Chapter 3

EBMT IN A CONTROLLED ENVIRONMENT

Reinhard Schäler
Reinhard.Schaler@ul.ie

Andy Way
away@computing.dcu.ie

Michael Carl
carl@iai.uni-sb.de

Abstract One key to the success of EBMT is the removal of the boundaries limiting the potential of translation memories (TMs). We discuss a linguistically enhanced TM system, a Phrasal Lexicon (PL), which takes advantage of the huge, underused resources available in existing translation aids. We claim that PL and EBMT systems can provide valuable translation solutions for restricted domains, especially where controlled language restrictions are imposed. When integrated into a hybrid and/or multi-engine MT environment, the PL will yield significant improvements in translation quality. We establish a future model of translation usage and anticipate that EBMT and the PL will have a central place in future hybrid integrated translation platforms.

Keywords: translation memory, phrasal lexicon, controlled translation, hybrid MT

1. Introduction

It is right for the wider MT community to acknowledge that there have been a number of success stories where rule-based MT (RBMT) systems have provided a general solution to the problems of translation. Examples include the Météo system (Chandioux, 1976) and PAHO's
ENGSPAN and SPANAM (Vasconcellos & León, 1985) systems, to name but two.

However, at the same time it is generally agreed that the overall quality of today's MT systems is somewhat short of what might be desired. Nevertheless, translators remain wary that MT software might pose a considerable threat to well established work practices, despite clear evidence to the contrary. Indeed, it is hard to know how such entrenched views may be overcome, despite the best efforts of MT educators and researchers to level the playing field in this emotive issue. It is clear that RBMT can be a solution: the problem is in knowing in what circumstances. We will address this issue in more depth later in the paper.

Despite the widespread view that RBMT will never be good enough to warrant serious consideration as systems capable of high quality general purpose translation, research and development in RBMT systems continues to this day. Furthermore, their deployment, especially in limited domains, cf. PaTrans (Ørnes et al., 1996), is growing, and real savings are being made where business professionals are prepared to keep an open mind and take the time to see where RBMT can be of use, and where it (probably) cannot.

At the same time, Translation Memory (TM) systems have rapidly come to be regarded as an extremely useful tool in the translator's armoury. Notwithstanding the widespread acceptance of such tools, the emergence of TM applications have continued to keep some translators on their guard. This is even harder to understand: TM systems do not translate. All they do is find close matches for the input string in their database of previously seen translations and display these matches together with their translations for the translator himself to manipulate into the final, output translation. At all stages in the translation process, the translator is the integral figure: he is free to accept or reject any suggested matches, and may or may not insert any suggested translations into the target document and the TM itself wherever he deems this to be appropriate.

In the view of some researchers, TM technology may be considered as some sort of sophisticated search-and-replace engine. For example, Macklovitch & Russell, 2000 liken TM systems to information retrieval tools, in that all the translator does in effect is search for 'documents' (aligned sentence pairs) which may help translate a given sentence. The TM system formulates the query itself from the input string to be translated, i.e. in most cases, the query is the source language sentence, while