

Strategies for developing machine translation for minority languages

5th SALTML Workshop on Minority Languages

Tuesday May 23rd 2006 (morning)

Magazzini del Cotone Conference Centre, Genoa, Italy

Organised in conjunction with LREC 2006: Fifth International Conference on Language Resources and Evaluation, Genoa, Italy, 24-26 May 2006

Workshop Programme

8:00 Registration and preparation of posters

8:15 Welcome: Briony Williams (University of Wales, Bangor, UK)

8:30 Stephen Krauwer (University of Utrecht, Netherlands)
"ENABLER, BLARK, what's next?"

9:00 Delyth Prys (University of Wales, Bangor, UK)
"The BLARK matrix and its relation to the language resources situation for the Celtic languages"

9:30 Christian Monson, Ariadna Font Llitjós, Roberto Aranovich, Lori Levin, Ralf Brown, Eric Peterson, Jaime Carbonell & Alon Lavie (Carnegie-Mellon University & University of Pittsburgh, USA)
"Building NLP systems for two resource-scarce indigenous languages: Mapudungun and Quechua"

10:00 Mikel Forcada (University of Alacant, Spain)
"Open source machine translation: an opportunity for minor languages"

10:30 Anna Sågvall Hein & Per Weijnitz (University of Uppsala, Sweden)
"Approaching a new language in machine translation"

11:00 COFFEE

11:30 Daniel Yacob (Director, Ge'ez Frontier Foundation)
"Unicode Development for Under-Resourced Languages".

12:00 Maja Popovic & Hermann Ney (RWTH Aachen University, Germany)
"Statistical Machine Translation with and without a bilingual training corpus"

12:30 POSTER SESSION

Tod Allman & Stephen Beale, "A Natural Language Generator for Minority Languages".

Saba Amsalu & Sisay Fissaha Adafre, "Machine Translation for Amharic: Where we are".

Carme Armentano-Oller and Mikel Forcada, "Open-source machine translation between small languages: Catalan and Aranese Occitan".

Arantza Casillas, Arantza Díaz de Illarraza, Jon Igartua, R. Martínez & Kepa Sarasola, "Compilation and Structuring of a Spanish-Basque Parallel Corpus".

Bostjan Dvorák, Petr Homola, Vladislav Kubon "Exploiting Similarity in the MT into a Minority Language".

Adrià de Gispert and José B. Mariño, "Catalan-English Statistical Machine Translation without Parallel Corpus: Bridging through Spanish".

Jorge González, Antonio L. Lagarda, José R. Navarro, Laura Eliodoro, Adrià Giménez, Francisco Casacuberta, Joan M. de Val and Ferran Fabregat, "SisHiTra: A Spanish-to-Catalan hybrid machine translation system".

Dafydd Jones and Andreas Eisele, "Phrase-based statistical machine translation between Welsh and English".

Svetla Koeva, Svetlozara Lesseva and Maria Todorova, "Bulgarian sense tagged corpus".

Mirjam Sepesy Mauèec and Zdravko Kaèiè, "Statistical machine translation using the IJS-ELAN Corpus"

Sara Morrissey and Andy Way, "Lost in Translation: The Problems of Using Mainstream MT Evaluation Metrics for Sign Language Translation".

Alicia Pérez, Inés Torres, Francisco Casacuberta and Víctor Guijarrubia, "A Spanish-Basque weather forecast corpus for probabilistic speech translation".

Kevin Scannell, "Machine translation for closely related language pairs".

Oliver Streiter, Mathias Stuflesser and Qiu Lan Weng, "Models of Cooperation for the Development of NLP Resources: A Comparison of BLARK and XNLRDF".

John Wogan, Brian Ó Raghallaigh, Áine Ní Bhriain, Eric Zoerner, Harald Berthelsen, Ailbhe Ní Chasaide and Christer Gobl, "Developing a Spoken Corpus and a Synthesiser for Irish".

Pavel Zheltov, "An Attribute-Sample Database System for describing Chuvash affixes".

13:30 END OF WORKSHOP