

# Book Reviews

## Natural Language Processing: A Paninian Perspective

Akshar Bharati, Vineet Chaitanya, and Rajeev Sangal  
(Indian Institute of Technology, Kanpur)

New Delhi: Prentice-Hall of India, 1995,  
xviii + 220 pp.  
Paperbound, ISBN 81-203-0921-9,  
Rs 95.00, \$20.00

*Reviewed by*  
Brendan S. Gillon  
McGill University

This book is an elementary introduction to natural language processing, but it is a very unusual one for two reasons. First, its subject languages are those of India; second, although the book sets out and discusses the better-known linguistic grammars that are commonly used in natural language processing, the grammar that the book actually uses is one based on the *Aṣṭādhyāyī* (pronounced 'ashtad'yayee'), humankind's earliest extant grammar.

To begin with, some background with regard to this grammar is in order. The *Aṣṭādhyāyī* is a grammar of the Sanskrit language, as it was spoken by the descendants of the Indo-Aryan tribe living on the Gangetic plain around the sixth century B.C. The grammar, written or compiled by Pāṇini, contains a little over 4,000 rules. Sanskrit, like its European cousin Latin, is a language in which the most salient grammatical device is inflection, while the least salient one is word order, and Pāṇini's grammar reflects this. The grammar is designed so that, together with a Sanskrit lexicon and a canonical specification of a situation, it is supposed to generate all and only the correct Sanskrit sentences expressing the situation. The assumption made is that each situation to be expressed can be framed as an action or event (*kriyā*) with associated factors (*kāraka*), roughly *valences*. Once a one-to-one correspondence has been established between the elements of a situation canonically framed, on the one hand, and underived root morphemes from the lexicon, on the other, the grammar operates nondeterministically to yield a set of morphologically well-formed words that together express the situation.

Up until the latter half of the twentieth century, Pāṇini's grammar was the most rigorous and comprehensive ever written for any language. Moreover, many, if not most, of the key ideas in modern linguistic theory have their origin in the ideas of his grammar (e.g., sandhi) or were anticipated by it (e.g., the theta criterion). The tradition that the grammar initiated made Sanskrit, until this century, the most thoroughly studied human language.

Though the languages of the Indian subcontinent belong to two distinct language families, Indo-European (those descended from Sanskrit) and Dravidian, they have much in common: in particular, they tend to be highly inflected. Hence, the treatment of inflection, and not word order, must play the most important role in the processing of such languages. English is a language in which inflection has only a marginal role, while word order has a central role. As a result, it is not surprising that other grammars—such as Government and Binding, lexical functional grammar, and head-

driven phrase structure grammar, all of which are largely inspired by the study of English—are not as well suited to analyze the languages of India as a grammar such as Pāṇini's. Hence, the authors' development and deployment of a Pāṇinian-style grammar is particularly apt and insightful, especially given the range of languages addressed.

The first chapter, entitled 'Introduction to NLP,' provides a clear and concise overview of the field of natural language processing, making clear the relevance of linguistics to this domain. The basic grammatical concept of modification is introduced in the second chapter, 'Language Structure and Language Analyzer,' together with a discussion of those aspects of computation relevant to natural language processing. Since, as was pointed out above, morphology plays a prominent role in the languages of the Indian subcontinent, it is inevitable that one whole chapter, entitled 'Words and Their Analyzer,' should be devoted to morphological analysis. The specifically Pāṇinian concepts of the grammar are set out in the fourth through sixth chapters, entitled 'Local Word Grouping,' 'Pāṇinian Grammar,' and 'Pāṇinian Parser.' An overview of Pāṇini's own grammar is set out in Appendix A, 'Pāṇini's Grammar and Sanskrit.' The next chapter turns to machine translation. By using look-up and a morphological analyzer and generator, a crude approximation of machine translation can be implemented, as has, in fact, been done between Hindi, an Indo-European language, and Kannada, a Dravidian language. The remaining six chapters of the book are devoted to a critical discussion and comparison of three major grammatical theories with the Pāṇinian grammar—lexical functional grammar, tree-adjoining grammars, and Government and Binding.

The book is an introduction to natural language processing. As such, it inevitably contains certain confluations, omissions, and abridgements. The most important abridgements are those in the presentations of the other grammatical theories, where the discussions are absolutely minimal. Fortunately, each chapter is terminated by a useful set of references to additional reading that can fill out the relevant grammatical theory. The most important omission, in my view, is that of compounding. Compounding is an important and productive feature of many Indian languages, yet it is not even mentioned. Pāṇini's grammar treats compound formation and indeed, his treatment is the one used by Robert Lees in his pioneering work *The Grammar of English Nominalizations* (1960). However, Sanskrit compounds, like English compounds, are probably better treated by context-free rules.

The most important conflation is that of complementation with modification. One of the central grammatical concepts used in the book is termed modification. This term is the English rendition of *viśeṣya-viśeṣaṇa-bhāva*, literally, the relation between what qualifies and what is qualified. As used by Pāṇini's successors and by the authors of this book, the relation is more akin to the relation of subordination, including, on the one hand, the relation of modification proper—that of an adjective to the noun it modifies—and, on the other, the relation of complementation—that of noun phrase and prepositional phrase complements of a verb to the verb.

None of these abridgements, confluations, or simplifications detracts from the book in any way. This book provides a concise and clear introduction to natural language processing, not only for anyone who wants to focus on the languages of India, but also for anyone who wishes to have a perspective on natural language processing broader than that attainable by a focus on English alone.

#### Reference

Lees, Robert (1960). *The Grammar of English Nominalizations*. Mouton.

*Brendan S. Gillon* earned an M.A. in Sanskrit and Indian Studies at the University of Toronto and his Ph.D. in Linguistics and Philosophy at MIT. His publications in semantics include "Towards a common semantics for English count and mass nouns" (in *Linguistics and Philosophy*, 1992), and "Peirce's challenge to material implication as a model of 'if'" (to appear in *Analysis*). He spent eighteen months at Poona University's Centre for the Advanced Study of Sanskrit and now teaches semantics and Sanskrit linguistics at McGill University. Gillon's address is: Department of Linguistics, McGill University, Montreal, Quebec H3A 1G5, Canada; e-mail: [gillon@langs.lan.mcgill.ca](mailto:gillon@langs.lan.mcgill.ca).