[Language Technology 9, Sept-Oct 1988]

TEACH YOURSELF MACHINE TRANSLATION

"Machine Translation: Linguistic Characteristics of MT Systems and General Methodology of Evaluation," Laurent Bourbeau and John Lehrberger (John Benjamins, Amsterdam and Philadelphia, 1988). 240 pages. Price: DFL 75 (US$ 32)

This book is based on reports prepared for the Canadian Secretary of State's Department from 1982 to 1984, at a time when that department was turning away from sponsoring MT development and toward evaluating commercial – and what are now referred to as "second generation" – MT systems.

It's designed around three chapters, dealing with three key aspects of MT: types of MT systems; their linguistic components; and their linguistic evaluation by the user.

The first two chapters would make a good teach-yourself intro to the field, illustrated as they are with plentiful, well chosen, English-French examples.

In chapter one, the authors deal with criteria to classify MT systems: degree of automation (batch, interactive, computer aids), depth of analysis (local vs full sentence), type of transfer (direct, indirect, interlingua), modularization, and domain dependency.

They make a strong case against local analysis – i.e. straight word substitution and local rearrangements – arguing that only full analysis of the sentence can yield consistently good results. This discussion gives ample opportunity for explaining the problems of MT and what solutions are available.

The second chapter describes the various linguistic components of MT systems. On the lexical level: a description of the content and structure of MT dictionaries, and a warning against "canned" idiom entries. And on pre- and postprocessing: how to handle various typographic features and break up text into significant processing units (words and sentences).

The section on morphology provides a clear overview of inflectional, derivational, and compositional morphology. The one on syntax deals with problems of homography, scope of conjunctions, and prepositional attachment, with special emphasis on complex noun phrases, illustrated by some stunning examples taken from the TAUM-Aviation corpus.

The semantics section discusses how semantic features are created, interrelated, and used for solving syntactic problems such as selecting alternative translations.

Throughout these two chapters, the authors also raise a number of design issues: modularization and sequencing of the analysis, transfer, and synthesis components; how to establish the boundaries between linguistic levels (for instance, should compound words be handled by morphology or syntax?); what to do when a complete sentence parse is not achieved (a typical TAUM issue); and how to extend a domain-based system to other domains (again a typical TAUM issue, since both TAUM's Météo and Aviation systems were designed for a particular sublanguage).

The authors also point out several limitations of second generation systems, such as sentence-limited context analysis and the absence of actual semantic representation of sentences. They fail to go on to discuss third generation technology – which, granted, is of secondary importance, since the book is about evaluating existing systems, none of which have yet broken through the third generation barrier.

The most salient feature of the book's last chapter – on system evaluation – is a discussion on how the user can feed sentences to a "black box" (a commercial MT system whose developer denies access to the grammars), analyze the output, and use it to get a fair idea of how the box's insides operate. If the user wants to know the system's real potential for growth – as opposed to what the vendor claims – this sort of insight is essential.

The book enumerates an impressive array of evaluation criteria, including an appendix reviewing various evaluation methods already in use, though it doesn't explain how these methods might be combined effectively. The authors rightly stress that evaluation is always specific to a particular situation – system + user + corpus of texts – and that a "cookbook approach" would be far too simplistic.

While I agree substantially that the right answer in MT is usually "It all depends," I still feel that there are dynamic interrelations between the various criteria that could be brought to light. In any case, whoever wants to pick up on system evaluation from there, can at least use this book as a comprehensive and authoritative starting point.

All in all, Bourbeau and Lehrberger have done a good didactic job in describing the intricacies of machine translation and demonstrating how multifaceted and complex its evaluation is. And any potential user will learn a lot – in terms of both wisdom and wariness – from it.

— Claude Bédard