SALTMIL SIG: Speech and Language Technology for Minority Languages

On May 23rd 2006, SALTMIL held a morning workshop on "Strategies for developing machine translation for minority languages". This was a satellite workshop preceding the biennial LREC (Language Resources and Evaluation Conference) in Genoa, Italy, and was chaired by Briony Williams. It was the latest in the series of SALTMIL workshops held as satellites to the biennial LREC conference.

The program was very full, beginning with seven invited talks, each of which stimulated many questions and discussions. This was then followed by a poster session, with sixteen contributed poster papers. About fifty people were present in total, from a wide range of countries, and representing work on a variety of minority languages.

One of the highlights was a talk by Steven Krauwer (University of Utrecht, Netherlands), the originator of the BLARK concept (Basic Language Resource Kit). His talk, entitled "ENABLER, BLARK, what's next?", focussed on recent developments such as the CLARIN initiative (see http://www.mpi.nl/clarin/) - Common Language Resources and Technology Infrastructure. This is a large-scale pan-European collaborative effort to create, coordinate and make language resources and technology available and readily useable. It is not a project proposal, but rather a proposal for a research infrastructure to be included in the European Roadmap for research infrastructures. He also introduced the concept of the "Blarkette" (a mini-BLARK) for languages with few digital resources.

The other speakers gave talks on the following topics:

* "The BLARK matrix and its relation to the language resources situation for the Celtic languages"
* "Building NLP systems for two resource-scarce indigenous languages: Mapudungun and Quechua"
* "Open source machine translation: an opportunity for minor languages"
* "Approaching a new language in machine translation"
* "Unicode Development for Under-Resourced Languages".
* "Statistical Machine Translation with and without a bilingual training corpus"

One of the participants recorded her impressions of the event as follows:

* There is an optimistic outlook for lesser-resourced languages.
* BLARK is very useful for explaining our needs to policy-makers and funders.
* There is a need to invest in basic language resources before end-user applications can be developed.
* Perhaps a Blarkette is appropriate for languages starting from nothing.
* We need to monitor developments in CLARIN.
* Great savings in time and money can be gained by piggy-backing off a closely-related language which is better resourced, and also by leveraging the work which has gone into a major language which is commonly paired with the lesser-resourced language
* A half day was sufficient for the workshop.
* The speakers were very high-quality in terms of content and delivery.
* The workshop was well-organised.
* I found it very beneficial!

Contributed by Briony Williams (University of Wales, Bangor, UK)
b.williams@bangor.ac.uk