

Readings in Machine Translation

Sergei Nirenburg, Harold Somers, and Yorick Wilks (editors)

(University of Maryland, Baltimore County, University of Manchester Institute of Science and Technology, and University of Sheffield)

Cambridge, MA: MIT Press, 2003,
xv+413 pp; ISBN 0-262-14074-8, \$55.00,
£36.95

Reviewed by
Elliott Macklovitch
Université de Montréal

This is a wonderful book—and not just for people actively involved in machine translation. Anyone with an interest in the history of computational linguistics will find much to relish and learn from in this weighty collection of articles past. Lest we forget, MT was one of the first nonnumerical applications proposed for the digital computer following the Second World War, and its often tumultuous 50-year history has had a significant impact on the entire field of computational linguistics. Indeed, this very journal can trace its lineage back to the journal whose original title was *Mechanical Translation*.

Though not a proper history of MT, *Readings in Machine Translation* is certainly a historical collection. The editors have sought to bring together in one volume “the ‘classical’ MT papers that researchers and students want, or should be persuaded, to read” (page xi) and that, alas, are often so difficult to find. For this alone, Nirenburg, Somers, and Wilks deserve our gratitude. The volume begins with the famous memorandum that Warren Weaver sent out to some 200 professional acquaintances in 1949, which is generally taken to mark the genesis of machine translation; and the most recent paper included dates back to the fourth MT Summit in 1993. The 36 articles that span the intervening period are said to constitute “MT’s communal inheritance.”

By what criteria were the papers selected? The editors cite three: personal taste; the aforementioned problem of availability, which is certainly a real one, particularly in the case of some of the early classics; and historical significance, which is said to be the main criterion for inclusion in the volume. The articles selected by the editors are supposed to represent “the most important papers from the past 50 years” of MT (p. xi). Well, as criteria go, that certainly sets a high standard! And yet many of these articles seem to meet it with ease. In addition to Weaver’s memorandum and a monumental state of the art published by Bar-Hillel in 1960, both of which absolutely must be read, there is a jewel of a piece written by Victor Yngve in 1957 that argues for what later became known as second-generation MT: systems that analyze the input text into an essentially syntactic intermediate representation that serves as the basis for transfer, rather than applying a bilingual dictionary directly to the input string, as first-generation systems did. There is Martin Kay’s “The Proper Place of Men and Machines in Language Translation”—an MT classic if ever there was one—in which the author derides the pursuit of fully automatic MT, not as a legitimate goal of basic research, but as a strategy promising a short-term solution to the burgeoning demand for translation; in its place, Kay proposes a modest, incremental program of machine aids for human translation: “little steps for little feet.” There is Makoto Nagao’s 1984 paper that launched the entire example-based MT paradigm (still very much alive and kicking), as well as a stellar piece by Jan Landsbergen laying out the advantages for

machine translation of Montague-style semantics (an approach all but moribund now, at least insofar as MT is concerned). One reads these papers today, decades after they were written, and one still cannot help but be impressed.

Needless to say, not all the articles included in *Readings in Machine Translation* come up to this high standard; that would be too much to expect. However, there are a fair number of papers that don't appear to even come close to the editors' stated selection criteria, unless of course one invokes the lame justification of personal taste. I won't bother to name names, out of respect for the elderly and the departed; but most of the papers I have in mind should be fairly obvious to all from a cursory perusal of the table of contents. ("Who is *that*?" is a good first indicator.) In other cases, one wishes the editors had made more liberal use of their prerogative to abridge. There are articles containing long tables filled with obscure codes and idiosyncratic terminology that can't possibly present any interest to the vast majority of contemporary readers.

Another reason for the excessive length of *Readings in Machine Translation* is that the book is divided into three distinct sections, each under the responsibility of one of the editors. The historical section is under Nirenburg's editorship and includes papers up to the late 1960s; Wilks's section is on theoretical and methodological issues; and Somers's is on system design. There are obvious overlaps between these divisions, in the sense that articles included in one section could just as well fit into another. The editors acknowledge this, and in itself it is not very serious. More tiresome, perhaps, is the fact that each section is prefaced by its own introduction (in addition to a common introduction to the entire volume) in which the editors sometimes marshal "their" articles in an attempt to argue for a certain perspective on machine translation. In his introduction, for example, Nirenburg cites numerous, often lengthy passages from the articles by the early MT pioneers that purportedly support his preferred approach to meaning-based MT. Well, maybe they do and maybe they don't; but either way, gathering grist for one's mill has a rather unseemly feel in this context.

A more serious criticism of *Readings in Machine Translation* is that the book is somewhat dated. This is a rather paradoxical charge for a collection of historical articles; what I mean by it is this: By the editors' own admission, the volume took much more time to bring to publication than they had originally anticipated. (In fact, I was sent a preliminary version by the publisher in 1999.) As a result, the editors' assessment of the most significant recent trends in MT is not entirely up to date. In the last few years, for example, there has been an impressive resurgence of activity in machine translation, particularly in the United States, where statistical methods drawn from speech recognition and various techniques borrowed from machine learning have proven remarkably successful. Had the editors been more aware of the profound impact of these new influences on the field, they would perhaps have modified their selection of articles. As it is, only two of the thirty-six papers in the collection explicitly address data-driven or statistical methods in MT: the seminal paper "A Statistical Approach to Machine Translation," published in 1990 by Peter Brown and his colleagues at IBM; and an earlier piece entitled "Stochastic Methods of Mechanical Translation" by Gilbert W. King.

Which brings me to my final criticism of this otherwise wonderful volume. Perhaps you recognized the name of Gil King (as Nirenburg calls him in his introduction), but I confess that I didn't. So I looked him up in John Hutchins's *Machine Translation: Past, Present, Future* and discovered (somewhat to my embarrassment) that King was the director of the project at IBM T. J. Watson Research Center in the late 1950s that eventually produced the Mark I system, later installed at the U.S. Air Force's Foreign Technology Division. Okay. . . . And where did the article included in this collection first appear? To find the answer to that question, and indeed to locate the source of

all thirty-six papers included in *Readings in Machine Translation*, one has to search the source notes that appear at the end of the volume: three dense pages of references whose order doesn't always correspond to that of the articles themselves. It would have been so much easier and more helpful to display this information on the first page of each contribution! Indeed, one wishes the editors had seen fit to include a short introductory note to each article, providing a few words of historical background on the author, or at least his or her affiliation at the time the paper was published.

But these are more or less minor quibbles, and they do not significantly detract from the value of this generous volume: generous in the size of its pages; generous in the quality of the paper and its cloth binding; generous above all in the intellectual caliber of the articles it offers for our delectation. In an age of skyrocketing book prices, MIT Press is to be commended for pricing *Readings in Machine Translation* so affordably.

References

Hutchins, John. 1986. *Machine Translation: Past, Present, Future*. Ellis Horwood, Chichester, England.

Elliott Macklovitch is coordinator of the RALI Laboratory at Université de Montreal and president of the Association for Machine Translation in the Americas (AMTA). His address is RALI-DIRO, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7; e-mail: macklovi@iro.umontreal.ca.