TEXT COHERENCE IN TRANSLATION

Bart Papegaaij and Klaus Schubert
(BSO/Research, Utrecht)

Reviewed by
Chrysanne DiMarco
University of Toronto

Machine translation (MT) has focused on the problems of syntax and semantics at the sentence level, but the real goal of MT is to translate texts, a fact that has been generally overlooked. There is a crucial difference between a text and a set of unrelated sentences, and in MT, one must avoid destroying the former by translating it into the latter. It is the coherence of text in particular that Papegaaij and Schubert address. They aim to examine the role of text phenomena in machine translation, to assess the feasibility of a number of suggested models of text coherence for MT, and to propose solutions.

The book consists of three chapters: introductory material, a chapter on the clues and devices of text coherence, and a chapter on text coherence in translation.

The introductory chapter (11 pages) summarizes the main approaches to text linguistics, reviews the Distributed Language Translation project (with which the authors are associated), and introduces the relevant terminology. Chapter 2 (139 pages) is the background chapter for understanding the techniques that can be used to render text coherent. It begins with a study of the kinds of decisions involved in a sample translation, and then surveys the maintenance of text coherence through deictic reference, word disambiguation (by means of a shared context), thematic progression, the structure-building properties of verbs, and, on a higher level, rhetorical patterns. This review of diverse devices for text coherence is illustrated by showing how they contribute to the coherence of a sample text. This chapter covers essentially the same ground as Halliday and Hasan (1976), but in less detail and with a slant to translation.

In Chapter 3 (41 pages), the authors follow up on this review to consider text coherence from the standpoint of translation. A systematic summary of the coherence devices discussed earlier is given, grouped under the headings: coherence of entities, coherence of focus, and coherence of events. This chapter also focuses more explicitly on the question of how to maintain text coherence in a MT system.

The strength (and the bulk) of the book is the quite thorough discussion of the text coherence devices that are relevant in MT. A reader unfamiliar with the prob-

lem of maintaining coherence in translation will receive a useful tutorial in the possible approaches. However, it should be noted that, as the authors themselves point out, "the reader [may be] somewhat unsatisfied and [have] a feeling of having been offered just another series of sample analyses and just another, sketchy, model of text structure. This may appear to be so, at least to a casual reader . . . [However], it has . . . been our aim to make some steps towards (preliminary) implementation possible right now, although we are aware of the fact that much more research of both a fundamental and an application-oriented nature remains to be done" (p. 198). The book will therefore be of interest to readers who want an introduction to a difficult problem in MT but who recognize that, as of now, solutions remain preliminary and still theoretical.

REFERENCE

Halliday, M.A.K. and Hasan, Ruqaiya 1976 *Cohesion in English.* London: Longman.

Chrysanne DiMarco is a doctoral student in the natural language understanding group, University of Toronto. Her thesis research is in stylistics in machine translation. DiMarco's address is: Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4. E-mail: cdi@ai.toronto.edu

---

WORD EXPERT SEMANTICS: AN INTERLINGUAL
KNOWLEDGE-BASED APPROACH

Bart C. Papegaaij, Victor Sadler, and A.P.M. Toon
Witkam (eds.)
(BSO/Research, Utrecht)

Reviewed by
Jonathan Slocum
Symantec Corporation

This book describes the Dutch firm BSO's machine translation system DLT at a very early stage in its development. The book is organized into four parts, each with its separate set of chapters. Part I is an introduction to the problems of NLP in general, and MT in particular, for the neophyte. Part II describes the "Semantic Word Expert System" in some (not exhaustive) detail, describing its workings primarily in the future tense. Part III describes the "Semantic Work Bench" development tool that is being developed for development of the DLT lexicons. Part IV discusses "Future Developments" in DLT, all the way up to "The Ultimate Aim," which turns out (p. 207) to be a system for multidirectional machine translation, multilingual information retrieval and document indexing,