Invited talks

Matxin: developing sustainable machine translation for a less-resourced language

Kepa Sarasola, Euskal Herriko Unibertsitatea, Spain

Following the strategy defined in IXA group for reusing linguistic resources and NLP tools, in year 2000 (but not before), we decided that we had enough languages resources and tools (bilingual dictionaries, morphological and syntactic analysers and parsers) that could be reused to build an RBMT system for the Spanish–Basque pair. The system built is called Matxin and is available at `matxin.sourceforge.net`. Since 2006 we are collaborating with DCU building a Spanish–Basque system based on EBMT and SMT paradigms. We could get better results with bigger parallel corpus, but it is difficult to get it for Basque, a minority language. Based on our work we have published a strategy for sustainable MT for lesser-resourced languages; it is based on incremental design, reusability, standardisation and open source. We have developed MT engines based on the three paradigms (RBMT, SMT and EBMT), so our position is optimal to experiment with hybrid systems and multi-engine systems.

Anusaaraka: An accessor cum machine translator

Amba Kulkarni, University of Hyderabad, India

India being multilingual, there is a demand for translation both among Indian languages as well as from English to Indian languages. Translation being not reliable, Anusaaraka aims to provide complete access to the source text in addition to translation. With an appropriate division of load between man and machine, Kannada-Hindi Anusaaraka, developed in early 90s, demonstrated that it is possible to reduce the language barrier considerably. However it is necessary for an Anusaaraka reader to undergo some training on the syntactic divergences and special notation used to handle the divergences in word-meaning mappings between the source and the target language. In the later version of Anusaaraka, in order to reduce the burden on a user, the state-of-the-art MT system formed an important component of it. Care was taken to develop the architecture in such a way that, it can cater to the needs of diverse requirements ranging from faithful access to the full fledged translation.