

CRL at New Mexico State University
Stephen Helmreich
Computing Research Laboratory
New Mexico State University
Las Cruces, NM 88003 USA

Reaching the goal of making digital libraries accessible to scholars and users around the world will require translation capabilities. Users, speaking many different languages, will want to access libraries whose contents are also in many different languages.

At the level of document retrieval, translation capabilities would be useful for retrieving appropriate documents and for helping the user judge the relevancy of the retrieved material.

Even if document retrieval can be accomplished through methods short of full translation, such as keyword correspondences, or by other methods, such as semantic indexes, once a relevant document has been retrieved, it must be translated in order to be of use, if the language(s) of the user and the language of the document are different.

Over the last dozen years, the Computing Research Laboratory (CRL) at New Mexico State University in Las Cruces, NM, has specialized in multilingual information technologies, including machine translation, information retrieval, information extraction, summarization, and combined applications.

The CRL brings to the table a variety of resources and skills that could be of use in Korean-US collaboration on digital libraries.

First, it offers a variety of approaches to translation. A knowledge-based, interlingual approach to translation offers in-depth, high-quality translation. This approach uses a large ontology and a lexicon that is hand-tuned for the data. See <http://crl.nmsu.edu/Research/Projects/mikro/index.html> for information on this project. Such an approach would be useful for smaller applications where accuracy and quality are paramount. Lexicons for Spanish, English, and Chinese are in existence or under development.

More recently, the CRL has also been developing translation approaches geared towards rapid development of systems for languages with limited on-line linguistic resources. The machine translation system developed within the Temple, TIDE, and Corelli projects and currently under further development for incorporation within the Expedition system, is a set of engines that, given appropriate input and data, can be rapidly configured into a transfer MT system (Zajac and Amtrup, 2000). With the provision of an interlingual dictionary, a morphological analyzer, and a grammar, translation of reasonable quality can be produced from a source language into English. Such a system would be useful for rapid translation of larger sections of text, where complete accuracy and fluency are not absolutely required. For Korean, CRL has both the dictionary and the morphological analyzer, and the grammar is in a beginning stage.

At the lowest level, the CRL can already provide a word-for-word “translation” from Korean into English. Such low-quality translation would be primarily useful in document retrieval systems or where only a gist is required. Such a system is currently available for Korean to English as part of CRL’s Keizai project (Ogden et al, 1999), an inter-active multi-lingual document retrieval system, with summarization and translation capabilities.

In addition to its capabilities in machine translation, other aspects of CRL’s research could be utilized within the framework of a Korean-US digital library project. The multi-lingual document retrieval system could be of obvious assistance within a digital library retrieval system. CRL’s summarization programs could also be of use in a digital libraries project. The ontology used within the interlingual MT system (Mahesh and Nirenburg, 1995) on its own would be a valuable resource for semantic indexing.

References

These and many other relevant papers are available for downloading at the CRL website: <http://crl.nmsu.edu/> (Look under the heading “Research” and then under the various projects: Mikrokosmos, Corelli, URSA.)

Mahesh, K. and S. Nirenburg. 1995. A Situated Ontology for Practical NLP. In *Proc. Workshop on Basic Ontological Issues in Knowledge Sharing, International Joint Conference on Artificial Intelligence (IJCAI-95)*, Aug. 19-20, 1995. Montreal, Canada.

Ogden, W., J. Cowie, M. Davis, E. Ludovik, S. Nirenburg, H. Molina-Salgado, N. Sharples. 1999. *Keizai: An Interactive Cross-Language Text Retrieval System*. Paper for the Workshop on Machine Translation for Cross Language Information Retrieval, held in conjunction with the MACHINE TRANSLATION SUMMIT VII, September 13-17, 1999, Singapore.

Zajac, R. and J. Amtrup. 2000. "Modular Unification-based Parsers". *Proceedings of IWPT2000, the 6th International Workshop on Parsing Technologies*. Trento, Italy, February 2000.

Biographical Note

Stephen Helmreich has been employed as a computational linguist at the Computing Research Laboratory for just over 12 years and has worked on a variety of projects there, including Machine Translation projects in German, Spanish, and Serbo-Croatian, automatic grammar induction, Collative Semantics, Pragmatics-based Machine Translation, and Metaphor-based Reasoning.

He is currently working on the Expedition project, including development of grammars for English and Spanish and of web pages for elicitation of source language information on closed-class lexical items. He received his Ph.D. in linguistics from the University of Illinois in 1996,