Pangeanic in the EXPERT Project:
EXPloiting Empirical appRoaches to Translation

Manuel Herranz, Alex Helle, Elia Yuste, Ruslan Mitkov, Lucia Specia
{m.herranz,a.helle,e.yuste}@pangeanic.com, r.mitkov@wlv.ac.uk,
l.specia@dcs.shef.ac.uk

EXPERT, a 48-month project funded by the People Programme (Marie Curie Actions) of the EU's 7th Framework Programme FP7/2007-2013/ under REA agreement nº 317471, aims to train young researchers to promote the research, development and use of hybrid language translation technologies. In practice, EXPERT aims at improving translation practices and enhancing the productivity of relevant actors in the translation market. EXPERT findings will set an agenda for new skills and jobs.

Promoting new job profiles based on empirical data from LSPs and academia. The true potential of MT remains to be exploited as a result of non-user-friendly interfaces, lack of awareness of translator's feedback, etc. Pangeanic has already created a web-based tool that is able to organize by domain, maintain and clean training material (TM) for MT. The tool already incorporates hybrid features, and these will be tested, expanded and improved upon during EXPERT.

Pangeanic will contribute to EXPERT at different project stages by (co-)hosting/supervising a number of post-graduate and post-doctoral researchers during their own research projects. These will focus on the implementation and evaluation of the new hybrid computer-aided translation technology, the investigation of translators' requirements from translation technologies, and the confidence estimation of corpus-based approaches to translation. During the project, Pangeanic will give unlimited access to its entire computing infrastructure and resources, including our massive data repositories and statistical language models, DIY platform, automated retraining features, etc.

Pangeanic will focus on the automatic generation of bilingual written texts for multiple language combinations, alignment, segment cleaning and selection for bilingual engine building. We will also look at what hybridisation language technology techniques need to be incorporated and improved to tackle reordering issues and other linguistic phenomena in unrelated languages. When dealing with language-specific issues, we will also delve into automatic quality metrics and how these can correlate to human, non-objective qualitative appreciations.

Pangeanic is one of the industry partners of the EXPERT Project Consortium, led by the Research Group in Computational Linguistics of the University of Wolverhampton, UK. Other partners include: Universidad de Málaga, Spain; University of Sheffield, UK; Universitaet des Saarlandes, Germany; Translated srl, Italy; Dublin City University, Ireland; Hermes Traducciones y Servicios Lingüísticos, SL, Spain; and Universiteit van Amsterdam, Netherlands. We shall be happy to represent the consortium and start disseminating initial project actions items and future work.