

Similar discussions apply to parsing. Fifty thousand words of the corpus have been parsed by hand, and this has been used to make a table of the relative frequencies of different syntactic constructions. Assuming that the correct parse tree is the one made of the most probable constituents (to greatly oversimplify in the interests of saving space), a program was written to parse with about 50% accuracy. Since the preparation of this book, continuing work by Eric Atwell and Geoffrey Sampson at Leeds has greatly improved on this figure, using a simulated annealing technique (see Sampson 1986).

Other chapters of the book discuss the history of corpora in linguistic research, a defense of probabilistic methods, a discussion of speech synthesis and an outline of a sophisticated spelling corrector. Not much has been done on speech synthesis, partly because we do not as yet have good data on the relation between syntax and prosody. The spelling corrector is aimed at errors of word selection, i.e., finding words that, although they appear in the dictionary, should not appear in the particular sentence being studied (e.g., "They kingdom come, thy will be done"). All these tools follow the same model: reliance on statistics from the corpus.

It is a great relief to read a book like this, which is based on real texts rather than upon the imaginary language, sharing a few word forms with English, that is studied at MIT and some other research institutes (see Postal 1988). It is amazing that computers, which are distinguished for their ability to deal with vast quantities of bytes and their incompetence with even simple patterns and models, have been used in linguistics primarily for the implementation of complex logical models. This book is a start on the exploitation of large database methods for linguistic information. It is remarkable for the performance of its methods combined with their simplicity. Unlike many books on linguistics, it is easy to understand; it makes one think of the Molière character who suddenly found out he had been speaking prose all his life.

I heartily recommend this book to anyone who wishes to process language for a useful purpose. Other workers such as John Sinclair (1987) and Yaacov Choueka (1988) have also used large text databases for deriving linguistic information. When I was an undergraduate, one of my professors said that "mathematical intuition means having seen the problem before." Similarly, there is no substitute in linguistics for knowing that a particular construction is likely because it has appeared many times. This book is a testimony to the superiority of experience over fantasy.

#### REFERENCES

- Choueka, Yaacov 1988 Looking for Needles in a Haystack. In *Proceedings of the RIAO 88*, 609–623.  
 Postal, Paul 1988 Advances in Linguistic Rhetoric. In *Natural Language and Linguistic Theory* 6:129–137.

Sampson, Geoffrey 1986 Simulated Annealing as a Parsing Technique. In *University of Leeds Working Papers in Linguistics and Phonetics* 4:43–60.

Sinclair, John 1987 *Looking up*. Collins, London, England; Glasgow, Scotland.

*Michael Lesk* is division manager of computer science research at Bell Communications Research, 445 South St., Morristown, NJ 07960. He uses machine-readable dictionaries in his research on text handling and retrieval. E-mail: lesk@wind.bellcore.com

---

#### SEMANTIC INTERPRETATION AND THE RESOLUTION OF AMBIGUITY

**Graeme Hirst**  
 (University of Toronto)

Cambridge, England: Cambridge University Press,  
 1987, xiv+263 pp.  
 ISBN 0-521-32203-0; (hb) \$39.50 [20% discount to  
 ACL members]

*Reviewed by*  
*Karen Sparck Jones*  
*University of Cambridge*

Hirst's book presents an approach to natural language interpretation, using as his vehicle a description of the experimental system he built. It therefore has to be evaluated as a contribution on how to build NLP systems from both theoretical and practical points of view. It also has to be considered for teaching purposes, since Hirst has vamped up what was originally a thesis with some pedagogic exposition and test exercises, as well as a substantial and useful bibliography.

Hirst is very clear about his aims and very honest about what he has tackled. He presents detail well and provides excellent summaries, so the essential properties of his work are well laid out.

His goal was to build an interpretation system that could handle serious lexical and structural ambiguity, and handle it in a principled way. His concern is thus essentially computational; he does not make any claims for the psycholinguistic relevance of what he is doing, but he is, on the other hand, willing to exploit psycholinguistically derived support for good processing strategies.

The system consists of a syntactic parser, **Paragram**, a semantic interpreter, **Absity**, and two disambiguation processors: the **Polaroid Word** (PW) subsystem for lexical disambiguation and the **Semantic Enquiry Desk** for structural disambiguation. The system builds an explicit meaning representation in the frame language **Frail**.

Hirst's design is motivated by two goals: to allow processes of different sorts to use different kinds of information but to interact to construct a sentence representation; and to do this in the theoretically well-founded way exemplified by Montague's work by doing

interpretation compositionally with formally proper semantic objects. The specific "non-standard" issues he addresses within this framework are handling the loose associative processes that appear to be needed for word sense selection, and allowing for emergent sense identification.

In sentence processing, therefore, the basic mechanism is provided by the Paragram parser that Hirst exploits, working in tandem with his Absity interpreter. Absity is strongly typed, with types referring to the various elements and constructs of (his extended) Frail: for example, generic and instance frames, slot names and slot filler pairs, and frame determiners. The tandem operation of parser and interpreter is thus designed to map syntactic types onto their corresponding semantic types: for instance, noun phrases into frame statements (i.e., knowledge base access statements), prepositions into slot names, or verb phrases into frame descriptors (i.e., complete descriptions of generic frames). However, as Paragram is essentially deterministic, Hirst has to compromise a little and sully his pure compositional ideology by using pseudo-words and fake objects.

Lexical disambiguation within this framework relies on the PW processors. These are syntactic type procedures (in principle for all types of word) individualized with lexical information for each word through which, in turn, it is possible to access the Frail knowledge base and hence information about the world. The procedures develop, i.e. disambiguate, their PWs in two ways: by exploiting the results of marker passing, which is an entirely autonomous, non-syntactic process operating on the frame data for input words and establishing links between and hence sense selections among these using slot and isa relationships; or, if the marker passing is not sufficiently selective, by heavily constrained interword checking for slot filler connections. These processes are activated by each incoming word, so PWs may be only partly developed at intermediate points in the sentence input.

For structural disambiguation, limited to prepositional phrase attachment and relative clause gap finding and filling, Paragram looks to the Semantic Enquiry Desk, which is supplied with relevant PW information for germane candidates and applies strategies relying heavily on slot filler restrictions and invoking, e.g., Crain and Steedman's *Principle of Referential Success*, which fit with the way semantic representation is handled through objects in the Frail knowledge base. Hirst is able to treat his two types of phenomenon in a fairly algorithmic way because he can call on the systematically organized and rich information available in the knowledge base.

With these lexical and structural procedures Hirst can, for example, resolve "The slug operated the vending machine" and "Ross included the book for Nadia", but not "The astronomer married the star" or "The women discussed the dogs at breakfast". The fact that he can successfully disambiguate "The deep philoso-

pher threw the peach pit into the deep pit" may, perhaps, be regarded as a triumph.

Hirst presents his own ideas in the context of extensive accounts of the problems to be solved, and of other approaches to them. But he unfortunately missed Alshawi's work (Alshawi 1987, but available as a technical report in 1984), which also combines more conventional syntactic and semantic processing with marker passing, and which attacks text processing above the sentence in a serious way. Carter (1987) recently described work on anaphor resolution in text which subscribes to the multipronged and side-effect philosophy of interpretation that Hirst espouses, though the more specific character of his work is different.

Hirst is so open about the limitations of what he has done—for example, about the semantic phenomena, like noun modifiers, that Absity cannot handle—that he to some extent preempts criticism. His final chapter is particularly disarming, since it cheerfully invites the reader, through speculations, questions, and suggested exercises, to think about the way his approach could be developed to tackle a whole range of interesting and challenging problems. But Hirst does refer to his achievements in terms that imply they have been demonstrated, and he emphasizes his claim that Frail objects are kosher semantic objects of the kind all right-thinking semantic theories ought to have.

However, though Hirst's approach is interesting, and his account of it is quite detailed, he does not really give us enough evidence, in the form of system data, behaviour traces, or output results to evaluate what he has done or could do; and, as an important part of this, to judge the validity of his argument that Frail objects rule OK in the way mere symbols don't. This is put in a robustly plausible way, but, in fact, too briskly to convince. Hirst appears to maintain that because Frail allows systematic operations on well-defined objects, and because these internal objects can be taken for external ones, attaining thereby a Germanic ding-hood, as it were, those nasty philosophical questions about the status of the representation language itself somehow disappear. They don't quite.

*Semantic Interpretation* is an interesting and enjoyable read. Hirst combines perceptive analyses with commonsense attitudes (for instance, to the definition of "polysemy") in an engaging way, and he has made a valuable attempt to synthesize some attractive, but hitherto distinct, ideas about language interpretation. This is food for thought for the research worker, and there is matter for digestion for students in the pedagogic surveys and in the way Hirst uses his system as a problem solving illustration. Many of the examples are a delight, and instructive not least because memorable (though I'm beginning to feel quite bothered about Nadia, publicly exhibited for our benefit in deplorable attitudes, like happily cleaning up after Ross, for instance, though Ross can quite clearly look after himself).

But the way Hirst has souped up his thesis with general expository matter has had not altogether satisfactory results. The textbook-like sections against which the accounts of his own work are set treat some topics, like structural ambiguity, at length, but do not form a very well-balanced or comprehensive whole. They nevertheless dilute Hirst's own work enough to prevent the reader from experiencing that feeling of excitement that good theses provoke, and to leave her disappointed in not getting the fuller and more concentrated account of Hirst's system as such, and its performance, she would have liked. The idea that language interpretation involves quite disparate processes is one deserving investigation, but it is not clear whether Hirst's particular way of combining such different lines as Montague and markers, via frames, is on the right track, and his book does not compel his reader, in a Pied Piper imperative, to follow him.

#### REFERENCES

- Alshawi, H. 1987 *Memory and Context for Language Interpretation*. Cambridge University Press, Cambridge, England.  
 Carter, D.M. 1987 *Interpreting Anaphors in Natural Language Texts*. Ellis Horwood, Chichester, England.

*Karen Sparck Jones* is co-editor of *Automatic Natural Language Parsing* (Ellis Horwood 1983). Her present research concerns user modeling. Sparck Jones's address is: Computing Laboratory, University of Cambridge, Pembroke Street, Cambridge CB2 3QG, England. E-mail: sparckjones@cl.cam.uk.ac

NOTE: The author of this book is also the book review editor of this journal. Therefore, this review was edited by James Allen, editor of *Computational Linguistics*.

#### THE FIFTH GENERATION FALLACY: WHY JAPAN IS BETTING ITS FUTURE ON ARTIFICIAL INTELLIGENCE

J. Marshall Unger

(University of Hawaii, Honolulu)

Oxford University Press, 1987, x+230 pp.  
 ISBN 0-19-504-939-X; (hb)

*Reviewed by*  
*Harold Somers*  
 UMIST

The main proposal of this book is that the motivation behind the Japanese Fifth Generation project is not a desire to push the barriers of computer technology research, nor even to attain economic superiority through advanced technology, but quite simply to overcome the problems of the Japanese writing system. In advancing this claim, Unger demonstrates that he is at least well read—an impressive array of literature in both English and Japanese is cited—and the book is well written, although sometimes in a journalistic rather than academic style. Despite that, this reviewer, and at least

four other informed colleagues that I have discussed this book with, find the claim quite preposterous. The book is an enjoyable read, in the way that the Letters page in an extremist newspaper is sometimes enjoyable, though I fear that only those who already have a fair background knowledge of the issues will get very far with it: if you don't know the difference between a kanji and a kana, then you might find it rather hard going.

The book is divided into three sections, each of two chapters, dealing with Japanese linguistics and orthography, political and cultural issues, and economics and technology. This is preceded by an Introduction which sets the scene, actually getting the book off to a very promising start, in which Unger virulently attacks the typical reaction to the Fifth Generation, as exemplified by the well-known Feigenbaum and McCorduck (1983) book, which exhibits paranoia, misunderstanding of Japanese attitudes to AI, the weak/strong AI debate, and the reasons for the Fifth Generation:

What the Japanese have in mind when they speak of AI, however, turns out to be something else yet again, and unless one dispels the fog of strong-AI hyperbole that surrounds the Fifth Generation project, it is impossible to appreciate its significance. (p.4)

It never occurs to [Feigenbaum and McCorduck] that, despite Japan's new-found affluence, its intellectual climate continues to suffer from such fundamental conditions as geographical and linguistic isolation, academic factionalism, and the overweening influence of government bureaucracy and giant corporations. (p.5)

It is as an answer to this misunderstanding that the remainder of the book is intended, and it begins with a detailed description of the Japanese writing system, and some of its practical consequences. As I suggested above, this whole section is probably too complex for the reader who does not know about it already, and this complexity is not helped by the unusual approach, starting from a romanization presented in alphabetical order and working towards the kana rather than vice versa. Some minor criticisms of this section would include the fact that the term *mora* (roughly, the consonant-plus-vowel syllable that each kana represents) is not included in the otherwise extensive glossary; the error (pp.21–22) of describing the Hepburn romanization as phonemic, when it is precisely not phonemic (e.g. /tu/ is written *tsu*), whereas the Cabinet (*kunrei*) romanization is; and there is no mention of a third romanization (especially used by Japanese) which is a simple transliteration of kana, e.g. *Toukyou* for *Tōkyō*. While we are on the subject of romanization, we mention also the quaint use by the publisher of the circumflex rather than the macron to indicate long vowels in the Japanese transcription.

Turning to the question of kanji, it is always difficult to convey accurately to readers used to alphabetic writing systems the pros and cons of a radically different orthography, and like most Western observers, Unger overestimates the problems they cause