

COCOSDA/WRITE/FLaReNet Workshop

**Language Technology issues for International
Cooperation**

Saturday May 22nd 2010

WORKSHOP ORGANISERS

**Nicoletta Calzolari
Khalid Choukri
Dafydd Gibbon
Joseph Mariani
Stelios Piperidis**

Workshop Program

9.30- 10.00: *The Workshop context and objectives*: Dafydd Gibbon and Nicoletta Calzolari

10.00-13.00: COCOSDA contribution: rapporteur overviews (10.30-11.00: Coffee break)

The COCOSDA rapporteurs will provide short overviews of current developments in their areas, concentrating on the COCOSDA remit of networking for international cooperation in speech and multimodal technologies resources and evaluation:

1.Regions (each talk 10 min)

1. *Asia* (Nick Campbell, for Oriental COCOSDA Trinity College, Dublin, Ireland)
2. *Americas* (Chris Cieri, Linguistic Data Consortium, USA)
3. *Europe* (Khalid Choukri, ELDA/ELRA, France)
4. *Africa* (Justus Roux, Potjestroom University, South Africa)
5. *Australia and Oceania* (tba)

2.Topics (each talk 15 min)

1. *Multimodal resources and assessment* (Nick Campbell, Trinity College, Dublin, Ireland)
2. *Multilingual multimodal resources - creating a levelling field in India* (Girish Nath Jha, Jawaharlal Nehru University, India)
3. *Resource infrastructures for endangered and other minority languages* (Peter Wittenburg, Max-Planck-Institut, Nijmegen, Netherlands)
4. *Speech technology developments for less resourced local languages* (Dafydd Gibbon, Universität Bielefeld, Germany)

13.00-14.30: Lunch break

14.30-18.00: FLaReNet – WRITE contribution (16.00-16.30: Coffee break)

14.30-15.00: **Reports from FLaReNet events**

- *Report from Poznan Workshop on Less-resources languages*: Joseph Mariani (LIMSI-CNRS & IMMI, France) and Khalid Choukri (ELDA, France)
- *Report from FLaReNet Forum in Barcelona*: Claudia Soria (ILC-CNR, Italy)

15.00-18.30: Infrastructural initiatives for sharing Language Resources (Data, Tools, Services)

They have different levels of generality, different purposes, but: How they compare? Which their respective roles? Synergies? Complementarities? Overlappings? How to profit from each other? Which lessons and priorities?

Each talk 15 minutes, followed by discussion among the presenters and with all the audience.

- *LanguageGrid*: Toru Ishida (Kyoto University, Japan) and Yohei Murakami (NICT, Japan)
- *Trends, priorities and strategies - from the CLARIN point of view*: Peter Wittenburg (MPI, The Netherlands) and Steven Krauwer (Utrecht University, The Netherlands)
- *Language Commons*: Ed Bice (Meedan, USA)
- *LDC & ELRA*: Chris Cieri (LDC, USA) and Khalid Choukri (ELDA, France)

- *U-Compare: an interoperable platform for sharing and evaluating text mining components:* Sophia Ananiadou (National Centre for Text Mining, UK)
- *The Panacea Platform:* Nuria Bel (UPF, Spain)
- *ALAGIN and language resources extracted from the Web:* Kentaro Torisawa (NICT, Japan)
- *LREC Map, what next?* Nicoletta Calzolari (ILC-CNR, Italy)
- *META-SHARE:* Stelios Piperidis (ILSP, Greece)

Language Technology issues for International cooperation

Two major international initiatives in the area of resource and evaluation coordination and standardisation, COCOSDA (together with WRITE) and FLaReNet, have agreed to submit a proposal for a joint workshop on Language Technology Issues for International Cooperation, as part of a series of event to discuss state of the art, international cooperation, roadmapping, etc. COCOSDA, the International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques, was established in 1991 to encourage and promote international interaction and cooperation in the foundation areas of Spoken Language Processing, especially for Speech Input/Output. COCOSDA has successfully facilitated and supported many international R&D initiatives in these and related areas, and is extending its mission to the coordination of multimodal resources and assessment, and to further cooperation with related initiatives (WRITE, FLaReNet). FLaReNet aims at providing recommendations for future initiatives in the field of Language Resources and Technologies: we are aware that it is important to discuss future policy and priorities not only on the European scene, but also in a worldwide context. This is true both when we try to highlight future directions of research, and – even more – when we analyse which infrastructures are needed. The growth of the field should be complemented by common effort that tries to look for synergies and to overcome fragmentation. It is an achievement, and an opportunity for our field, that recently a number of strategic-infrastructure initiatives have been debated and some are going to start, all over the world. This is also a sign that funding agencies recognise the strategic value of our field and the importance of helping a coherent growth also through a number of coordinated actions. Cooperation is an issue that needs to be prepared. FLaReNet may be the place where these – and future – initiatives get together to discuss and promote collaboration actions.

The aim of this joint COCOSDA/WRITE & FLaReNet strategic planning workshop is to start a discussion on the usefulness and the interest of promoting international cooperation among various initiatives and communities around the world, within and around the field of Language Resources and Technologies. This should lead to discussing the modalities of how to organise the cooperation.

On the basis of a number of contributions, some key topics will arise for discussions: for instance, COCOSDA contribution to the workshop will consist of regional rapporteur contributions on current activities in the field on a worldwide basis (Africa, Asia, Europe, North America, South America and Oceania) as well as topical rapporteur contributions on speech and multimodal technology resources and assessment. There will be a strong emphasis on resources and evaluation methods for sparse data based systems for underresourced languages.

FLaReNet will contribute with reports on its International forums and workshops. In particular FLaReNet will bring in both the issues debated at the Vienna (and the coming) Barcelona events and the recommendations that were drawn by the participants. Both events (and other workshops e.g. Brandeis, November 2009) addressed (some of the) technological, market and policy challenges posed by the Information society and the major role that HLT and LRS can play.

Both COCOSDA/WRITE and FLaReNet aims at boosting a constructive and fruitful debate. Examples of the questions to be addressed are:

1. Cluster 1:

- Which topics mostly deserve/require a worldwide effort for building consensus on priorities?
- Multilingualism as a unifying issue?

- How to boost the Interoperability issue?
- How to draw priorities per language? How to conduct a BLARK/ELRAK for each?

2. Cluster 2:

- Common Strategies for the next years? In which framework: local, regional, international frameworks?
- Relation between political and technical issues ? Inventory of the relevant initiatives in the various regions/countries? And directory of leading players.
- Cooperation through the initiatives or by people (Consensus/ cooperation through discussion or also through actions)?
- Which (realistic) modalities of cooperation among worldwide initiatives can be established, assuming different gaps/language? What is the best model for cooperation? Formal structures? Flexible models? Set up a permanent Forum?
- Or only bottom-up initiatives of cooperation are successful? See FLaReNet- SILT (EC-NSF) cooperation. How to be proactive? ?? Is it useful/possible to establish some form of “formal” cooperation between funding agencies as was done by the past ... ? Are there issues on which the respective agendas could be usefully coordinated?
- Space for international cooperation programs? Are there instruments to be used?
- Interaction between research communities and funding agencies