Foreword

Following two successful EBMT workshops in 2001 at the MT Summit VIII in Santiago de Compostela, Spain, and in 2005 at the MT Summit X in Phuket, Thailand, this is the third workshop of its kind. Many things have happened since 2005. The last few years have witnessed a decline in example-based machine translation (EBMT) research and statistical machine translation (SMT) has almost completely taken over the corpus-based machine translation arena, with many EBMT practitioners moving into hybrid approaches integrating EBMT with other approaches, mostly (but not only) SMT. Not having a clear definition of what EBMT is has also contributed to this lack of visibility. In fact, research that would have been considered EBMT has been published without the EBMT label.

Is the success of SMT due to the fact that it is the best way to do corpus-based machine translation or is it because many SMT software packages are readily available to researchers under free/open-source licences that allow use as well as collaborative improvement? Shouldn't EBMT practitioners start to think about putting together their tools, their engines and their data and releasing them under open licenses to extend their use both in academia and industry?

The pressure on machine translation researchers to prove their results through detailed empirical evaluation is growing. But the validity of empirical results hinges on reproducibility. Turning our experimental research into packages and tools that other researchers can use and improve is a challenge but it is not infeasible as SMT practitioners have shown.

The response of the community to the call for papers for this conference was very encouraging. We received fifteen papers addressing the main theme or other aspects related to a potential strengthening of EBMT and its real-world applications. Of these, 11 have been accepted for presentation at the workshop.

The papers have been grouped in four sessions.

In the first session, three papers present hybrid approaches to EBMT. James Smith and Stephen Clark present a new EBMT–SMT hybrid, Felipe Sánchez-Martínez and colleagues show an EBMT–RBMT hybrid, and Declan Groves and colleagues evaluate syntax-driven approaches to phrase extraction for machine translation.

The second session groups papers addressing the main theme of the Workshop, featuring free/open-source EBMT software: Aaron Phillips and Ralf Brown describe their Cunei MT platform, David Farwell and Lluís Padró describe the Freeling analyser suite and outline possible applications in EBMT, and Adrien Lardilleux and colleagues discuss the sampling-based alignment performed by their software.

The third session deals with what one could call "Pure EBMT". Harold Somers and colleagues review the use of proportional analogies in EBMT, Maarten van Gompel and colleagues describe the extension of their memory-based translation formalism to phrases, and Vincent Vandeghinste and Scott Martens describe their top-down transfer strategy to EBMT.

Finally, the fourth session deals with applications of EBMT. Julien Gosme and colleagues discuss the translation of sublanguage using subgrammars and Marian Flanagan describes the use of EBMT to translate film subtitles.

The papers in these proceedings will therefore give a wide but focussed view of the current status of EBMT research and seem to hint at a revival of the whole field after years of languishing.

I will finish this foreword by thanking the authors who sent papers; the reviewers, who did an excellent work and gave many suggestions to improve the papers that were finally accepted; and the local organizers. I thank also the European Association for Machine Translation, the Office of the Vice-President for Research and the Centre for Next Generation Localisation of Dublin City University, and Science Foundation Ireland for their crucial support.

Dublin, November 2009

Mikel L. Forcada
Workshop co-chair
CNGL, School of Computing,
Dublin City University,
Dublin, Ireland.