SUDDENLY IT'S DISCOURSE ANALYSIS

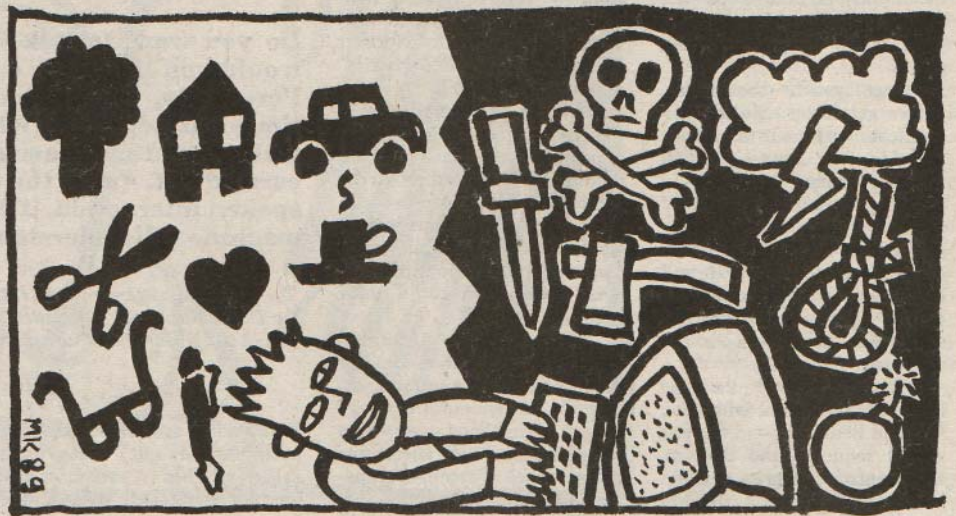ILLUSTRATION BY MAX KISMAN

"Text Cohesion in Translation," Bart Papegaaij and Klaus Schubert (Foris Publications, Dordrecht, Netherlands/Providence, RI, USA, 1988) Price: US$15

By Claude Bédard

Most second generation MT systems have never attempted to cross sentence boundaries, having their hands already full with sentence-level grammar. They therefore typically translate texts as if they consisted of bunches of unrelated sentences. (And most of them don't even do that particularly well.)

Of course, we all know that every text has its own organization and is filled with pointers that relate sentences and words into a broader picture, and that a proper translation should respect this fact. But then, how do we formalize this process for a machine?

This third book in an MT series, by two members of the Dutch-based DLT machine translation project, gives an outlook on how it *might* be done, drawing on the still scarce literature in the field of discourse analysis (aka text linguistics or text grammar).

The problems addressed in most detail by Papagaaij and Schubert are pronominal reference (finding the right

antecedent for pronouns) and content word reference (various words or expressions in the text can refer to the same entity of meaning).

The main types of clue examined include: thematic progression, or how concepts flow from sentence to sentence; theme-rheme, where each sentence says something (or a rheme) about something else (a theme) and those elements can be identified as to their positioning in the sentence; focus, or the essential point of a sentence; logical connectives (words such as "therefore," "but," "moreover," etc.); and grammatical features such as the role of the definite article in referring back to previous occurrences of the same entity.

A refreshing aspect of the book is that the authors actually bother to consider how real people operate: how the writer creates text coherence (economy vs explicitness, new vs shared knowledge, etc.) and how the reader understands it (expectations, short-term and long-term memory).

Even the translator is granted some attention – a rare attempt in computational linguistics literature. Papagaaij and Schubert look at how text coherence is conveyed, and sometimes recreated, in the target language, and by what means, including various compensations due to differences in the linguistic material available in each language.

In their description of how high-quality human translation is achieved, the authors reach for the level of FAHQT (fully automatic high-quality translation).

The main part of the book is straightforward and concrete, and clearly aims at popularization. It is based on a single real-life sample text rather than on a series of "forged" examples, which gives more credibility to the description – at the price of demonstrating somewhat less successfully the principles discussed.

The book's final chapter deals with the intricate issue of how to integrate text coherence features into machine translation. The level of discussion is more abstract than in the rest of the book, ending on the interesting question of how much understanding (including pragmatics) is necessary for machine translation.

All in all, a stimulating introduction to discourse analysis in machine translation.

Claude Bédard is a Montréal-based freelance translator and MT consultant.