



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



IQLA-GIAT Summer School in
Quantitative Analysis of Textual Data
University of Padua, 16-20 September 2013

Final Report

The first edition of the IQLA-GIAT Summer School in *Quantitative analysis of textual data* took place at the University of Padua, Department of *Philosophy, Sociology, Education and Applied Psychology* (FISPPA) – Sociology buildings, from Monday 16th to Friday 20th September 2013.

The Summer School was funded by the University of Padua, coordinated by prof. Arjuna Tuzzi (University of Padua) and organized by GIAT – *Interdisciplinary Text Analysis Group* (www.giat.org) in collaboration with IQLA – *International Quantitative Linguistics Association* (www.iqla.org) and the Department of *Computational Linguistics* of the University of Trier (www.uni-trier.de).

The teaching staff included scholars, researchers and IT experts from 10 European universities: Toulouse II, ENS de Lyon (France), Trier (Germany), Athens (Greece), Neuchatel (Switzerland), Kraców (Poland), Bologna, Padua, Trieste, Rome “La Sapienza” (Italy).

In terms of the number and quality of participants, the Summer School achieved an exceptional success: 35 applications for 20 available positions. The selected participants mainly belonged to three scientific branches - linguistics, mathematics and computer sciences - and came from 11 different countries: Albania, Bulgaria, China, Croatia, Germany, Greece, Italy, Netherlands, Romania, Russia and Spain (but also from Finland, France, UK, and USA if we consider the institutions where they are working). The participants' CVs were excellent and well suited for the Summer School objectives: In terms of their specific research interests, experiences and publications, the participants formed a strongly motivated and adequately skilled group, and this proved to be a great advantage for both the participants and the teaching staff.

The lecturers discussed with participants the strengths, weaknesses, opportunities and threats of various quantitative methods for text analysis. In their turn, the participants had the opportunity to develop the necessary expertise to exploit integrated approaches to text analysis for a better understanding of the role played by theoretical issues, expected results and data and software selection. The lecturers shared information on software, corpora, relevant literature and research results about:

1. corpora and textual data collection;
2. quantitative linguistics;
3. authorship attribution;
4. content mapping and data visualization;
5. software for computer-aided analysis of textual data.

Teaching activities raised questions that can be answered thanks to quantitative methods implemented within a text analysis framework and other procedures aimed to identify and compare the characteristics of texts. Teaching activities included lectures and lab sessions, as well as workshops illustrating software and tools. The lab sessions took place in a computer lab of the Department FISPPA. Each participant was assigned a PC in which all the necessary software had already been installed (the limit of 20 participants in the Summer School was determined by the number of PCs in the lab).

At the end of the classes all participants filled in a questionnaire to assess the main aspects of the Summer School (e.g. organization, teaching, materials, facilities and equipment, expectations, satisfaction rate, suggestions etc.) and five of them decided to take a written test in order to receive ECTS credits (3 CFU). The organizers were very pleased with the assessment because the participants' opinions were very good and those who took the test passed it with excellent results.

Of course the Summer School program also included social events, i.e. two lunch breaks and an evening cocktail in the beautiful garden of the Department building, a pizza-party in a restaurant in the city centre and a guided tour of Palazzo del Bo (the ancient main building of the University of Padua).

Many thanks...

... to the University of Padua and its International Relations Office for funding this first edition of the IQLA-GIAT Summer School;

... to the Department FISPPA for the great support provided by technical and administrative staff;

... to the members of the GIAT group, the IQLA association and Reinhard Köhler's research team from Trier for their support and friendship;

... to all colleagues, scholars, and lecturers for joining the Summer School and sharing with participants their expertise, research experiences and knowledge;

... to all candidates for their application and all selected participants for joining the Summer School and showing great enthusiasm and willingness to learn.

And, of course, see you at the next edition!



Program of the first edition (2013)

General Topics

1. Building corpora: caveats, opportunities and examples (Stefano Ondelli, University of Trieste, Italy)
2. Collecting corpora on the Web: the dubious joys of scraping web pages (Alberto Cammozzo, IT expert, Italy)
3. Qualitative and quantitative approaches (Alberta Contarello, University of Padua, Italy)
4. Methodology: measuring and testing (Reinhard Köhler, University of Trier, Germany)

Quantitative Linguistics

1. Measures of lexical richness (Andrei Beliankou, University of Trier, Germany)
2. Linguistics laws (Reinhard Köhler, University of Trier, Germany)
3. Quantitative syntax analysis (Sven Naumann, University of Trier, Germany)

Authorship Attribution

1. Distance-based models in authorship attribution (Jacques Savoy, University of Neuchatel, Switzerland)
2. Machine learning methods in authorship attribution and profiling (George Mikros, University of Athens, Greece)
3. Mathematical models of textual data with applications to authorship attribution (Mirko Degli Esposti, Alma Mater University of Bologna, Italy)

Software (Workshop and Lab-Sessions)

1. LAB – Introduction to Iramuteq (Pierre Ratinaud, University of Toulouse II, France)
2. LAB – Introduction to TXM (Serge Heiden, ENS de Lyon, France)
3. LAB – Introduction to TaLTac2 (Isabella Chiari, University of Rome “La Sapienza”, Italy)
4. LAB – Introduction to R package ‘stylo’ (Maciej Eder, Kraków, Poland)
5. LAB – Introduction to WordSmith tools and AntConc (Maria Teresa Musacchio, University of Padova, Italy)
6. LAB – Problems and practical solutions in quantitative linguistics (Andrei Beliankou and Sven Naumann, University of Trier, Germany)
7. LAB – Lab session on profiling (George Mikros, University of Athens, Greece)