

# Dynamic Translation Memory: Using Statistical Machine Translation to improve Translation Memory Fuzzy Matches

Ergun Biçici<sup>1</sup> and Marc Dymetman<sup>2</sup>

<sup>1</sup> Koç University (Istanbul, Turkey)  
ebicici@ku.edu.tr

<sup>2</sup> Xerox Research Centre Europe (Grenoble, France)  
marc.dymetman@xrce.xerox.com

**Abstract.** Professional translators of technical documents often use Translation Memory (TM) systems in order to capitalize on the repetitions frequently observed in these documents. TM systems typically exploit not only complete matches between the source sentence to be translated and some previously translated sentence, but also so-called *fuzzy matches*, where the source sentence has some substantial commonality with a previously translated sentence. These fuzzy matches can be very worthwhile as a starting point for the human translator, but the translator then needs to manually edit the associated TM-based translation to accommodate the differences with the source sentence to be translated. If part of this process could be automated, the cost of human translation could be significantly reduced. The paper proposes to perform this automation in the following way: a phrase-based Statistical Machine Translation (SMT) system (trained on a bilingual corpus in the same domain as the TM) is combined with the TM fuzzy match, by extracting from the fuzzy-match a large (possibly gapped) bi-phrase that is dynamically added to the usual set of “static” bi-phrases used for decoding the source. We report experiments that show significant improvements in terms of BLEU and NIST scores over both the translations produced by the stand-alone SMT system and the fuzzy-match translations proposed by the stand-alone TM system.

## 1 Introduction

Translation Memory (TM) systems [1, 2] have become indispensable tools for professional translators working with technical documentation. Such documentation tends to be highly repetitive, due to several factors, such as multiple versioning of similar products, importance of maintaining consistent terminology and phraseology, and last but not least, simplification of the translation process itself. TM systems typically exploit not only *complete matches* between the source sentence to be translated and some previously translated sentence, but also so-called *fuzzy matches* [3], where the source sentence has some substantial commonality with a previously translated sentence. These fuzzy matches can be