations or diagrams that are many pages away, necessitating a tedious search. Finally, neither the section numbers nor the page headers include the chapter number, making it hard to find one's way about (though this was probably not the author's decision).

I have actually found this quite a hard book to review. As a textbook, it provides a valuable map of the field, and yet, as far as I can detect, there is nothing here that is notably innovative or controversial; there is very little to focus on except the minor flaws. But taken as a whole, the book certainly provides a timely and informative overview of automatic summarization, and all researchers with an interest in this important field will wish to have it on their bookshelves.

References

- Luhn, H. P. 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2):159–165. Reprinted in Inderjeet Mani and Mark T. Maybury, editors, *Advances in Text Summarization*, MIT Press, 1999, pages 15–21.
- Mathis, B. A., J. E. Rush, and C. E. Young. 1973. Improvement of automatic abstracts by the use of structural analysis. *Journal of the American Society for Information Science*, 24(2):101–109.
- Rush, J. E., J. R. Salvador, and A. Zamora. 1971. Automatic abstracting and indexing. II. Production of indicative abstracts by application of contextual inference and syntactic coherence criteria. *Journal of the American Society for Information Science*, 22(4):260–274.
- Skorochoďko, E. F. 1972. Adaptive method of automatic abstracting and indexing. In C. V. Freiman, editor, *Information Processing 71: Proceedings of the IFIP Congress 71, Ljubljana, Yugoslavia*, North Holland, 1179–1182.

Chris Paice teaches in the Computing Department at Lancaster University, U.K. He has research interests in various aspects of information retrieval and text processing and has published several papers on automatic abstracting, the first appearing in 1981. Paice's address is: Computing Department, Lancaster University, Bailrigg, Lancaster LA1 4YR, U.K.; e-mail: cdp@comp.lancs.ac.uk.

Efficient Processing with Constraint-Logic Grammars Using Grammar Compilation

Guido Minnen
(Motorola Labs)

Stanford: CSLI Publications (Stanford monographs in linguistics), 2001, viii+255 pp; distributed by University of Chicago Press; hardbound, ISBN 1-57586-305-7, \$55.00, £35.00; paperbound, ISBN 0-57586-306-5, \$20.00, £13.00

Reviewed by
Suresh Manandhar
University of York

1. Book Content

This monograph should be of interest to researchers working on building practical and efficient methods for processing highly abstract declarative constraint-based grammars, primarily head-driven phrase structure grammar (HPSG) (Pollard and Sag 1994). The work should also be of interest to researchers in the logic-programming community. The monograph is accessible to anyone with a background in logic programming. Background in grammar formalisms or HPSG is not essential to follow the monograph. Minnen has done a commendable job in making the monograph relatively easy to follow by using numerous well-explained examples. A number of typographic errors in the example programs, however, and a crucial missing figure (Figure 4.17) make the reading somewhat more difficult than it need have been.

Declarative constraint-based grammars are notorious for being highly inefficient from a processing point of view. Minnen's techniques, several of which are explored in the monograph, automatically transform the input grammar into a more specialized grammar that efficiently realizes the user's goal. The techniques can be viewed as performing equivalence transformations on the input logic program to derive a logic program that is more efficient with respect to the input goal. These transformations range from simple strategies such as *literal rearrangement* to more complex ones such as *building recursion reversal* and *magic templates*. Although most of these techniques are closely related to work by other researchers (Strzalkowski 1994, Ramakrishnan 1991), Minnen has extended them to make them suitable for dealing with feature-based grammars.

The monograph is divided into three parts: top-down control, bottom-up control, and lexical rules. Whereas the top-down and bottom-up control chapters primarily deal with grammar rule compilation, the lexical-rules chapter deals with methods for processing lexical rules.

Central to the extraction of control information are two notions:

- *the adornment of a literal*, which identifies those arguments in a literal that are bound.
- *the degree of nondeterminism (DoN)* of a literal, which is the number of alternatives or choice points available when evaluating the literal.

Given a user-specified goal and a user-specified global DoN, Minnen's grammar compilation procedures attempt to transform the grammar nondeterministically into an equivalent grammar until the transformed grammar meets the global DoN requirement.

Minnen's method can be understood generally in terms of two mutually recursive methods:

- Adornment of a goal is used to perform a static abstract interpretation of the program to determine the DoN of each literal.
- The information gained from abstract interpretation is employed to perform a program transformation.

Minnen couples program transformation with relatively sophisticated tabulation techniques to store partial solutions and hence minimize the cost of processing recursive goals with shared subgoals. His tabulation method stores failed goals, successful goals, and currently "opened" goals.

Chapter 3 describes literal rearrangement, in which the literals of a clause are rearranged to make the goal more deterministic. Literal rearrangement employs adornment information specified in the user goal to determine the best literal rearrangement to achieve the specified DoN. Literal rearrangement, although fairly simple in principle, requires a recursive search through the clauses of each literal and hence can be costly. Tabulation becomes essential here. In addition, Minnen describes a local heuristic that chooses the literal with the most arguments (or feature paths, in the case of constraint-based grammars) instantiated to minimize the cost of this search. A simple iterative deepening search over the literal rearrangement procedure starting with DoN = 1 finds the literal rearrangement program transformation with the smallest possible DoN.

When information from two literals is simultaneously required to successfully constrain the processing task, such as in certain treatments of German partial verb phrase topicalization (Example (1)), then literal rearrangement alone is insufficient. In fact, in such cases, Minnen states that *coroutining* or parallel processing of literals is required. He then goes on to show that coroutining can be simulated by *unfolding* the literals and applying literal transformation to the resultant clause.

(1) [Anna lieben]_i wird ..._i Karl.

The second problem that literal rearrangement cannot solve is left recursion, since in a grammar such as the one shown in Figure 1, described by Minnen (p. 72), it is not the literal order that is problematic but that the base case needs to be processed as early as possible. Chapter 4 explores techniques for automatically detecting when such a *building recursion reversal* (BRR) transformation can be applied and describes methods for implementing it correctly. This type of transformation is more involved than the simpler literal rearrangement transformation of the previous chapter, as it involves analyzing argument instantiation patterns called *argument sequences*. For example, in the program in Figure 1, one possible argument sequence (if we trace the second argument of vp/4) is the sequence <Comps, [Comp|Comps]> (given in simplified form here). Roughly speaking, a building recursion reversal changes the order of argument sequences from a sequence such as <Comps, [Comp|Comps]>, which builds structure, to the reversed sequence <[Comp|Comps], Comps>, which consumes structure.

Part 2 covers bottom-up processing using *magic transformations*, which transform the original program by adding an additional literal, known as a *magic rule*, to the

```

vp(Subj, Comps, VSem, P0, P):-
    vp(Subj, [Comp|Comps], VSem, P0, P1),
    np(Comp, P1, P0).

vp(Subj, Comps, VSem, P0, P):-
    v(Subj, Comps, VSem, P0, P).

v(Subj, [Obj,IObj], bring(Subj,IObj,Obj),[bring|P],P).

np(john, [john|P], P).
np(flowers, [flowers|P], P).
np(mary, [mary|P], P).

```

Figure 1

Example program from Minnen (p. 86).

start of the right-hand side of each clause. In the example, magic transformation of the predicate `vp/4` of the program shown in Figure 1 with respect to the goal shown in Figure 2 would result in the clauses shown in Figure 3. The magic rule `magic_vp/4` acts as a guard by instantiating variables and filters out subgoals that cannot be part of the main goal. In a naive bottom-up strategy, all facts are used to deduce the goal. Naive bottom-up processing is expensive in terms of both space and time. The magic rule provides a top-down filtering component and hence makes an otherwise purely data-driven control more goal driven. Minnen explores both the Earley deduction strategy and an improved seminaive bottom-up processing strategy and concludes that the two are very similar, with the semi-naive strategy being slightly better in terms of space requirements. He concedes, however, that to ensure termination, both a subsumption check and an abstraction function are necessary. Subsumption checks are computationally expensive, and abstraction functions have to be user specified, which is a big disadvantage.

The final part of the monograph deals with the treatment of HPSG lexical rules. This work builds upon work reported by Meurers and Minnen (1997). The nice part is that, in Minnen's setup, lexical rules can be viewed as definite clauses, so that techniques from the previous chapters directly apply to lexical rules. As in Meurers and Minnen (1997), lexical rule interaction (through which the output of a lexical rule can be the input of another lexical rule) is modeled by means of a finite-state automaton, computed off-line, that precompiles all the possible interactions between lexical rules. Minnen shows that nondeterminism in lexical rule expansion can be minimized by combining *partial unfolding* with lexical rule interaction. In this way, a large number of choice points that lead to failure can be eliminated at compile time.

2. Final Analysis

Minnen's monograph provides a refreshing entry point for someone wanting to pursue a research program in efficient implementation of constraint-based grammars. Minnen's work complements work on compilation techniques for typed-feature hi-

```

vp(Subj,Comps,bring(john,flowers,mary),P0,P1)

```

Figure 2

User goal.

```

vp(Subj, Comps, VSem, P0, P):-
    magic_vp(Subj,Comps,VSem, P0, P1),
    vp(Subj, [Comp|Comps], VSem, P0, P1),
    np(Comp, P1, P0).

vp(Subj, Comps, VSem, P0, P):-
    magic_vp(Subj,Comps,VSem, P0, P1),
    v(Subj, Comps, VSem, P0, P).

magic_vp(Subj,Comps,bring(john,flowers,mary), P0, P1).

```

Figure 3

Application of magic transformation to vp/4 from Figure 1.

erarchies (cf. Fall 1996, Wintner and Francez 1995). Coupling Minnen's techniques with compilation methods for typed-feature hierarchies should provide the necessary mechanisms for efficient implementation of large HPSG grammars.

A fair amount of work still remains: For an automated grammar-compilation system, it is essential that as much of the control information be extracted automatically as possible. Minnen's work, however, falls short of achieving this objective. His top-down processing strategy employing literal transformation and BRR transformation comes close to being a fully automated strategy, but I suspect that space requirements from tabulation becomes a factor, and hence (heuristic) techniques for making tabulation decisions need to be explored before the approach can be made practical. Somewhat surprisingly, Minnen does not explore the behavior of other variations on the basic top-down strategy, such as *deterministic closure*. It was not clear to me why the BRR transformation could not be used along with top-down control.

To ensure termination, Minnen's bottom-up control requires additional user-supplied control information in the form of parse types and delay patterns, which is not very desirable. Either automated generation of such control information or methods to eliminate it are needed. Although it is clear that Minnen's transformation techniques apply to typed-feature-structure grammars, there are hardly any examples of this in the monograph. Results of evaluation on a realistic grammar are only glossed over or missing, making it impossible to assess performance on large HPSG grammars such as the LinGo grammar (Copestake and Flickinger 2000).

References

- Copestake, Ann and Dan Flickinger. 2000. An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the Second Linguistic Resources and Evaluation Conference*, Athens, Greece, pages 591–600.
- Fall, Andrew. 1996. *Reasoning with Taxonomies*. Ph.D. dissertation, Department of Computer Science, Simon Fraser University, July 1996.
- Meurers, Detmar and Guido Minnen. 1997. A computational treatment of lexical rules in HPSG as covariation in lexical entries. *Computational Linguistics*, 23(4):543–568.
- Pollard, Carl and Ivan Andrew Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press and Stanford: CSLI Publications.
- Ramakrishnan, Raghu. 1991. Magic templates: A spellbinding approach to logic programs. *Journal of Logic Programming*, 11:189–216.
- Strzalkowski, Tomek, editor. 1994. *Reversible Grammar in Natural Language Processing*. Kluwer Academic.
- Wintner, Shuly and Nissim Francez. 1995. An abstract machine for typed feature structures. In *Proceedings of the Fifth Workshop on Natural Language Understanding and Logic Programming*, Lisbon, pages 205–220.

Suresh Manandhar is a lecturer in Computer Science at the University of York. He has worked on the implementation and formalization of constraint-based grammars. His published work includes papers on constraint logics and efficient compilation methods for constraint-based grammars, unsupervised learning of categorial grammars, learning of WordNet relations, and applications of inductive logic programming to natural language processing. Manandhar's address is: Department of Computer Science, University of York, Heslington YO1 5DD, York, U.K.; e-mail: suresh@cs.york.ac.uk.

The Language of Word Meaning

Pierrette Bouillon and Federica Bosa (editors)

(University of Geneva and LingoMotors Inc.)

Cambridge University Press (Studies in natural language processing, edited by Branimir Boguraev), 2001, xvi+387 pp; hardbound, ISBN 0-521-78048-9, \$69.95

Reviewed by

Mari Broman Olsen

Microsoft Corporation

1. Introduction

The papers collected in *The Language of Word Meaning* resemble nothing so much as a holiday celebration in a large, heterogeneous family, with echoes of old feuds, marginally relevant contributions from distant relatives, and fresh stories from recent friends. Although the feuds are entertaining (opening the festivities as Section I, “Linguistic Creativity and the Lexicon”), and the outsider perspectives insightful, from a computational-linguistic perspective many of the most valuable contributions come from guests who have traveled far, and with other companions, before finding common ground at this gathering.

The volume’s title conceals the specificity of its subject: a toast and roast of Generative Lexicon theory (henceforth GL), originally proposed by James Pustejovsky (henceforth JP) (Pustejovsky 1993, 1995, inter alia), who contributed two chapters and the preface. The editors also contribute a joint article, and Bosa is coauthor in a second. Others, certainly, would be included in a general discussion on “the language of word meaning.” Indeed (with rare and sometimes confusing exceptions) the discussion assumes GL as the “semantic vocabulary” (p. xv) and focuses on the converse—“the word meaning of language”¹—specifically, how and whether GL’s finite number of generative devices and rules can be used to construct semantic expressions compositionally. Several contributors (Fodor and Lepore, papers in Section II; Vossen, p. 373) raise interesting ancillary questions: how (and whether) to divide word meaning from world knowledge.

Pustejovsky (Chapter 7) and Bosa and Bouillon provide especially clear reviews of basic GL concepts. These (and other) surveys would be more useful, however, as a condensed introduction for newcomers (which this review will present with great reduction as part of Section 2). The book claims (according to JP’s preface) that the GL approach to language synthesizes traditions and ideas from ordinary language philosophy (a focus on words and word use), analytic semantics (formalization of rules and types), and generative linguistics (infinite generation of meanings from finite, recursive resources). It claims that GL can tackle empirically difficult problems, both for computational applications and for other formal, even descriptive, systems: specifically, how words vary in context, how new senses emerge, and what the underpinnings are of the systematic mapping of semantic types onto syntactic forms across languages.

¹ Here I’m agreeing with my six-year-old that this would “make more sense” as the title.

The editors begin the volume and each of four sections with short introductions (somewhat confusingly numbered as chapters). The introductions bring harmony and clarity to the four sections (except in Section III, the introduction to which takes issue with Kilgarriff's contribution). Nevertheless, the sections do not clearly form a whole, beyond their family relation to GL, computational lexicography, and/or lexical semantics. JP points out that not all contributors even agree with the generativity of lexical senses and composition. Some who do, but in a different framework (Moravcsik) provide explicit mappings from their representation to GL; others (Asher and Lascarides, Hobbs) leave relationships implicit. Several papers focus on practical natural language processing (NLP) (Section IV and others) and a few on questions of philosophical reality, which, though an interesting diversion, leads to further diffusion of focus.

One frequently wants more explication of criteria. When are new subrelations or attribute values required (cf. Ruimy, Gola, and Monachini)? Why is *book* a physical object in *I am waiting for her next book* (p. 137)? Why does *waiting for the car* anticipate a construction event, but not *repairing the car, washing the car, or waiting for the car to be returned [by a borrower]* (p. 159)? How do we know when senses are different and require different generative mechanisms (p. 367)? Why are some predicted senses rare (pp. 153–154)?

In some chapters, editorial lapses and typos are frequent enough to be distracting: non-English *finish to read* for *finish reading*, and *on one side* for *on one hand*; Schank and Abelson cited as 1997 instead of 1977; unexpanded acronyms (*LCS* by Saint-Dizier, *AFT* by Moravcsik) and multiply expanded acronyms (*JP* by Fodor and Lepore); and a missing abstract (Moravcsik). A few sentences are uninterpretable (the sentence in Yorick Wilks's abstract beginning "It is argued . . ."; a couple of sentences by Piek Vossen (p. 36) on "disjunctive" labels). A shared bibliography would have improved the book's value as a GL reference and reduced the number of pages (all but one paper cites Pustejovsky 1995, for example).

2. Generative Lexicon Theory

GL employs four levels of linguistic representation. ARGUMENT STRUCTURE encodes obligatory and optional arity of predicates. EVENT STRUCTURE describes the aspectual nature of the events, whether states, processes, or transitions (cf. Mourelatos [1978], *inter alia*), and their subevents, if any, as well as the temporal relations between subevents. QUALIA STRUCTURE links the first two representation levels by assigning arguments to participant slots in (sub)events. The four qualia roles, inspired by Moravcsik's interpretations of Aristotle (p. 56), minimally encode the linguistic behavior of lexical items, primarily nominals and predicates with which they compose. The CONSTITUTIVE role of an object represents the relation between it and its parts (and what it is part of, according to Busa and Bouillon), the FORMAL role categorizes it within a larger domain, the TELIC role represents its purpose, and the AGENTIVE role its origins. Finally, LEXICAL INHERITANCE STRUCTURE relates lexical structures to other structures in the type lattice. Generative devices, including type coercion, subselection, and co-composition, connect all four levels in restricted ways, providing for compositional interpretation of words in context (p. 56).

Qualia structure work has proved especially fertile for lexical semantic research, including many clear, data-rich papers in this volume. The qualia roles challenge researchers to consider how far a minimal lexical representation can go toward accounting for semantic productivity, much of which had been assigned to world knowledge. For example, JP suggests (Pustejovsky 1993) that both possible meanings for *I began the book* ('began to read/write') derive from qualia structure, one from the AGENTIVE

(books come into being by writing), and one from the TELIC (books are intended to be read).

Julius M. Moravcsik aligns his fourfold structure (m-factor, s-factor, a-factor, and f-factor) with the four qualia roles above, respectively, to partially decompose metaphors. Salvador Climent argues that the CONSTITUTIVE role accounts for the behavior of Spanish partitives; Patrick Saint-Dizier augments the TELIC role with a small set of rules relating predicates and arguments for more explanatory representation of adjectival modification, as well as some sense variations, metaphors, and metonymies; Jacques Jayez finds qualia structure inadequate to account for the distributional properties of several French synonym classes—for example, to encode the idea that *suggérer* ‘suggest’ permits both animate and inanimate subjects, but the former only if the complement is a “choice” among items in the propositional context. Adam Kilgarriff concludes from the SENSEVAL data (Kilgarriff and Palmer 2000) that qualia structure fails to account for many senses (perhaps metaphorical) not encoded in dictionaries, for which GL would seem to be naturally suited.

3. Section I: Family Feuds

The first section, “Linguistic Creativity and the Lexicon,” considers whether words have internal syntax and inferential meaning (as in GL) or whether they are concepts as particulars with denotational semantics. James McGilvray, in “Chomsky on the Creative Aspect of Language Use and Its Implications for Lexical Semantic Studies,” situates GL as Relation R in Chomskyan syntax, as an internalist semantics, “a branch of syntax, broadly speaking” (p. 19). His leap from the usefulness to the innateness of qualia structure (p. 17) is jarring, as is the sudden appearance of minimalism 14 pages into the 23-page chapter (p. 18). But he clearly and cogently defends Chomsky as grounding a computational approach: obvious to some (p. xii), yet hotly debated in some computational circles.

As they have in the past,² Jerry A. Fodor and Ernie Lepore, in “The Emptiness of the Lexicon: Critical Reflections on J. Pustejovsky’s ‘The Generative Lexicon’”, challenge JP on the possibility of the GL enterprise (or indeed any nonatomic representation structure). Fodor and Lepore (henceforth F&L) lack faith that lexical generalizations are ultimately interesting or powerful and not merely one of “all sorts of ways” in which “the meanings of words can partially overlap” (p. 48). Claiming that the null hypothesis (p. 29) is as “everybody always thought: the lexicons of natural languages are just lists” (p. 44),³ they question whether GL reaches its aims: to account for semantic well-formedness and generativity and to provide evidence that meaning is inferential (rather than atomistic and denotative). JP responds (“Generativity and Explanation in Semantics: A Reply to Fodor and Lepore”) that F&L misread him in substance and detail and are “negative and unconstructive” (p. 51), artificially raising the bar for semantic theory while avoiding the theoretical issues raised by the data. According to F&L, GL is descriptively inadequate, both over- and undergenerating interpretations. JP claims that “the modes of inheritance for concepts associated with linguistic expressions are not arbitrary” (p. 61). F&L say those given are incomplete. F&L’s alternative (*want a beer* → *want to have a beer* → *want to drink a beer* derived by

² See, for example, discussion between Fodor and Weinberg in CUNY (1998), as well as references in the papers by Fodor and Lepore and Wilks in the volume under review.

³ Although, as F&L point out, much effort has been expended studying semantic relations among lexical items, for example, synonymy and antonymy, which are phenomena that they foresee will turn out also to be conceptual rather than lexical.

logical-form rules) should be subjected to the same adequacy standard to which they subject the GL discussion of *bake* (on which JP says, in a footnote on p. 64, that they miss the point anyway).

Wilks stands with JP in “The ‘Fodor’-FODOR Fallacy Bites Back,”⁴ rejecting the F&L position as untenable for analytic reasons (there is no one-to-one mapping between words and meanings) and unhelpful in organizing a useful dictionary, for NLP or otherwise. Although F&L don’t claim “real” representations should be useful, Wilks also disagrees with F&L and JP that an adequate theory of the lexicon should be able to judge semantic well-formedness.

4. Section II: Finding Common Ground

The papers in the second section, “The Syntax of Word Meaning,” refine and gently challenge the vocabulary, framework, and crosslinguistic data analysis required for an explanatorily adequate, language-independent lexical semantics. JP, in “Type Construction and the Logic of Concepts,” a revision of a 1998 paper, discusses well-formedness directly and convincingly, arguing that constraints on concepts (i.e., thought) are revealed through lexicalization strategies. He proposes a tripartite-concept subtyping—natural, functional, and complex, discussed further in Section IV by Busa, Calzolari, and Lenci—and addresses criteria for choosing among competing representations and deciding about feature admission, appropriateness, and composition.

In “Underspecification, Context Selection, and Generativity,” Jayez shows that context provides cues to polysemous word meaning as well as imposing restrictions on synonym selection (French *suggérer* and *attendre* ‘wait’ classes). It is not clear that the data presented are inherently incompatible with the GL framework, as he argues. He also makes the general point that GL practitioners must demonstrate generative adequacy not only against sense enumeration lexicons (the current focus in the literature) but also against other decompositional and underspecified lexical theories.

The next two papers propose extensions to qualia structure. For Bouillon and Busa, in “Qualia and the Structuring of Verb Meaning,” qualia roles are dynamic, varying by speaker and context; for example, the constitutive relation has subtypes (IS PART OF, HAS AS PART, IS IN, LIVES IN, IS A MEMBER OF, HAS AS MEMBER). They argue that the restrictions on the variety of noun phrases that can occur as objects of *attendre* can be derived compositionally: The verb selects an event, and some nominals encode events, for example, creation (such as *journal*, *book*, *car*), possession (all nominals with telic roles, inter alia), and beginning or culminating. Saint-Dizier (“Sense Variation and Lexical Semantics: Generative Operations”) augments the telic role, the “most productive role to novel sense variations” (p. 168).

In “Individuation by Partitive Constructions in Spanish,” Climent argues that the semantics of portions is grounded in the entailment that entities (individuated, mass, etc.) are composed of something. He argues that qualia are syntactic, with compositional operations creating selectionally constrained complex nominals. Through interesting (though constructed) data, in “Event Coreference in Causal Discourses,” Laurence Danlos argues persuasively that causatives can be interpreted from a linguistic perspective with internally complex events, although she rejects the extended event structure for unaccusatives.

⁴ Wilks says F&L claim to be part of informational role semantics (IRS), as he is. This is confusing, since in their chapter they express doubt that IRS can be sustained, challenge the cogency of arguments that take it as a premise, and go on to address JP, who purports to provide an argument for the complexity of lexical entries that do not presuppose IRS (p. 29).

5. Section III: Forbears and New Friends

The gratingly titled third section, "Interfacing the Lexicon" (with what?), examines metonymy and metaphor in the GL framework. The stimulating studies in this section decompose structures generally considered lexically atomic. In "Metaphor, Creative Understanding, and the Generative Lexicon," Moravcsik uses GL's polysemy account (whose qualia spring, in part, from Moravcsik's work) to uncover the lexical information structure by means of metaphor. Moravcsik argues that simile also can be treated "as a literal statement" (p. 260) of comparison.

Nicholas Asher and Alex Lascarides examine "Metaphor in Discourse" in terms of lexical semantic structures, inspired by GL and others (Copestake and Briscoe 1995), focusing on change-of-location verbs, metaphorical extension of properties from objects to persons, and the relation between metaphoric interpretation and discourse structure. They constrain interpretations by means of lexical rules that restrict meaning shifts and a discourse structure theory in which truth conditions can be evaluated.

Jerry Hobbs presents his interpretation-as-abduction framework in "Syntax and Metonymy." He accounts for Nunberg's (1995) metonymic data (*I'm parked out back*) by means of a set of coercion relations, selected on the basis of saliency by least-cost abductive interpretation. He extends his account to a wide range of intriguing data, old and new, including extraposed modifiers (*A jolly old man arrived with an armload of presents*), Bolinger's (1988) ataxis (*John smokes an occasional cigarette = John occasionally smokes cigarettes*), container nouns (cf. Jayez), and disguised small clauses (*I have a sore throat = My throat is sore*).

Kilgarriff, in "Generative Lexicon Meets Corpus Data: The Case of Nonstandard Word Uses," brings to the discussion a refreshing quantitative, corpus linguistics perspective. He asks to what extent GL is able to generate, for a sample set of words, senses found in corpora but not represented in a dictionary ("nonstandard" senses). As mentioned above, he argues that contextual sense variation is too large and unpredictable for such a model of lexical structures. Some might question his strategy for determining nonstandard senses and his rigorous interpretation of GL mechanisms, but his methodology is clear, reproducible, and a fair challenge to GL, potentially accounted for by Bouillon and Busa's expansion of the qualia relations, for example.

6. Section IV: Practical Matters

The fourth section, "Building Resources," chronicles practical experiences for a range of NLP applications within the SIMPLE and EuroWordNet frameworks. The papers in this section, as the introduction states, leave open the question of whether they are useful or usable; they furthermore neglect to evaluate whether they are practical to construct: Most involve significant manual work.

In "Generative Lexicon and the SIMPLE Model: Developing Semantic Resources for NLP," Busa, Nicoletta Calzolari, and Alessandro Lenci highlight the need to capture richness of language (apparently independent of applications) within a testable model for building large-scale resources. They employ qualia roles for structuring the semantic/conceptual types, assuming that words vary in complexity (expanding on Pustejovsky in Section II). Nilda Ruimy, Elisabetta Gola, and Monica Monachini, in "Lexicography Informs Lexical Semantics: The SIMPLE Experience," describe how they use GL to categorize concepts, without defending or critiquing the theory or their particular attribute-value assignments. They argue that encoding their methodology serves to explicate the theory.

Piek Vossen describes his work on “Condensed Meaning in EuroWordNet,” focusing on the challenge of developing GL types of relations from resources that are sense-enumerative. He sets forth how EuroWordNet links semantic representations across languages via an Inter-Lingual-Index that connects their structures.

7. Conclusion

The Language of Word Meaning offers glimpses into the variety of work spawned by GL: philosophical, computational, theoretical, and multilingual. Section introductions and GL overviews by Pustejovsky (Chapter 7) and Busa and Bouillon lay out important themes in computational and theoretical lexical semantics. Other chapters offer practical, often entertaining exemplars of lexical work unfolding—like family celebrations—somehow, noisily, and with great diversity.

References

- Bolinger, Dwight. 1988. Ataxis. In Rokko Linguistic Society, editors, *Gendai no Gengo Kenkyu* [Linguistics Today], Kinseido, Tokyo, 1–17.
- Copestake, Ann and Ted Briscoe. 1995. Semi-productive polysemy and sense extension. *Journal of Semantics*, 12(1):15–67. Reprinted in James Pustejovsky and Branimir Boguraev, editors, *Lexical Semantics: The Problem of Polysemy*, Oxford University Press, 1996, 15–67.
- CUNY. 1998. *Proceedings of the 11th CUNY Conference on Human Sentence Processing*, Rutgers University, New Brunswick, NJ.
- Fodor, Jerrold. 1970. Three reasons for not deriving *kill* from *cause to die*. *Linguistic Inquiry*, 1:429–438.
- Kilgarriff, Adam and Martha Palmer, editors. 2000. Special Issue on SENSEVAL: Evaluating Word Sense Disambiguation Programs. *Computers and the Humanities*, 34.
- Mourelatos, Alexander. 1978. Events, processes, and states. *Linguistics and Philosophy*, 2:415–434. Also published in Philip Tedeschi and Annie Zaenen, editors, *Syntax and Semantics 14: Tense and Aspect*, Academic Press, 1981, 191–212.
- Nunberg, Geoffrey. 1995. Transfers of meaning. *Journal of Semantics*, 12:109–132.
- Pustejovsky, James. 1993. Type coercion and lexical selection. In James Pustejovsky, editor, *Semantics and the Lexicon*. Kluwer Academic, pages 73–96.
- Pustejovsky, James. 1995. *The Generative Lexicon*. MIT Press.

Mari Broman Olsen has worked as lexical semanticist in Microsoft Corporation’s Natural Language Group since 1999. From 1996 to 1999, she was a postdoctoral fellow at the University of Maryland Institute for Advanced Computer Studies, researching interlingual lexical representation for Chinese–English machine translation, inter alia. Her Ph.D. is from Northwestern University (1994); her dissertation, *A Semantic and Pragmatic Model of Lexical and Grammatical Aspect*, was published by Garland Press. Olsen’s address is: Microsoft Corporation, One Microsoft Way, Redmond, WA 98052; e-mail: molsen@microsoft.com.

Empirical Methods for Exploiting Parallel Texts

I. Dan Melamed

(New York University)

Cambridge, MA: MIT Press, 2001,
xi+195 pp; hardbound, ISBN
0-262-13380-6, \$32.95, £22.95

Reviewed by

Ted Pedersen

University of Minnesota, Duluth

Parallel translations of written texts have long been useful tools for human students of language and have begun to serve as an intriguing source of data for corpus-based approaches to natural language processing. A source text and its translation can be viewed as a coarse map between the two languages, and an industrious student or clever computer program may wish to refine that mapping so that it shows which sentences, phrases, and words are translations of one another.

Humans are very adept at finding such relations in parallel text. This is true even when one or both of the languages is unfamiliar, as can be seen in a simple but convincing exercise by Knight (1997). Although there was considerable early success in automatically identifying sentences in parallel text that are translations of each other (e.g., Brown, Lai, and Mercer 1991, Gale and Church 1993), a variety of challenging problems has emerged since that time.

Empirical Methods for Exploiting Parallel Texts is a revision of author I. Dan Melamed's 1998 Ph.D. dissertation (University of Pennsylvania) and succeeds in capturing the range of problems inherent in parallel text. It presents a variety of techniques for finding translation equivalents and demonstrates that once these are available, they can be used to align text segments, detect omissions in translations, identify noncompositional compounds, and discriminate among word senses.

The book is arranged in three parts, the guiding organizational principle of which is the distinction between tokens and types. (A token represents each occurrence of a linguistic entity in a text, whereas a type consists of every occurrence of identical tokens.) The author casts his own work in terms of pattern recognition techniques that acquire information about specific tokens and statistical learning methods that induce generalized models of word types on the basis of token data.

Part I consists of three chapters that focus on tokens and pattern recognition. Chapter 2 presents an algorithm that finds a token-by-token mapping of a parallel text in which each token in the source text is aligned with its translation in the target text. This algorithm is presented in terms of pattern recognition concepts such as signal generation, filtering, and search.

The signals generated are candidate token alignments, and these come from cognates identified in the parallel text and optionally from a user-supplied seed lexicon of word translations. High-frequency words create noise that is filtered by means of a localized search procedure that allows the algorithm to consider the parallel text piece by piece as it performs token alignment. The next two chapters (3 and 4) show that once such a mapping is available, it can be used to find segment alignments in parallel text and to detect portions of text that have been erroneously omitted from a translation.

Part II is a two-chapter interlude that touches on both tokens and types. Chapter 5 is a discussion of the issues involved in counting co-occurrences in parallel text. Although Melamed's conclusions may seem fairly intuitive, he shows that previous work in translation modeling has often been based on less-than-optimal counting schemes. In Chapter 6, Melamed describes the creation of a manually word-aligned version of the Bible in French and English. This is a nontechnical chapter that captures the difficulties of manual alignment in general and in dealing with the Bible in particular.

The three chapters in Part III consider word types and statistical learning techniques. Chapter 7 develops three statistical models of translation that represent unordered word-by-word translation. These models rely on the observation that a source word will most often translate into a single target word (as opposed to multiple words) or may have no translation equivalent at all. Parameter estimation schemes are developed that capture these properties, and empirical results show that the inclusion of parameters developed according to these schemes steadily improves the translation models, all of which significantly outperform the historically important IBM Model 1 (Brown et al. 1993).

Melamed argues that word-by-word translation models can also be used to identify linguistic entities that are normally not translated word for word. In Chapter 8 he shows how noncompositional compounds such as *beat a dead horse* can be discovered in parallel text by means of such a model. In Chapter 9, an unsupervised word sense discrimination algorithm is introduced that is evaluated by treating the different discovered senses of a word as distinct types. The incorporation of these additional sense distinctions as types is shown to improve the quality of a translation model.

Several of the chapters in this book have appeared in preliminary form in this journal (Chapters 2, 3, and 7) and conference proceedings (Chapters 4 and 8). As a result, the individual chapters are well polished and clearly the product of careful reviewing and rewriting. Since this material retains its original style and form, each chapter is relatively self-contained, and readers may wish to approach this book as a collection of related papers rather than as a work that must be read from start to finish.

Readers not familiar with corpus-based methods, pattern recognition, and statistical learning may struggle at times. There is not a great deal of preliminary or background material presented in each chapter before moving into more advanced concepts. This is particularly true of Chapters 2 and 7, which present the token alignment algorithm and translation model development. This should not discourage the novice from approaching this book, however. Since each chapter can be read independently, it is possible to appreciate the related applications described in Chapters 3, 4, 8, and 9 without fully understanding all of the technical details in Chapters 2 and 7. The author also includes sufficient discussion of related work throughout to provide a general picture of the field.

To conclude, this is a useful book for anyone interested in parallel text. It offers enough improvements and clarifications over previously published material that it will appeal to a reader already familiar with the author's work. It also has a wide scope and strikes a reasonable balance between theory and application, so newcomers will get a sense of the possibilities and problems associated with parallel text. These readers may wish to supplement this book with pattern recognition and statistical-learning readings, as found in Duda and Hart (1973), for example. Finally, in an age in which academic texts tend to be expensive and poorly printed, this book is a notable exception, as the production quality is high and the price is reasonable.

References

- Brown, Peter F., Stephen Della Pietra, Vincent Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.
- Brown, Peter F., Jennifer Lai, and Robert L. Mercer. 1991. Aligning sentences in parallel corpora. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 169–176, Berkeley, CA.
- Duda, Richard O. and Peter E. Hart. 1973. *Pattern Classification and Scene Analysis*. Wiley.
- Gale, William and Kenneth Church. 1993. A program for aligning sentences in bilingual corpora. *Computational Linguistics*, 19(1):75–102.
- Knight, Kevin. 1997. Automating knowledge acquisition for machine translation. *AI Magazine*, 18(4):81–96.

Ted Pedersen is an Assistant Professor of Computer Science at the University of Minnesota, Duluth. His research focuses on word sense disambiguation. He is the recipient of a National Science Foundation CAREER award. Pedersen's address is: Department of Computer Science, University of Minnesota, Duluth, MN 55812; e-mail: tpederse@d.umn.edu.

Sentence Comprehension: The Integration of Habits and Rules

David J. Townsend and Thomas G. Bever

(Montclair State University and University of Arizona)

Cambridge, MA: MIT Press (Language, speech, and communication series), 2001, xi+445 pp; hardbound, ISBN 0-262-20132-1, \$65.00, £44.95; paperbound, ISBN 0-262-70080-8, \$24.95, £16.00

Reviewed by

Matthew W. Crocker

Saarland University

The last two decades have witnessed a great deal of activity in the development of computational theories of human sentence processing. The aim of such research is to explicitly identify and model the architectures and mechanisms that underlie human comprehension (see Crocker, Pickering, and Clifton [2000] for a recent collection on the topic). Although there is often relatively little interaction between the *computational psycholinguistics* and *natural language processing* (NLP) communities, both have been concerned with how various symbolic, connectionist, probabilistic, and hybrid techniques can be applied to the task of recovering interpretations of ambiguous linguistic input according to some set of underlying linguistic rules and principles. The difference between the two enterprises concerns their respective goals. Models of human sentence comprehension are concerned with replicating human behavior during the on-line, incremental (word-by-word) processing of an utterance as it unfolds. Particular emphasis is therefore placed on the resolution of local or temporary ambiguities, which occur as each initial substring of an utterance is encountered. Identifying human preferences in the face of local ambiguity is not a goal in itself but provides critical insight into the kinds of mechanisms (parallel, serial, underspecified) and information sources that are used (e.g., syntactic, semantic, visual, probabilistic). Traditional NLP systems, on the other hand, are typically concerned with trying to resolve global ambiguities (language models for speech recognition being a notable exception) and often strive only to recover partial interpretations (e.g., keyword finding, chunking, template filling). Interestingly, the work reviewed here can be seen as arguing for greater convergence between psychological and technical models in that it argues for the existence of a shallow, partial parser within the human sentence-processing mechanism.

Sentence Comprehension: The Integration of Habits and Rules is an ambitious work that aims to integrate ideas and evidence from sentence-processing research that have accrued over the last 40 years. This is essentially two books in one: an introductory text and a research monograph. The need for these two subbooks arises from the highly interdisciplinary nature of the topic: The development of an algorithmic model of human sentence processing based on state-of-the-art linguistic theory necessarily draws upon background in linguistics, computational linguistics, and psycholinguistics. The first 150 pages of the book are devoted to providing this. After a brief introductory chapter sketching the perspective and goals of the book, we are presented with a rather historical review of early theories of sentence processing (Chapter 2) and an introduction to some of the key aspects of current linguistic theory (Chapter 3). The latter concludes

with an overview of minimalist syntactic theory (Chomsky 1995), which the authors later assume in developing their own model.

Of these introductory chapters, Chapter 4 is worthy of particular mention. It gives a truly comprehensive and relatively detailed review of the prevailing theories of sentence processing. Not only are the models well presented, but similarities, differences, advantages, and disadvantages are also clearly identified. The chapter begins with the deterministic parser of Marcus (1980) and Frazier's Garden Path theory (1978) and moves on through principle-based parsing and reanalysis (e.g., Pritchett 1992; Crocker 1996). The final half of the chapter then considers more-recent probabilistic models, which I loosely take to encompass statistical parsing (Jurafsky 1996), connectionist, and competitive constraint-based approaches (Spivey and Tanenhaus 1998) as well as hybrid models (Stevenson 1994). This part of the book is particularly interesting in revealing how probabilistic mechanisms can be used in extremely different ways and how such models face fundamental challenges, for example, in establishing the "level of grain" at which statistical frequency information is accrued (Mitchell et al. 1995), and the scalability of recent connectionist proposals (Spivey and Tanenhaus 1998).

We now turn our attention to the model proposed by the authors, which is introduced in Chapter 5. The central hypothesis of the book, and core idea of the model, is that sentence comprehension involves associative and syntactic components. An associative component is assumed to be necessary for rapid, shallow recognition of major phrases. Configurational surface schemata, or *canonical sentence templates*, and verb-frame information are then exploited to construct a preliminary conceptual representation. The syntactic component, in turn, is responsible for recovering the full cyclic structure of the sentence, including the longer-distance relationships introduced by movement transformations. The syntactic component takes the shallow "quick and dirty" analysis as its input, derives a complete syntactic parse, and then checks that this is still consistent with the input string. The result is the so-called analysis-by-synthesis model.

This two-tiered model of processing is broadly motivated by an apparent duality in the evidence from human language processing. On the one hand it is clear that people are sensitive to certain kinds of probabilistic information in building interpretations of ambiguous linguistic input and even in generating predictions of what is to come next (and recovering material that was unclear in the original, noisy input signal). On the other hand, people clearly demonstrate the ability to apply rich and productive linguistic principles in dealing with complex and novel input. The authors argue that trying to recover the interpretation of a sentence directly by means of the grammar is intractable. In particular, they argue that reversing the transformational derivations of the grammar incrementally would simply not be possible, given the recursive complexity of language. The idea is that the associative component "isolates a candidate analysis based on surface cues, and syntactic processes check that the analysis can derive the sequence" (p. 164, para. 1). Since full parsing follows an initial syntactic and semantic analysis, it is argued that the tremendous search space required for full parsing can be substantially reduced.

The following chapters further develop the *Late Assignment of Syntax Theory* (LAST). Chapter 6 presents a wide range of empirical data, which are interpreted as supporting the model, and also takes the opportunity to introduce a range of experimental psycholinguistic techniques for the uninitiated. Chapter 7 delves into greater detail concerning the associative component and the use of canonical sentence templates in particular. Chapter 8 then situates the model of sentence processing within language processing more generally (e.g., discourse and conceptual understanding). Finally, Chapter 9 considers a range of other, more peripheral issues, such as evidence

from language acquisition and neurological support for the model. Chapter 10 gives a concise summary of the theory and its motivation and some further rationalization.

To their credit (and the benefit of the seriously interested reader), these chapters cover a tremendous amount of psycholinguistic data in building support for the authors' proposals. It must be accepted, of course, that empirical *facts* are typically consistent with many possible *interpretations*, and it will be up to readers to decide whether they are persuaded by the authors' attempt to interpret the facts presented as supporting the LAST model. One serious criticism of the empirical discussion, however, is that the book considers only evidence from English. Sentence processing has seen an increasing emphasis on crosslinguistic research over the last 10 years, as it became evident that earlier models were "overfitted" to English. LAST seems to fall prey to this in its heavy reliance on two properties of English, namely that (1) verbs precede their complements and (2) clauses have a relatively fixed word order. The former is essential for the prediction of complements on the basis of the verb's most frequent subcategorization frame (which the authors propose guides the associative component), and the latter seems crucial in order for *canonical sentence templates* to have any substantive value. An additional (and important) consequence of the model is that it sacrifices full, word-by-word incremental interpretation. Although the authors attempt to justify this move, it is certainly at odds with current consensus in the sentence-processing community. Experimental evidence suggests that people can and do rapidly integrate the full complement of linguistic constraints that follow from a left context in processing the words that follow. I remain unconvinced that the associative stage of the LAST model would be able to account for the full range of facts here.

Although the theory's ability to explain the empirical phenomena it is intended to elucidate will be crucial to its ultimate success, it is worth making some observations on the theory itself. It has to be said that, despite its being couched in terms of modern minimalist linguistic theory, and despite all the attention paid in the book to the full complement of modern models (symbolic, probabilistic, and connectionist), the proposed model has a decidedly "retro" feel to it. In particular, it resurrects notions like the old N-V-N (agent-action-patient) canonical sentence templates (Fodor, Bever, and Garrett 1974), not to mention something closely resembling the transformational theory of grammar, in that the syntactic component is assumed to perform some direct form of reverse transformation operations. The rather literal interpretation of transformations has all but been abandoned even in principle-based models (Crocker 1996). I was also disappointed to discover that the claimed reconciliation and "integration of symbolic-computational and associative-connectionist" approaches amounted to simply having one of each. Although this is a possible position to take, it is certainly at odds with Occam's razor to have two mechanisms doing essentially the same task. A theory that can explain the evidence with a single mechanism would surely be preferable. Whereas earlier models indeed failed to account adequately for both symbolic and probabilistic behavior, recent models have shown that they can be integrated by applying ideas from probabilistic language modeling, such as hidden Markov models and stochastic parsing (Jurafsky 1996; Crocker and Brants 2000).

Furthermore, although the model might be considered algorithmic, it has not been implemented computationally, in contrast with many competing models. Given the complexity of human language comprehension and the range of interacting mechanisms and information sources, it is difficult to ensure that the theory is being invoked consistently and uniformly in accounting for the full range of empirical findings. This potential pitfall seems especially important for LAST given that two almost autonomous mechanisms are involved. The only way to achieve the necessary consistency in evaluating a model, and to make clear predictions, is through implementation.

To conclude, *Sentence Comprehension* functions as an excellent introduction to the history of sentence processing, the methods used, and prevailing theories. It strives to identify both the truths that have been revealed and the outstanding conundrums and presents evidence and concepts in both a detailed and comprehensible fashion. The book is well written and organized, including a detailed bibliography, as well as subject and author indices. Other psycholinguists may not always be convinced by the authors' interpretation of competing theories and empirical data, but that will always be the case. This book nonetheless provides exceptional coverage and would function as a thorough introduction for newcomers and a valued resource to those already in the field. It could certainly be used as a text, although, given the emphasis on what I take to be a rather controversial proposal, I would probably prefer to use it alongside other material. Despite the criticisms detailed above, the authors' proposals certainly offer numerous insights and put into sharp focus the importance of considering probabilistic, experience-based aspects of language use as well as general linguistic principles.

References

- Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press.
- Crocker, Matthew W. 1996. *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language*. Kluwer Academic.
- Crocker, Matthew W. and Thorsten Brants. 2000. Wide coverage probabilistic sentence processing. *Journal of Psycholinguistic Research*, 29(6):647–669.
- Crocker, Matthew W., Martin Pickering, and Charles Clifton Jr. 2000. *Architectures and Mechanisms for Language Processing*. Cambridge University Press.
- Fodor, Jerry A., Thomas G. Bever, and Merrill F. Garrett. 1974. *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. New York: McGraw-Hill.
- Frazier, Lyn. 1978. *On Comprehending Sentences: Syntactic Parsing Strategies*. Ph.D. dissertation, University of Connecticut. Published by the Indiana University Linguistics Club, Bloomington, IN.
- Jurafsky, Daniel A. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20:137–194.
- Marcus, Mitchell P. 1980. *A Theory of Syntactic Recognition for Natural Language*. MIT Press.
- Mitchell, Don C., Fernando Cuetos, Martin M. B. Corley, and Marc Brysbaert. 1995. Exposure based models of human parsing: Evidence for the use of coarse-grained (non-lexical) statistical records. *Journal of Psycholinguistic Research*, 24:469–488.
- Pritchett, Bradley L. 1992. *Grammatical Competence and Parsing Performance*. University of Chicago Press.
- Spivey, Michael and Michael Tanenhaus. 1998. Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24:1521–1543.
- Stevenson, Suzanne. 1994. Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research*, 23:295–322.

Matthew W. Crocker is a Professor of Psycholinguistics in the Department of Computational Linguistics, Saarland University. His research is concerned with computational models of human sentence processing, focusing in particular on the development of probabilistic models of sentence comprehension, statistical language processing, and experimental psycholinguistics. Crocker's address is: Department of Computational Linguistics, Building 17, Saarland University, 66041 Saarbrücken, Germany; e-mail: crocker@coli.uni-sb.de; <http://www.coli.uni-sb.de/~crocker/>.