

Report from SALT MIL

LREC, Marrakech, Morocco

1. Report on activities since 2006
 - a) Workshops
 - b) Other
2. Current committee
3. Looking ahead

Activities since 2006 (Workshops)

2006 Workshop: “Strategies for developing machine translation for minority languages”. Satellite workshop of LREC 2006, Genoa, Italy, May 2006.

7 talks, 16 posters

- "ENABLER, BLARK, what's next?" (Steven Krauwer)
- "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 MT: an opportunity for minor languages"
- "Approaching a new language in machine translation"
- "Unicode Development for Under-Resourced Languages".
- "Statistical MT with and without a bilingual training corpus"

Activities since 2006 (Workshops)

2007 Special Session at Interspeech 2007, “Speech and language technology for less-resourced languages”. *15 posters, 4 demos*

1. A Morpho-graphemic Approach for the Recognition of Spontaneous Speech in Agglutinative Languages – like Hungarian
2. A Semi-Supervised Learning Approach for Morpheme Segmentation for An Arabic Dialect
3. Accelerating the Annotation of Lexical Data for Less-Resourced Languages
4. On Web-based Speech Resource Creation for Less-Resourced Languages
5. Building an Information Retrieval System for Serbian - Challenges and Solutions
6. Bootstrapping Morphological Analysis of Gĩkũyũ Using Unsupervised Maximum Entropy Learning
7. The VoiceTRAN Machine Translation System
8. MuLAS: A Framework For Automatically Building Multi-Tier Corpora
9. Creating multimedia dictionaries of endangered languages using LEXUS
10. IceNLP: A Natural Language Processing Toolkit for Icelandic
11. Phonotactic spoken language identification with limited training data
12. Automatic Speech Recognition for an Under-Resourced Language
13. Information Retrieval Strategies for Accessing African Audio Corpora
14. Morfessor and VariKN machine learning tools for speech and language technology
15. Towards Better Language Modeling for Thai LVCSR

Activities since 2006 (Workshops)

2008 Workshop: “Collaboration: interoperability between people in the creation of language resources for less-resourced languages”.
Satellite workshop of LREC 2008, Marrakech, Morocco, May 2008.

1. Icelandic Language Technology Ten Years Later
2. Human Language Technology Resources for Less Commonly Taught Languages: Lessons Learned Toward Creation of Basic Language Resources
3. Building resources for African languages
4. Extracting bilingual word pairs from Wikipedia
5. Building a Basque/Spanish bilingual database for speaker verification
6. Language resources for Uralic minority languages
7. *Eslema*: Towards a Corpus for Asturian

Activities since 2006 (Other)

1. Website: <http://ixa2.si.ehu.es/saltmil/>
Completely re-designed (by student at EHU).
Better navigation design.
Now uses “Joomla” for easy content management.
2. Discussion list:
Active “pruning” of subscribers.
Very low traffic: needs more discussions!
3. ISCA grant:
1000 Euros for start-up of new SIG.
This sum recovered from accounts of ITE after much correspondence and delay.
Now held by ISCA for future use by SALT MIL SIG (for e.g. publicity costs).

Current committee and organisers

Chair: Briony Williams (Bangor University, UK)

Secretary: Mikel Forcada (Universitat d'Alacant, Spain)

Other workshop organiser: Kepa Sarasola (University of the Basque Country, Spain)

Review panel for 2008:

- Atelach Alemu Argaw (Stockholm University, Sweden)
- Julie Berndsen (University College Dublin, Ireland)
- Shannon Bischoff (Universidad de Puerto Rico, Puerto Rico)
- Lori Levin (Carnegie-Mellon University, USA)
- Climent Nadeu (Universitat Politècnica de Catalunya, Spain)
- Juan Antonio Pérez-Ortiz (Universitat d'Alacant, Spain)
- Bojan Petek (University of Ljubljana, Slovenia)
- Oliver Streiter (National University of Kaohsiung, Taiwan)

Looking ahead

- Activities for 2009???
- New committee members?
- More members of SALTMIL SIG needed.