Speech-to-speech translation: A massively parallel memory-based approach

Hiroaki Kitano
(Carnegie Mellon University and Sony Computer Science Laboratory, Tokyo)

Boston: Kluwer Academic Publishers
(The Kluwer international series in
electrical engineering and computer
science; natural language processing
and machine translation, edited by
Jaime Carbonell), 1994, xvii + 193 pp.
Hardbound, ISBN 0-7923-9425-9, $85.00,
£63.75, Dfl 180.00

Reviewed by
Nigel Ward
University of Tokyo

This book, based on the author's Kyoto University dissertation, describes a system able to translate spoken Japanese to spoken English in real time. For this work, as the back cover notes, Kitano was awarded IJCAI's Computers and Thought Award in 1993, and it is not hard to guess what impressed the award committee. This work not only incorporates many of the key ideas of artificial intelligence—including parallel search algorithms, memory-based reasoning, plan recognition, and meaning-based parsing—it brims with excitement, with statements like "the basic approach taken in traditional [machine translation] systems faces a serious dead-end, and needs a dramatically different paradigm" (p. 35) and "the ideas grown out of this work lead me to propose massively parallel artificial intelligence, which is now being recognized as a distinct research field" (p. xvi).

This review will first discuss the basic idea of the system, then the book itself, and finally the research strategy.

The most significant contribution of the work, according to Kitano, is that it "demonstrated that real-time spoken language translation at the order of milliseconds is attainable" (p. 176). This was achieved by using memory-based reasoning. It is easy to see how this works in the simple case: a system can store all possible input/output sentence pairs in memory, one per processor, enabling it, when given an input to translate, to quickly recognize which sentence matches, and output the appropriate translation. Moreover, this allows efficient speech recognition: after some of the input has been seen, it is easy to predict what might come next, which provides a constraint on the phonemes the speech recognizer has to consider.

But, of course, a system that can only translate sentences that are already stored in its memory is not very useful. How can a memory-based model cope with novel sentences? The book provides three possible answers to this crucial question.

Answer 1 is the old argument that, since there are no infinitely long sentences, the number of sentences in any language is finite, and so a sufficiently big memory will suffice. This argument is backed up by reference to an unnamed "empirical study" which "shows that only 0.53% of possible sentences are observed in large corpora" (p. 40); no explanation is given of what this could mean.

Answer 2 is a "memory network" that augments the store of sentences with: their constituency structures, abstract versions of the sentences, and the plans that underlie utterances of the sentences. Given an input, the process of finding which of the nodes in this network are relevant is performed by marker-passing, specifically, by a variant of Riesbeck and Martin's (1985) Direct Memory Access Parsing algorithm. Unfortunately, there is no explanation or illustration of exactly how this enables the system to parse, represent, or generate novel sentences.

Answer 3 is "a method to fuse constraint-based and case-based approaches" (p. 68). Here the constraint-based approach is presented as similar to unification, and, therefore, is presumably computationally powerful enough to handle novel sentences; but again, there is no explanation or illustration of exactly how.

Even taken together, these answers do not seem to suffice; there is no indication that the implemented system could handle novel inputs. Indeed, the only specific statement about performance, apart from timing results, is the enigmatic assertion that "at least over 300 sentences in the [ATR Conference Registration] corpus have been covered" (p. 146) in one version of the system.

Now to the book itself. Chapter 1 introduces the problem of speech-to-speech translation. One point to note is that, as the title suggests, this is not a book about "interpretation" but about "speech-to-speech translation" (a distinction that other writers sometimes obscure). What this system does is parse a sentence, translate it, and pronounce the result; there is no pretence that this is adequate for real interpretation, as performed by human interpreters, with all its complicating, or perhaps simplifying, aspects.

Chapter 2 discusses some related research, although only a small fraction of the relevant work is cited. Even that which is mentioned is scarcely analyzed; for example a lot of interesting and relevant work is curtly dismissed with the sentence: "[Saito and Tomita, 1988] [Kita et. al., 1989] and [Chow and Roukos, 1989] are examples of approaches to integrate speech with unification-based parsing, but, unfortunately, discourse processing has not been incorporated" (p. 104, punctuation as supplied). Incidentally, despite the 1994 copyright, there is no reference to work after 1991, so there is no mention of, for example, the NEC and SRI spoken language translation systems (Hatazaki et al. 1992; Rayner et al. 1993).

Chapter 3 outlines the philosophy behind the system design. The arguments are fairly clear, although the story is oversimplified. For example, it is not true that "memory-based processing ... has not been investigated in past studies on machine translation" (p. 30); Tomabechi (1987) worked on exactly this. Indeed, this book draws very heavily on the ideas in one "past study" (Tomabechi et al. 1989).

Chapter 4, the system description, comprises over a third of the book. In addition to the basic marker-passing parser, it explains mechanisms for speech processing, plan-based dialogue understanding, incremental generation, and cost-based ambiguity resolution. These mechanisms are all fairly familiar and all seem plausible, although none is explained or illustrated adequately. Part of the problem is the focus on the implementation level, namely, the behavior of markers in a network—and these markers are quite intricate, including extra information such as propagation path constraints, addresses, and feature-structures—while the algorithms that the markers implement (shift-reduce parsing, for example) are mentioned in passing, if at all. Another problem is that discussion of how the mechanisms are integrated, or how well they worked together, is missing.

Chapters 5 and 6 describe two implementations of the system on parallel machines. Not only do they sometimes help to clarify previous discussion, they sometimes repeat it verbatim, or contradict it; and sometimes the relationship is less clear: for example, pseudocode for the basic marker-passing algorithm appears in three places (pp. 57, 124, and 147), slightly different each time, but with no indication as to whether the differences are significant.

Chapter 7 proposes some minor elaborations to Sato's memory-based translation model (Sato and Nagao 1990), and illustrates how it can translate simple sentences, including *You gave me a direction*. This chapter is an unintegrated conference paper; it neither builds on, nor casts light on, the rest of the book.

In general, the book is not very reader-friendly. It requires one to fill in gaps of logic and exposition; to make sense of passages like "we extend the idea of the semantic-oriented grammar to allow direct encoding a surface string sequence into a specific case of utterances" (p. 52); to understand Japanese example sentences appearing in kanji without transliteration or gloss; and to deal with typos, creative spellings, erroneous cross-references, and a skimpy index.

Finally, the wishlist. I am a big fan of mucking around with programs; I believe that a good research strategy is to take some half-baked ideas, implement them, experiment, improve, and iterate, until you know more than when you started. This book seemed, at first, to describe an exemplar of such exploratory systems-building. But the book never says what was learned in the course of building the system or observing its performance, nor how the system and the theory evolved in the light of experience. I was disappointed. Not only is such discussion lacking, but this book doesn't communicate how the system relates to other approaches, how it works, or even whether it works.

But maybe I expected too much. After all, on an emotional level, the book is fairly satisfying, at least to old-time AIers like myself. Many of us would like to believe that traditional models of syntax are irrelevant, that new computer architectures will solve our problems, that parallel natural language processing is the way to go, that hard work programming is the way to get there, and that the field is on the verge of greatness. What this book does is give flesh to these dreams, in the form of specific technical proposals and claims. It's just unfortunate that these claims are without support.

References

Hatazaki, Kaichiro; Noguchi, Jun; Okumura, Akitoshi; Yoshida, Kazunaga; and Watanabe, Takao (1992). "Intertalker: An experimental automatic interpretation system using conceptual representation." *1992 International Conference on Spoken Language Processing*, Banff, Canada, 393–396.

Rayner, Manny; Bretan, Ivan; Carter, David; Collins, Michael; Digaliakis, Vassilios; Gambäck, Björn; et al. (1993). "Spoken language translation with mid-90's technology." *Third European Conference on Speech Communication and Technology (Eurospeech '93)*, Berlin, 1299–1302.

Riesbeck, Christopher K. and Martin, Charles E. (1985). "Direct memory access parsing." Technical Report YALEU/DCS/RR 354, Department of Computer Science, Yale University.

Sato, Satoshi and Nagao, Makoto (1990). "Toward memory-based translation." *Proceedings, 13th International Conference on Computational Linguistics*, Helsinki, 247–252.

Tomabechi, Hideto (1987). "Direct memory access translation." *Proceedings, 10th International Joint Conference on Artificial Intelligence*, Milan, 722–725.

Tomabechi, Hideto; Kitano, Hiroaki; Mitamura, Teruko; Levin, Lori; and Tomita, Masaru (1989). "Direct memory access speech-to-speech translation." Technical Report CMU-CMT-89-111, Center for Machine Translation, Carnegie Mellon University.

Nigel Ward is the author of *A connectionist language generator* (Ablex, 1994). He is now interested in the role of syntax in speech understanding, among other things. Ward's address is: Mechano-Informatics, Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan; e-mail: nigel@sanpo.t.u-tokyo.ac.jp; URL: http://www.sanpo.t.u-tokyo.ac.jp/people/nigel.html.