

Workshop on
**Adaptation of language resources and tools for closely related languages and
language variants**

13 September 2013

(<http://c-phil.informatik.uni-hamburg.de/view/Main/RANLPLangVar2013>)

followed by a session of COST Action IC 1002 **MUMIA**

8.50 - 9.00: Opening

9.00 - 10.00: **Invited talk: Preslav Nakov** (Qatar Computing Research Institute). *Combining, Adapting and Reusing Bi-texts between Related Languages: Application to Statistical Machine Translation.*

Session 1: Machine translation

10.00 - 10.30: **Ernesto López, Luis Chiruzzo and Dina Wonsever.** *Adaptation of a Rule-Based Translator to Río de la Plata Spanish.*

10.30 - 11.00: Coffee break

11.00 - 11.30: **Barry Haddow, Adolfo Hernandez, Friedrich Neubarth and Harald Trost.** *Corpus development for machine translation between standard and dialectal varieties.*

11.30 - 12.00: **Cristina Vertan and Walther v. Hahn.** *Language diversity and implications for Language technology in the Multilingual Europe.*

12.00 - 13.00: Lunch

Session 2: Language processing

13.00 - 13.30: **Pavlina Fragkou.** *Text segmentation for Language Identification in Greek Forums.*

13.30 - 14.00: **Yves Scherrer and Benoît Sagot.** *Lexicon induction and part-of-speech tagging of non-resourced languages without any bilingual resources.*

14.00 - 14.30: **Andjelka Zecevic and Stasa Vujicic-Stankovic.** *The Mysterious Letter J.*

14.30 – 15.00: Coffee break

MUMIA WG2 subsession: presentations

15.00-15.10: **Christian Boitet.** *Presentation of MUMIA-WG2 and recap on WG2 at Tallinn in June 2013¹.*

¹ (1) Software architectures in CLIR for patents: what we have and what we want, (2) OMNIA architecture for crosslingual search in multimodal collections, (3) adapting SECTra_w to support evaluation of IR techniques in partially multilingual patent corpora, (4) possibility to use the cloud to implement CLIR systems such as Watson Q/A (by IBM and CMU).

15.10-15.30: **Lingxiao Wang, Christian Boitet.** *Online production of HQ parallel corpora and permanent task-based evaluation of multiple MT systems: both can be obtained through iMAGs with no added cost.*

15.30-15.50: **Todor Tsonkov, Ivan Koychev.** *Detecting Irony in texts from the social networks: the Bulgarian language case*

15.50-16.10: **Galia Angelova, Ivelina Nikolova, Irina Temnikova.** *Enriching Patent Search with External Keywords: a Feasibility Study*

16.10-16.30: **Lingxiao Wang, Christian Boitet.** *Creation and incremental specialization of Moses-based MT systems for the CLEF 60Gb patent collection, to navigate in it through iMAGs in various languages.*

16.30 – 17.00: Coffee break

MUMIA WG2 subsession: demos and discussion

17.00-17.30: Demos **Lingxiao Wang, Christian Boitet**

Demo of iMAG MUMIA-CLEF on a subset of the 60 Gb CLEF patent collection:

- using Google for the non-translated segments, in the main "CLEF" languages (a lot in English, French, German, a few in Spanish, some only in the other languages)
- using Moses-based systems created mainly from the bisegments extracted from a subset of the 60 Gb CLEF patent collection uploaded in the AXiMAG corpus database.

17.30-18.00: **Discussion** (at least MUMIA members, others are welcome to contribute!).