Ph.D. Program in Linguistics - University of Pavia

Title of the Course: QUANTITATIVE METHODS AND CORPORA IN LINGUISTIC RESEARCH

Organizer: Prof. Dr. Silvia Luraghi, Dipartimento di Linguistica Teorica e Applicata, Università di Pavia, email: silvia.luraghi@unipv.it

Objectives and Features of the Course:

The course on Quantitative Methods and Corpora in Linguistic Research aims to equip students with skills that will enable them to choose and/or create the appropriate corpus for a given study, the appropriate tools to extract and analyze relevant data, and the appropriate notions for interpreting statistical data, based on theories underlying the distributional paradigm. It will have a strong hands-on flavor, and will feature the following units: (a) Statistical methods (R, ANOVA, multidimentional scaling) and their application to historical linguistics, linguistic typology and language acquisition research; (b) Different types of corpora and their different uses: corpora and annotations will be both discussed from a theoretical point of view, and dealt with practically, via the use of customizable tools for corpus annotation and processing; (c) Acquiring and representing distributional information: students will be presented with methods, techniques and measures used for the acquisition of semantic information from corpora; focus will then move to the problems that arise translating the acquired semantic information into linguistic representation; (d) Data extraction from corpora: course participants will learn to use software for data extraction and processing in user-friendly environments.

Invited International Instructors:

Michael Cysouw, Ludwig-Maximilans-Universität-München, Research Unit 'Quantitative Language Comparison', cysouw@lmu.de, http://www.qlc.sprachwiss.uni-muenchen.de/people/project_leader/michael_cys/index.html
Jan Hajic, Univerzita Karlova v Praze, Institute of Formal and Applied Linguistics, Jan.Hajic@mff.cuni.cz, http://ufal.mff.cuni.cz/~hajic/
Olga Lashevskaja, Universitetet i Tromsø, Dept. of language and linguistics, olesar@mail.ru, http://olesar.narod.ru/en/about.html
Magnus Sahlgren, Swedish Institute of Computer Science, Stockholm, mange@gavagai.se, http://www.sics.se/~mange/research.html

Other Instructors in the Course:

Maria Freddi, Professore Associato, Dipartimento di Linguistica Teorica e Applicata, Università di Pavia, email: maria.freddi@unipv.it
Elisabetta Jezek, Ricercatore confermato, Dipartimento di Linguistica Teorica e Applicata, Università di Pavia, email: jezek@unipv.it
Malvina Nissim, Ricercatore confermato, Dipartimento di Studi Linguistici e Orientali, Università di Bologna, e-mail: malvina.nissim@unibo.it
Stefano Rastelli, Ricercatore a tempo determinato Università di Pavia, e-mail: stefano.rastelli@unipv.it

Duration of the Course: four units of 16 hours, total: 64 hours of activity. Given the hands-on nature of the course, there will be no lectures in the strict sense, but only workshop activities.

ECTS credits: 8 ECTS credits

Period: September 10-21, 2012

Location of the Course: Aula L3 San Tommaso, P.zza del Lino, 2, 27100 Pavia

Additional information:
The course will be attended by all students enrolled in the PhD program in Linguistics (about 15 distributed over the three-year program) and by four to six international PhD students from the research groups of the Invited International Speakers. The course will also admit additional students from other PhD programs in Lombardy.