A Road Guide to MT Research

by Tony Whitecomb

A SECOND OPINION ON HUTCHINS' RECENT BOOK

The fact that LT, in its first issue, gave attention to Hutchins' "Machine Translation — Past, Present and Future" (W.J. Hutchins, Ellis Horwood/ Halstead Press 1986) is of itself praiseworthy, but the actual review by Brian Dommisse was less than the book deserved. Let me explain why.

Machine Translation is a field of language technology with a history of four decades, during which a multitude of projects of various sizes and significance have come into existence. It is a history of little success and many failures, of repeated efforts to overcome persistent (semantic) barriers. The field is booming now, at least in terms of worldwide funding. And after years of taboo, it is again a subject which a scientist or a university professor is not ashamed to be involved with.

Until now, books about MT were limited in scope: monographs on individual researcher's theories or the particular approach in a specific system, sometimes furnished with a general MT introduction. Or one had to content oneself with chapters in AI or language books. Since the late 1970s, the handbook of Bruderer ("Handbuch der maschinellen und machinenunterstützten Sprachübersetzung," H.E. Bruderer, Saur KG, München 1978) was of some use, but this was largely an address book and a collection of project profiles on standard poll forms (with many entries, such as "memory used," which were of doubtful value).

Against this background, Hutchins' recent book appears to be the standard work the MT community has been waiting for. It is a reference work (except for addresses) and a textbook of high quality standards as to the completeness, precision and clarity of the material presented.

I strongly disagree with Brian Dommisse's criticism that Hutchins described "without attempting to shed light on their precise structures and workings" mere results of research instead of the research itself. The opposite is true: the book is an intricate account of MT exploratory efforts, rather than their results. This is quite in accordance with the simple fact that MT history, including recent, is one of many attempts and few products. Hutchins' description of these attempts is accurate, informative and at the same time entertaining.

The typical account of a project-group's work includes its particular approach, its goals and creeds, the claims and disclaimers made, insights won, crosslinks with the work of other groups, etc. Technical methods and structures (parsing techniques, tree transducers, transfer representations, sets of primitives) are described in sufficient detail and are adequately illustrated with examples. The essential principles are always made clear. Jargon is avoided or demystified by proper explanation; in only a few cases has Hutchins adopted an unfelicitous term of common use in MT (e.g. homograph resolution) — but has duly warned his readers.

Hutchins painstakingly documents his descriptions with numerous references and quotations, many of historic importance. This makes the book attractive to the scholar as well as the newcomer in the field. In addition, it is particularly interesting to read how certain techniques, now generally attributed to well-known names, can be traced back to lesser known predecessors: Kuno's Predictive Syntactic Analyzer as the embryonic ATN; the thesaurus approach by Masterman and Sparck Jones at CLRU as the jumping-off point for later work by Martin Kay and Yorick Wilks, and so on. In a field as complex as MT research, it is interesting to see how and when old ideas are revived.

That Hutchins is not involved in an MT project himself has obviously worked out to his advantage, not disadvantage: he does not address fellow-researchers, he is at the side of the reader, constantly observing and explaining. Yet, judging from the scope and quality of his book, he ought to be considered an MT expert rather than a "very well-informed layman," as Brian Dommisse suggests. He plays his role of independent expert extremely well, pointing out similarities, trends, gaps and prospects in MT research. His eloquent comments in the final chapter (such as "MT is a marriage of practical needs and theoretical idealism") are likely to be quoted often in the future.

Hutchins' book has only a few imperfections. Its coverage of Japanese projects is weak, at least compared to Western ones. There is no mention of Guzman's project in Bolivia, with Aymara as interlingua. Of course, the book reports only on work known as of mid-1985. The general increase in MT activity, especially in the last few years, will make supplements desirable from time to time: we can expect these in the form of invited papers at MT conferences, as Hutchins has done before.

Producing the standard work for the field is an endeavour that equals the contribution of a major research project: Hutchins has accomplished this feat for MT. In the recent past, some large MT projects have partly justified themselves by pointing at the raising of a generation of MT and computational-linguistics specialists. From now on, no serious student of MT who wants to know what came before can afford not to have "Hutchins" permanently on his desk. ●

Tony Whitecomb is the nom de plume of a software developer involved in natural language understanding and machine translation.