

Trajectories Through Knowledge Space: A Dynamic Framework for Machine Comprehension

Lawrence A. Bookman

(Sun Microsystems Laboratories)

Boston: Kluwer Academic Publishers
 (The Kluwer international series in
 engineering and computer science:
 natural language processing and
 machine translation, edited by Jaime
 Carbonell), 1994, xx+271 pp;
 hardbound, ISBN 0-7923-9487-9, \$89.95,
 £62.95, Dfl 170.00

Reviewed by

Jean-Pierre Corriveau
 Carleton University

In the foreword to this recent book by Larry Bookman, Wendy Lehnert writes, "this volume is without question a milestone in language processing scholarship." To put it simply, I agree. In 236 pages, Bookman develops at length the two-tier model of semantic memory that he first introduced in his doctoral dissertation on text comprehension. The relational (symbolic) tier captures a set of dependency relationships between concepts. The analog semantic-feature (ASF) tier represents (at the subsymbolic level) common or shared knowledge about the concepts in the relational tier, expressed as a set of statistical associations. This hybrid approach to NLU, named LeMICON (Learning Memory Integrated Context), stands out first and foremost because it claims to address the bottleneck of hand-coding the knowledge required for interpretation. More specifically, LeMICON offers both automatic acquisition of knowledge from on-line text corpora and integration of a text's interpretation into the knowledge base. Before commenting on this breakthrough, let me summarize the work.

Much like the model of Lange (1993) and my own model (Corriveau 1995), LeMICON tackles most facets of text comprehension, in contrast with PDP models, such as that of Miikkulainen (1993), which are typically more specialized. Bookman first skillfully presents, in a 21-page introduction, the essential characteristics of his work. LeMICON ignores syntax, but considers several psycholinguistic and neurolinguistic problems that are typically ignored by others. For example, it models the loss of information in working memory, and the time-course of inferences as well as their construction process. Throughout this discussion, the spreading-activation process, so typical of local and hybrid connectionist architectures such as LeMICON, is thought of as constructing *time trajectories* through the space of ASFs. That is, "energy" propagates through ASFs to form, over time, *stable* chains corresponding to inferences. The notion of ASFs proceeds from Waltz and Pollack's (1985) use of "microfeatures". More specifically, each concept exists as a named node at the relational tier, but is expressed as a set of ASFs; the network of these ASFs forms the associational tier. Furthermore, "comprehension is more than just a passive trace through concept space": activation of *relations* between concepts "sets up an expectation of what patterns of knowledge are likely to follow". Consequently, as a result of its processing of a text, LeMICON generates two sources of knowledge: a set of ASF trajectories and an *interpretation graph*.

The set of ASF trajectories produced is a representation of the fine-grained knowledge that LeMICON has accumulated in working memory as a result of its processing of an input text. Conversely, the interpretation graph captures a “coarse-grained view [of understanding as] a set of explicit graded semantic relationships that can be used to reason about the meaning of a text.” Clearly, this dual nature of an interpretation stems directly from the hybrid nature of the model. The introduction concludes with a short account of the historical evolution of LeMICON and a useful plan for the rest of the book.

The second chapter provides, in 26 pages, a brief overview of connectionist and probabilistic research in NLU. Psycholinguistic considerations, organized around the work of Jurafsky (1992), are quite laconic. Conversely, the discussion of how connectionist and probabilistic models combine and analyze information is not only original in its presentation, but also thorough and yet highly readable. Not surprisingly, the conclusion reached by Bookman is that “there is much to be gained in combining the approaches of connectionist and computational linguistics models of language.”

Chapters 3 to 8 then detail, in 140 pages, the nature and workings of LeMICON. First, the memory architecture is studied. On the one hand, the relational tier reduces to a semantic network. On the other, ASFs are essentially microfeatures (augmented to include background knowledge) and act as the prespecified conceptual primitives of the system. But Bookman stresses that ASFs differ from both semantic features (of symbolic models) and the microfeatures of PDP models. More specifically, ASFs are based on and organized according to the conceptual categories and classifications used by the Roget thesaurus, with some additional ASFs added to account for the specific domain at hand (namely, stock market ‘stories’). However, the total number of categories in a thesaurus is extremely large. Thus, in LeMICON, a subset of 454 of the categories from Roget are used to represent ASFs. Consequently, each concept is associated with a background frame consisting of a vector of 454 ASFs! Working memory consists of three short-term buffers: an input, a reactive, and a case-role buffer. The case-role buffer stems from the fact that the actual input consists of a *pre-parsed* input in which each concept and its filled case slots are replaced by their respective ASF encodings, that is, by vectors of dimension 454!

As previously mentioned, a fundamental characteristic of the model is that as new trajectories are triggered by the input, LeMICON attempts to integrate this new information with antecedent knowledge. To achieve this, the model subscribes to Haberlandt and Graesser’s (1990) delayed-integration hypothesis: “strongly activated patterns will remain active for a longer time and should therefore have a greater chance of being successfully integrated.” Bookman does an excellent job of explaining how this process is carried out in LeMICON. It is regrettable, however, that he fails to consider the work of Gernsbacher (1990) on this topic.

In chapter 4, the comprehension algorithm is further detailed. In particular, section 4.3 provides a step-by-step description of the *modus operandi* of LeMICON: Each clause activates concepts in semantic memory. Case-role patterns are propagated, and then competing assemblies are inhibited. Finally, relational *novelty* is computed and working memory is updated to integrate relevant new information. This comprehension algorithm relies on notions such as *ASF closeness* and *relational closeness*. However, these measures are not hardwired, but rather mathematically defined with respect to concepts and ASFs! The representation of the input is also discussed further. For example the sentence:

The stock market declined 50 points yesterday.

is “translated” into:

Clause 1: (decline (OBJ stock market) (VALUE 50 points) (TIME yesterday))

which, in turn, is encoded using a vertical concept assembly relating a high-level concept to its relevant case-roles. LeMICON uses 16 case relations selected from previous research on this topic. Bookman must be commended for the completeness and readability of this thorough description, which includes computational details, treatment of bindings, and links of LeMICON to psychology and neurophysiology. The chapter concludes with an interesting comparison to several other text understanding systems.

In chapter 5, Bookman elaborates on the ability of LeMICON to both skim and perform in-depth understanding. The presentation of some simple stories processed by the system along with their corresponding summaries may remind the reader of Dyer’s (1983) seminal book. But quickly, Bookman moves from this qualitative analysis to more specific issues, such as quantifying the volume of inferences, analyzing time-dependent interactions, characterizing working memory closeness, and so on. Most interestingly, LeMICON manages to build clusters of trajectories for similar stories *without* requiring the plethora of knowledge structures found in Dyer’s book.

Chapter 6 proceeds with a discussion of the uses of the interpretation graph for reasoning about the interpretation of a text. In particular, this graph can serve to identify the basic events (called the *conceptual roots*) of an interpretation and their importance. These roots are essential to the summarization task.

Chapter 7 is first devoted to the acquisition of knowledge from on-line corpora. LeMICON uses the *Wall Street Journal* database. The notion of *mutual information* is applied to pairs of concepts to determine their statistical relationship. The key point to understand here is that ultimately the whole approach rests on detecting co-occurrences of concepts in the corpus. The rest of the chapter addresses how memory is updated as it processes the input and constructs the interpretation. In chapter 8, Bookman takes a close look at the kind of knowledge that is learned and introduces the notion of knowledge *effectiveness* to compare different text understanding systems with respect to the use and complexity of the information they process.

To conclude, Bookman recapitulates in chapter 9 the important contributions of his work, and they are numerous. Then, in chapter 10, future directions are proposed: dealing with more than 100 concepts, learning deeper semantic relationships, handling contradictory input, and so on. Finally, the book includes several appendices detailing the ASFs used, the complexity of the different algorithms, the nature of the pre-parsed input and some additional results.

LeMICON is not perfect. Bookman explains that LeMICON ignores syntax and metaphoric text, as well as complex inferences (such as social ones). And it uses a limited form of temporal reasoning. Most importantly, LeMICON must preprocess the background knowledge for each new text, “a severe limitation for any real-time practical text system” by Bookman’s own admission. In my opinion, there are three problems with such a preprocessing strategy. First, any non-automated translation process is unavoidably subjective and typically eliminates important facets of comprehension (such as structural ambiguity). Indeed, it is not even clear whether or not LeMICON does address genuine lexical and referential ambiguities. Second, both the structure of the input format and the specific primitives used in the actual input (i.e., the ASF encodings) must be known *a priori*. In other words, the processing strategy relies on a fixed prespecified representational scheme that makes specific epistemological commitments. This is most noticeable in reading that the 454 ASFs were partly chosen with respect to a particular domain, the stock market. Third, it appears that

the model has poor space complexity, because each concept of the input text requires a large vector of features.

And now for the crucial claim: that LeMICON learns the knowledge base that it uses for comprehension. In my opinion, this is only partly true. First, one could object to the use of co-occurrence for genuine learning (see Corriveau (1995, chapter 2) for details). But more fundamentally, one must understand that, in LeMICON, the knowledge base is not constructed entirely from corpus. In fact, each concept of the relational tier consists of a *prespecified* schema defined by the possible case relations known to the system: LeMICON uses “approximately 100 concepts [that] were chosen based on how representative they were for the given stock market domain”. Only the relationships between these subjectively chosen concepts are learned, not the concepts themselves. And all ASFs are hardwired *a priori* in LeMICON: no ASF can be learned. Thus, a “new” concept for LeMICON is merely a concept with a *known* ASF encoding, but not yet included in the relational tier. There is apparently no learning on the ASF tier, the most important of the two tiers of the model. For example, *hospital* is always related to *emergency*, *admission*, *care*, and *discharge*. Viewing a hospital in a different way (e.g., as a workplace for nurses and doctors) cannot be learned, regardless of the number of texts processed by LeMICON. It follows that LeMICON constructs the interpretation of a text from a set of *fixed* meanings. From this standpoint, the system is still missing the key ability to modify what it has learned statistically from a corpus in light of the texts it processes. More specifically, beyond the integration of parts of an interpretation into the knowledge base, the meaning of words must be diachronic, that is, evolve over time. And the meaning of a word in a sentence—that is, its set of features relevant in that particular context—must be constructed from its interactions with surrounding words, not merely recognized. We still lack a computational theory for such models. We are still far away from Rastier’s (1991, p. 114) truly contextual view of meaning where every “semantic occurrence is a hapax.” But Bookman must receive immense credit for addressing how the knowledge base is initially constructed in text comprehension systems, a fundamental requirement for scalability.

References

- Corriveau, Jean-Pierre. (1995). *Time-constrained Memory: A Reader-based Approach to Text Comprehension*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gernsbacher, Morton Ann. (1990). *Comprehension as Structure Building*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Haberlandt, Karl F. and Graesser, Arthur C. (1990). “Integration and Buffering of New Information.” In: *Inferences and Text Comprehension*, edited by Arthur C. Graesser and Gordon Bower. NY: Academic Press.
- Jurafsky, Daniel. (1992). *An On-line Computational Model of Human Sentence Processing*. Ph.D. thesis, Computer Science Divisions, University of California at Berkeley.
- Lange, Trent E. (1993). Massively-parallel Inferencing for Natural Language Understanding and Memory Retrieval in Structured Spreading-activation Networks. *Proceedings, 11th National Conference on Artificial Intelligence (AAAI-93)*, Washington, DC, 144–149.
- Miikkulainen, Risto. (1993). *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon and Memory*. Cambridge, MA: The MIT Press.
- Rastier, François. (1991). *Sémantique et recherches cognitives*. Paris: Presses Universitaires de France.
- Waltz, David L. and Pollack, Jordan B. (1985). “Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation.” *Cognitive Science*, 9: 51–74.

Jean-Pierre Corriveau’s research interests include massively parallel natural language understanding systems and models of human memory for text comprehension. He received a Ph.D. in

artificial intelligence from the University of Toronto and recently published a book entitled *Time-constrained Memory: A Reader-based Approach to Text Comprehension* (Lawrence Erlbaum Associates, 1995) that presents a computational model of text comprehension. Corriveau's address is: School of Computer Science, Carleton University, Colonel By Drive, Ottawa, Ontario, Canada K1S 5B6; e-mail: jeanpier@scs.carleton.ca