interdisciplinary approach to the analysis of spatial prepositions. The result is an intriguing, sometimes incomplete, often speculative account of how language and space are related. Often the guiding philosophical maxims are contradicted by assumptions made later in the book. For example, the ideal meanings of words often appear to be Platonic objects, something obviously inconsistent with the tenets in the preface of the book. Elsewhere, the geometric description functions seem to embody mental representations, a view not comfortably condoned by a neo-Heideggerian analysis. In general, however, the work is an enjoyable and well-written foray into a very basic subset of language.

References

Appelt, D. 1985 *Planning English sentences*. Cambridge, England: Cambridge University Press.

Carnap, Rudolf. 1967 *The logical structure of the world*. University of California Press.

Charniak, Eugene 1983 Passing markers: A theory of contextual influence in language comprehension. *Cognitive science* 7(3): 171–190.

Dreyfus, Hubert 1979 *What computers can't do*. Revised edition. Harper Colophon Books.

Fass, Dan 1986 *Collative semantics*. Technical report MCCS–86–56, Computing Research Laboratory, New Mexico State University.

Goodman, Nelson 1951 *The structure of appearance*. Dordrecht, The Netherlands: Reidel.

Halliday, M.A.K. 1985 *A short introduction to functional grammar*. London: Arnold.

Hayes, Patrick 1985 Naive physics: Ontology for liquids. In *Formal theories of the commonsense world*, J. Hobbs and R. Moore (eds.), Ablex.

Hirst, Graeme 1987 *Semantic interpretation and the resolution of ambiguity*. Cambridge, England: Cambridge University Press.

Jackendoff, Ray 1983 *Semantics and cognition*. Cambridge, MA: The MIT Press.

Lakoff, George 1987 *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

McDonald, David 1987 Natural language generation. In *Encyclopedia of artificial intelligence*, S. Shapiro (ed.), New York: Wiley.

Moravcsik, Julius 1981 How do words get their meanings? *The journal of philosophy* 78:(1).

Rieger, C.J. and S. Small 1979 *Word expert parsing*. Technical Report TR-734, Department of Computer Science, University of Maryland.

Rosch, E. 1977 Human categorization. In *Advances in cross-cultural psychology*, vol. 1, N. Warren (ed.) 1–49, London: Academic Press.

Schlick, Moritz 1918 Allgemeine Erkentnisslehre, Berlin.

Searle, John 1979 Literal Meaning. In *Expression and Meaning*, Cambridge, England: Cambridge University Press.

Talmy, Len 1978 Figure and ground in complex sentences. In *Universals of human language* 4, J. Greenberg et al (eds.) Stanford, CA: Stanford University Press.

Tarski, A. 1943 The semantic conception of truth. *Philosophy and phenomenological research* 4: 341–375.

Wilks, Yorick 1975 A preferential pattern-seeking semantics for natural language inference. *Artificial Intelligence* 6: 53–74.

Winograd, Terry 1980 What does it mean to understand language? *Cognitive science* 4: 209–241.

Winograd, Terry and Fernando Flores 1986 *Understanding computers and cognition*. Norwood, NJ: Ablex and Reading, MA: Addison-Wesley.

Wittgenstein, Ludwig 1963 *Philosophical investigations* (translated by G. E. Anscombe). Oxford: Blackwell.

James Pustejovsky is the author of several papers on language generation and lexical semantics. His address is: Department of Computer Science, Brandeis University, Waltham, MA 02254. E-mail: jamesp@brandeis.edu

---

The Form of Information in Science: Analysis of an Immunology Sublanguage

Zellig Harris, Michael Gottfried, Thomas Ryckman, Paul Mattick Jr., Anne Daladier, T. N. Harris, and S. Harris

Dordrecht: Kluwer Academic Publishers, 1989, xiii + 590 pp.
(Boston Studies in the Philosophy of Science 104)
ISBN 90-277-2516-0; $124.00 (hb)

Reviewed by
Stephen B. Johnson
Columbia University

There is wide agreement that domain knowledge plays an important role in natural language processing systems, and, at the same time, that acquisition of domain knowledge is an extremely difficult problem. The work under review offers a rigorous method for knowledge acquisition in scientific and technical domains, based on a formal analysis of the texts written by domain experts. The set of texts in a restricted domain is known as a **sublanguage**. The method, which may be termed **sublanguage analysis**, reveals a formal structure in the sentences of the texts, **sublanguage formulas**, which are similar to the formulas of logic, but with certain extensions (which will be described below).

The sublanguage formulas described by the authors constitute a form of knowledge representation, and suggest interesting possibilities for the design of flexible and expressive databases or knowledge bases. The strength of the sublanguage approach lies in basing the knowledge representation on the analysis of actual texts. The significance of this approach to computational linguistics is that the initial phase of sublanguage analysis establishes a direct relationship between surface sentence forms and the semantic representation (formulas). This mapping serves as a basic design for text processing algorithms.

A striking feature of the book is that the authors have carried out a thorough test of their technique on real data: 14 full-length research articles from the field of immunology, published in the period 1935–1970. The formulas obtained and the methods used in producing them are given in meticulous detail. (The appendices that give examples of the formulas actually exceed the length of the narrative portion of the book). The methods employed are founded on Operator Grammar (Harris 1982) and are carried out in a general theoretical framework that portrays the organization of information in natural language (Harris 1988).

A second feature that makes this book a rarity is that the analysis of the immunology texts is presented as an "experiment" with a testable hypothesis: Do the results (sequences of formulas) obtained by objective analysis of the immunology articles correlate with known changes in knowledge of the field during the given period? The documents were selected for the study by immunologists (authors T. N. Harris and S. Harris), on the basis of the historical coverage of this period of immunological research. Confirmation of the correlation is provided by directly comparing the formulas (appendices 1 and 2) to the historical discussion given by the immunologists (Chapter 8), and by the discussion in the first two sections of Chapter 3.

A key aspect of the sublanguage method is that it is objective, relying only on structural features of texts and not ad hoc semantic judgements. This property insures that the analysis is repeatable, and the authors demonstrate this fact by performing an independent analysis of French immunology reports from the same period (Appendix 2). This second analysis served to verify that the resulting sublanguage formulas were the same, regardless of the host language employed by scientists.

Chapter 1 describes the sublanguage method, a form of knowledge acquisition that has been applied primarily in text-processing applications (e.g., Sager et al. 1987, Sager 1986, Sager 1978, Hirschman et al. 1976) but which can be used in a more general way as a means of unearthing the information structure of a domain through analysis of texts written by experts. The last three sections of Chapter 3 define this structure using the intriguing concept of a "grammar of science." The purpose of the method is to establish classes of objects relevant in the domain, and classes of relations in which the objects participate. The technique groups different *arguments* of sentences (grammatical subjects or objects) into a class according to their occurrence in the texts with the same *operator* (main verb, adjective, or preposition). Operators are grouped into classes according to their occurring with the same classes of arguments. When the analysis is carried out on a sample of sufficient size, argument classes are found to correspond to domain objects, and operator classes to domain relations.

Chapter 2 presents the classes and formulas obtained for the immunology domain. Formulas are well-formed expressions made up of an operator class and one or more argument classes, and correspond to the "events" of a domain. The argument classes established by the authors include *antibody* (A), *antigen* (G), *cell* (C), *tissue* (T), and *body part* (B). Operator classes include *inject* (J), *move* (U), and *present in* (V). Examples of formulas and the sublanguage sentences they represent are:

G J B "antigen was injected into the foot-pads of rabbits"

A V C "antibody is found in lymphocytes"

G U T "antigen arrives by the lymph stream"

The sublanguage method need not be limited to analysis of texts and can be adapted to incorporate data elicited directly from domain experts. An exciting prospect arises in automating portions of the sublanguage analysis, to create tools to assist the linguist in setting up classes and formulas (e.g., Hirschman et al. 1975, Sager 1975, Grishman et al. 1986), and to interact with domain experts to gather supplementary information and to confirm hypotheses made by the tools.

Chapter 4 discusses the informational properties of the formulas presented in Chapter 2. Sublanguage formulas are a compact notation for knowledge representation that employ a number of devices to enrich the basic structure of operator-argument predication. Modifiers can be placed on operator and argument classes as superscripts. On arguments, they function as unary operators or as quantifiers. Modifiers of operators include negation, quantity, aspect, and direction (of movement). Subclasses of operator and argument classes are indicated by subscripts, e.g., *cell* (C) has subclasses *lymphocyte* (C_l) and *plasma cell* (C_z). A rich set of connectives can join pairs of formulas.

Formulas can be implemented in a fairly straightforward fashion using Prolog terms, relational database tables, a semantic net, an object-oriented system, or a frame-based representation. The choice of implementation would obviously depend on the complexity of the sublanguage being processed and on the application that will make use of the data.

The application of Operator Grammar to sublanguage offers many exciting possibilities for text processing systems. Operator Grammar bears many similarities to Categorial Grammar and shares with combinatorial logic the avoidance of bound variables (cf. Steedman 1989). Chapter 5 describes the transformations of Operator Grammar used in the analysis phase to paraphrase variant sentence forms into the canonical formulas. This does not imply that the recognition algorithm must be a traditional, inefficient transformational system. In fact, the nature of Operator Grammar encourages the design of algorithms which map free text directly into a structured representation (Johnson 1987). The constraints afforded by domain-specific classes are well known, and algorithms can exploit these constraints in a simple way for considerable gains in efficiency. Chapter 7 gives an informal description of procedures for rapid recognition of formulas in free text, and for generation of English sentences from the formulas.

The book is a unique and remarkable contribution, so it goes without saying that the methods will be new to many readers, necessitating a fair amount of unfamiliar terminology. Terms are well explained and used in a clear and consistent manner, but the book suffers for lack of an index. An even greater deficit is the absence of a comprehensive bibliography. One could receive the false impression that the book is a first work by the group (it builds directly on Harris 1982, 1988), or that they are the only group working in sublanguage (cf. the collections by Kittredge and Lehrberger (1982) and Grishman and Kittredge (1986). The absence of references to related work in theoretical or computational linguistics makes the book much less accessible to readers unfamiliar with the sublanguage approach. This is truly unfortunate since there are many fruitful correspondences.

In summary, the book offers a clear description of a much-needed methodology for knowledge acquisition, and a concise, formulaic representation for science information. It is highly recommended to anyone developing text-processing applications in restricted semantic domains.

References

Grishman, R. and Kittredge, R. (eds.) 1986 *Analyzing language in restricted domains: Sublanguage description and processing*. Erlbaum Associates, Hillsdale, New Jersey.

Grishman, R.; Hirschman, L. and Nhan, N. 1986 Discovery procedures for sublanguage selection patterns: Initial experiments. *Computational linguistics* 12(3): 205–215.

Harris, Z. 1982 *A grammar of English on mathematical principles*. Wiley-Interscience, New York.

Harris, Z. 1988 *Language and information*. Columbia University Press, New York.

Hirschman, L.; Grishman, R.; and Sager, N. 1975 Grammatically-based automatic word class formation. *Information processing and management* 11: 39–57.

Hirschman, L.; Grishman, R.; and Sager, N. 1976 From text to structured information: Automatic processing of medical reports. *AFIPS conference proceedings* 45, AFIPS Press, Montvale, NJ: 267–275.

Johnson, S. 1987 *An analyzer for the information content of sentences*. Ph.D. dissertation, New York University.

Kittredge, R. and Lehrberger, J. (eds.) 1982 *Sublanguage— Studies of language in restricted semantic domains*. De Gruyter, New York.

Sager, N. 1975 Computerized discovery of semantic word classes in scientific fields. *Directions in artificial intelligence: Natural language processing*. Courant Computer Science Report 7, Courant Institute of Mathematical Sciences, New York University: 27–48.

Sager, N. 1978 Natural language formatting: The automatic conversion of texts to a structured data base. In: M. Yovits (ed.), *Advances in computers* 17, Academic Press, New York: 89–162.

Sager, N. 1986 Sublanguage: Linguistic phenomenon, computational tool. In: Grishman and Kittredge 1986: 1–18.

Sager, N.; Friedman, C.; and Lyman, M. 1987 *Medical language processing—Computer management of narrative data*. Addison-Wesley, Reading, MA.

Steedman, M. 1989 Combinators and grammars. In: Oehrle, R.; Bach, E.; Wheeler, D. (eds.), *Categorial grammar and natural language structures*. Reidel, Dordrecht: 417–442.

Stephen Johnson is an associate research scientist in the Center for Medical Informatics at Columbia University, conducting research in active databases and natural language processing for medical applications. He holds a doctorate in computer science from New York University, where he was a member of the Linguistic String Project. His dissertation presented an implementation of a parsing algorithm based on Operator Grammar. Johnson's address is: Center for Medical Informatics, 161 Fort Washington Ave., AP-1310, New York, NY 10032. E-mail: adjohnson@cuhsda.bitnet

---

Generalized Quantifiers: Linguistic and Logical Approaches

Peter Gärdenfors (ed.)
(Department of Philosophy, Lund University)

Dordrecht: D. Reidel, 1987, vii + 307 pp.
(Studies in linguistics and philosophy 31)
ISBN 1-55608-017-4, $74.00, £59.00, Dfl 158.- (hb)

Reviewed by
Kenneth Ray
Trent University

This collection of 10 papers incorporates proceedings of the 1985 Lund conference on generalized quantifiers (GQ). Research on GQ was brought into natural language analysis in 1981 by Barwise and Cooper in their "Generalized quantifiers and natural language". The aim was to elevate model-theoretic analysis of NL phenomena from a sterile exercise in formalization to a valuable stimulus in development of linguistic theory. Following Montague's PTQ, generalized quantifiers were treated not as the determiner expressions in a noun phrase (NP), but rather as the entire NP construction. An NP determiner functions to select a family of sets from the head noun's extension as the denotation for the NP. Interpretation of the noun denotation as a restriction upon the domain of quantification then allows for uniform semantics for NPs, encompassing non-logical determiners, (e.g., *most*, *a few*), along with the traditional logical determiners such as *every* and *some*.

The articles in this volume pick up on this theme by extending the GQ analysis to many of the syntactically varied forms of NP constructions. At least four distinct approaches to semantic interpretation are considered, showing the interest in exploring alternatives to the possible worlds interpretations of Montague.

Jon Barwise and Robin Cooper each have contributions that incorporate interpretive structures from situation semantics in building an alternative to Montague's model-theoretic interpretations. Situation semantics offers a more intuitive and simplified domain of individuals, properties, and facts for model-theoretic construction than the intensional domain of functions across possible worlds. Because the focus is on logical investigations into the semantic properties of GQs, however, computational issues concerning implementation of the proposed semantic models are not explored. Accordingly the volume is directed to those with well-developed research interests in formal methods, focusing on modeling a variety of NP phenomena. Other themes explored within these fine-grained treatments of quan-