A review of the PRESEMT project

George Tambouratzis, Marina Vassiliou, Sokratis Sofianopoulos
ILSP/Athena R.C., 6 Artemidos & Epidavrou Str., Athens, 15125, Greece
{giorg_t; mvas; s_sofian}@ilsp.gr
www.presemt.eu

Description

PRESEMT (Pattern REcognition-based Statistically Enhanced MT) was a project funded by the EU under the FP7 topic "ICT-2009.2.2: Language-based Interaction". PRESEMT has led to a flexible and adaptable Machine Translation (MT) methodology, overcoming well-known problems of existing MT approaches, e.g. compilation of extensive bilingual corpora.

In order for PRESEMT to be easily amenable to new language pairs, only relatively inexpensive, readily available language resources, in the form of monolingual corpora as well as bilingual lexica, are used. Since for the majority of language pairs the amount of available parallel corpora is very limited, PRESEMT extracts modelling information from monolingual resources. Only a very small bilingual corpus is used to provide information for structural modifications from SL to TL. Translation context is modelled on syntactic phrases, as they have been proven to improve the translation quality. Phrases are produced via a semi-automatic and language-independent process of morphological and syntactic analysis, removing the need for compatible, in terms of output, NLP tools per language pair. So a flexible MT system has been developed, which is enhanced with pattern recognition techniques supporting the development of a language-independent analysis.

To compare PRESEMT to other MT methodologies, both objective and subjective evaluations have been performed. All evaluation tasks have shown that PRESEMT has a lower yet comparable translation quality to that achieved by established MT systems (Google Translate, Bing Translator). This is expected as the proposed methodology by design avoids incorporating a priori grammatical knowledge and uses only inexpensive resources from which to extract knowledge as well as publicly available tools (such as parsers and taggers). The evaluation outcome also reflects the much shorter development time available for PRESEMT. However, PRESEMT consistently generates translations which convey the intended meaning of input text. As such, PRESEMT can be considered successful on the basis of its design brief, namely generating translations suitable for gisting. Currently, work is continuing on a number of aspects so as to improve further the PRESEMT accuracy, with latest results indicating a dynamic for advancing the translation quality.

References.

Sofianopoulos, S., Vassiliou, M. & Tambouratzis, G. (2012) Implementing a language-independent MT methodology. Proceedings of the First Workshop on Multilingual Modeling, held within the ACL-2012 Conference, Jeju, Republic of Korea, 13 July 2012, pp.1-10.

Tambouratzis, G., Sofianopoulos, S. & Vassiliou, M. (2012) Evaluating the translation accuracy of a novel language-independent MT methodology. In Proceedings of the COLING-2012 Conference, Mumbai, India, 9-14 December 2012, pp.2569-2584.

Sima'an, K., Forcada, M.L., Grasmick, D., Depraetere, H., Way, A. (eds.)
Proceedings of the XIV Machine Translation Summit (Nice, September 2–6, 2013), p. 437.