

Machine Translation

Theoretical and Methodological Issues

by **Sergei Nirenburg (ed.)**, in the series "Studies in Natural Language Processing", Cambridge University Press, 1987. \$17.95.

by Steven Krauwer

Nirenburg's book is a typical example of a "bottom-up book:" the editor starts out with a collection of interesting papers, and then tries to make a coherent book out of it. One should not underestimate the difficulty of this task: the bulk of the text is given and fixed, and the only weapons the editor can resort to are ordering, grouping, insertion of new chapters and writing an introductory

chapter to explain why this collection of texts is indeed a coherent and consistent whole. In very hard cases there is the additional possibility of adding an epilogue, in which these points from the introduction are repeated, just in case the reader had forgotten all about them. I wonder why Nirenburg didn't take the trouble to write such an epilogue.

The book contains 17 chapters,

most of which are revised versions of papers given at the Conference on Theoretical and Methodological Issues in Machine Translation (Colgate University, August 14-16, 1985). Those who own a copy of the proceedings of this conference (edited by Nirenburg) should therefore not expect to find many surprises.

According to the preface, the book is meant to be used either as a text, as a reference source for those who become machine translation researchers, or as a potential source of definitive statements about the various theories and methodologies currently employed in the field.

The book is, in Nirenburg's own view, naturally subdivided into 6 parts: I. The State of the Art in Machine Translation, II. Machine Translation and Linguistic Theory, III. Methodologies for Machine Translation, IV. Machine Translation and Artificial Intelligence, V. Research Tools for Machine Translation, and VI. Case studies of Machine Translation Projects.

Part I presents a statement of the problem of machine translation, the choices one has to make in order to arrive at tangible results (by Sergei Nirenburg), and an overview of the state of the art (Allen B. Tucker). It is quite informative for the newcomer in the field.

The title of Part II promises more than it contains: only one of the two chapters deals with the relation between linguistics and machine translation. The author (Victor Raskin) makes the observation that there is a big discrepancy between the goals of linguistics and the needs of natural language processing, and that linguistics has not

resulted in complete descriptions of any language. This is hardly contentious, but one wonders whether it is justified to completely ignore both the important role that is played by the descriptive devices developed by modern linguistics, and certain more recent developments in linguistics like LFG and GPSG, where there is a clear convergence of goals and needs of linguistics on one side, and natural language processing on the other.

The second chapter in Part II is dedicated to the significance of sublanguage for automatic translation (by Richard Kittredge). It is not clear to me why this chapter was put here, and in fact I am not sure whether it is an issue at all. This may very well be one of the symptoms of the bottom-up genesis of the book.

Considering the title of the book, one would expect Part III (Methodologies for Machine Translation) to be one of the central and most carefully prepared parts of the book. The human-machine interaction is taken up in 3 out of 5 chapters. It is certainly an important issue, but one chapter (e.g. an extended version of Alan Melby's contribution) would have done.

Knowledge-based translation is the other issue in this part of the book, in chapters by Jaime Carbonell and Masaru Tomita, and by Nirenburg, Raskin and Tucker. Again I would say that one single chapter going into more depth would have been preferable. Instead, much space is absorbed by a discussion of the interaction issue, and superfluous pictures and illustrations taken from the authors' own translation systems. It would have been much more interesting to focus on the central

question of how one arrives at non-arbitrary, extensible ways of organizing and describing knowledge.

The contribution on EUROTRA (Doug Arnold and Louis des Tombe) mostly deals with methodological issues that are internal to this project.

Part IV, Machine Translation and Artificial Intelligence, contains three chapters, each of which deals with what I would call a real issue: Processing ill-formed language (Ralph M. Weischedel and Lance A. Ramshaw), Discourse analysis (James Pustejovsky), and Natural language generation (David D. McDonald).

The content of the chapter by Weischedel and Ramshaw can be best characterized by the first word of the full title: reflections. Interesting and relevant, but not very deep. The chapter on discourse analysis is probably not easy to read for the uninitiated, but fortunately it contains many references. I found it interesting and qualitatively much better than the chapters on knowledge-based translation.

The last chapter, on generation, gives a good overview of the field, and will certainly convince newcomers in the field that generation is at least as difficult as analysis, and should not be underestimated in a machine translation project. I found this part of the book by far the best, and closest to what the title promises.

Part V, Research Tools for Machine Translation, could have been left out as far as I am concerned: there is a chapter on the research environment in the METAL project, and one on knowledge resource tools for accessing large text files. Many researchers can survive for quite a while without having

read them.

In Part VI we find the inevitable case studies. Although I think that it is extremely useful to know what is going on in the field, and to know how specific approaches lead to specific problems and solutions, I fail to see why such case studies should be part of this book.

My overall judgement is that the book has become a clear victim of the bottom-up syndrome: the collection of chapters is too heterogeneous and lacks coordination. It would have been better if Nirenburg had decided to leave out two thirds of the contributions, and to give stronger directives to the authors of the remaining parts.

In spite of all this I would still recommend it. Most contributions are well written (except in some cases where the reader has to swallow rather irrelevant details of specific systems). In addition, I feel very tempted to suggest to Sergei Nirenburg that he do another book. Here I do not mean just another book, but rather the book that is implicit in the 5 questions he raises in the introductory chapter:

1: What is the meaning of a text? 2: Does it have any component structure? 3: How does one represent the meaning of a text? 4: How does one set out to extract the meaning of a text?, and 5: Is it absolutely necessary to extract meaning (or at least all of the meaning) in order to translate?

Wouldn't that be a wonderful book (provided it were generated in a top-down manner)?

Steven Krauwer is a researcher in natural language processing, and is head of the Dutch research unit of the EUROTRA project.