

Machine-aided Translation and Lexical Strategies

Frank Knowles

University of Aston
Birmingham, UK.

ABSTRACT

The context of this paper is that of a translator wishing to develop dictionaries for the purposes of machine-aided translation (MAT). A description is given of the ways in which lexical items in running text are statistically "patterned", depending on whether these so-called "types" are left unaltered as they are extracted from the text or whether they are immediately mapped onto the corresponding dictionary look-up form ("lemma") for the purpose of statistical analysis. It is obvious, of course, that for translation purposes it is necessary to establish appropriate entry-points into the MAT dictionary, but this is a secondary problem.

There are two dimensions which can assist the machine-assisted translator to a considerable extent. One such factor is any degree of homogeneity - the greater, the better - in the texts he wishes to process. Translators specialising in certain subject areas and types of discourse are at an advantage if they wish to use an MAT system. The second factor is that of the so-called "multi-word unit". Although all languages have multi-word units, which are semantically atomic, they are particularly important in English, and even more so in English technical terminology. Frequency studies of multi-word units, although they generate large listings of types, can be very useful for MAT.

The machine-assisted translator is faced with the need to view his work as consisting of two distinct modes: dictionary elaboration and text transaction. The second mode, of course, provides important feed-back to guide the first. One thing is clear: the translator must be his own lexicographer to a great extent, at least until the time when software houses realise the commercial value of such "static" data as general bi-lingual high-frequency dictionaries and the potential "constellation" of carefully designed and delineated bi-lingual glossaries of technical terminology!