

Presentation

Sylviane CARDEY

Coordinator, IM in NLP & HLT,
Centre Tesnière, Université de Franche-Comté, France

This special number of the BULAG contains research project papers of special merit drawn from those written by students of the second intake (2009) of the European Erasmus Mundus Masters Course, IM in NLP & HLT - International Masters in Natural Language Processing and Human Language Technology (EMMC 2008-0083).

The papers have been written as a requirement of the International Masters students' research projects which they presented at the 2nd International Workshop on Natural Language Processing and Human Language Technology, Universidade do Algarve, Faro, 8-9 June, 2011.

The International Masters, which is a research masters, is administered by a consortium of four European universities: Université de Franche-Comté (coordinator), France, Universidade do Algarve, Portugal, Universitat Autònoma de Barcelona, Spain, University of Wolverhampton, United Kingdom. The students study in 2 or 3 countries - one learns a language better if one practises it in the country where it is spoken. Furthermore, by means of its language policy which extends to the course of study, the course provides competence in more than one language, an essential for the Computational Linguist.

Being a research masters, the course content has input from the consortium's research findings. The theories developed and used by the different research centres in the consortium are varied – thus there is no attachment to some specific 'school' for this Masters and this avoids a closed view and ensures that the students are open minded. The papers collected here indeed reflect this 'open' policy. Each of the students, in conducting research as an integral part of the course programme, has been co-supervised by two teachers attached to research centres associated with two of the consortium members where they have studied.

The papers presented here involve not only various research themes but also associated linguistic analyses, including comparative, of several languages for various ends; English, Finnish, French, Persian, Portuguese, Russian, and Spanish are represented.

These analyses have in common natural language processing and human language technology for very diverse applications.

This diversity of applications is exemplified by the methodology proposed by Sanja Štajner for investigating language change, and in the evaluation of linguistic adequacy where Biljana Drndarević conducts a study towards software for Basic Literary English.

In the domain of multilinguism and in particular translation and machine translation we find papers which are varied and deal with different problems and their proposed solutions. José Guilherme Camargo de Souza addresses coreference resolution, Sheila Castilho Monteiro de Sousa the translation of DVD sub-titles and Sean Healy that of controlled languages. Aurora Marsye contributes to translation memory systems, and Mitra Shahabi studies verb phrase ellipsis. Gilda Tataei contributes to parallel corpus building, and Maria Vorobey conducts a contrastive study of pragmatemes.

Concerning disambiguation, Gaia Olivia Fernandes addresses the *-mente* ending adverbs in Brazilian Portuguese, Tatiana Lesnikova is concerned with classifying homonymous city names and Yoon Mi Oh studies light verbs and phrasal verbs in a Question/Answer system.

Finally, in respect of normalization, Paul Brilliant Feuto Njonko addresses the problem of text normalization and Alice Kaiser-Schatzlein's contribution concerns context-sensitive word errors in speech recognition transcripts.