IQLA-GIAT Summer School in

Quantitative Analysis of Textual Data

University of Padua, 9-13 September 2019

4th edition

Final Report

The fourth 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 9th to Friday 13th September 2019. The Summer School was funded by the University of Padua, coordinated by prof. **Arjuna Tuzzi** (University of Padua, Dept. FISPPA) and organized by GIAT – *Interdisciplinary Text Analysis Group* (www.giat.org) in collaboration with IQLA – *International Quantitative Linguistics Association* (www.iqla.org).

The teaching staff included ten lecturers from four European countries, nine different universities and one research center: University of Roma "Tor Vergata", University of Venezia "Ca' Foscari", University of Trieste and University of Trento (Italy), University of Toulouse II Jean Jaurès and University Paul Valéry of Montpellier (France), Polish Academy of Sciences and Jagiellonian University of Kracόw (Poland), National and Kapodistrian University of Athens (Greece) and Italian Natural Language Processing Lab, CNR (Italy). In terms of the number and quality of participants, the Summer School achieved good success. The 19 selected participants (out of 36 applications) belonged to five scientific branches - linguistics, computer sciences, economics, political sciences, and psychology - and came from 9 different countries: Finland, Denmark, France, Greece, Italy, Egypt, Republic of the Philippines, U.S.A. and Turkey. The participants' CVs were excellent and well suited for the Summer School objectives: in terms of their specific research interests, experiences and interdisciplinary viewpoints. 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.

Teaching activities

"Distant Reading", "Digital Methods", "Computational social sciences" and "Statistical learning from textual data" are general terms that refer to a wide range of methods that have a common aim: retrieving information from texts by means of computer-aided tools. Today, computer-aided text analysis is an umbrella term referring to a number of qualitative, quantitative and mixed-methods

approaches. It is an object of research in many sectors of linguistics, computer sciences, mathematics and statistics. Furthermore, computer-aided text analysis is used as a research tool within several disciplines such as psychology, philosophy, sociology, sociolinguistics, education, history, political studies, literary studies, communication, and media studies. The recent evolution of information technologies (IT) and computational methods has led to a number of distinct but interrelated sectors (e.g. computational linguistics, information retrieval, natural language processing, text mining, text analytics, stylometry, sentiment analysis, opinion mining, topic extraction, etc.) with interesting industrial applications, such as electronic dictionaries, artificial intelligence, plagiarism detection and similar.

Recent studies have stressed the need for developing, adopting and sharing interdisciplinary approaches. The IQLA-GIAT Summer School is the ideal environment for developing innovative analytical tools by pooling together the research methods from different disciplines.

The IQLA-GIAT Summer School has been characterized by three main elements:

1. a general part devoted to quantitative methods;
2. a special issue that has changed over time (2019: Data Science and Data scientists in Humanities and Social Sciences);
3. several lab-sessions dedicated to computer-aided analyses 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 packages were available (the limit of 20 participants in the Summer School was determined by the number of PCs in the lab). One afternoon was dedicated to a workshop, that this year concerned Quantitative Methods in Forensic Linguistics.

At the end of the classes, the 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 a self-assessment test in order to check what they got from the school. The organizers were very pleased with the assessment because the participants' opinions were very positive towards the classes and the school in general. Moreover, they provided useful insights to improve the performance of the next editions.

Enjoyable details

The Summer School program also included lunch for each day of the school, social dinner, coffee breaks and a visit to Palazzo Bo.

All participants and teachers received a conference kit including gadgets (bag, folder, maps, pen drive, stationery), brochures and flyers about the hosting town. Besides this, the Summer School webpages provided indications for accommodation and travel plan.

What's next?

Keep an eye out for the next edition…

Many thanks to…

- University of Padua for funding this third edition of the IQLA-GIAT Summer School;
- International Relations Office and Department FISPPA for the great support provided by technical and administrative staff;
- members of the GIAT group and the IQLA association;
- colleagues, and scholars, for joining the Summer School and sharing with participants their expertise, research experiences and knowledge;
- candidates for their application and selected participants for joining the Summer School and showing great enthusiasm and willingness to learn.

<div align="center">

Programme

4th *edition 2019*

</div>

LECTURES (15h)

Fabio Ciotti (Uniersity of Roma "Tor Vergata", Italy) – *Quantitative Literary studies. Chimera or Phoenix?*

Sascha Diwersy (University Paul Valéry of Montpellier, France) – *PhraseoRom: on the corpus-driven analysis of literary genre*

Stefano Ondelli (University of Trieste, Italy) – *Before the Analysis: examples of corpus compilation and preprocessing*

Dominique Brunato (Italian Natural Language Processing Lab, CNR, Italy) – *NLP-based stylometry for author profiling with a focus on gender*

Giuseppe Alessandro Veltri (University of Trento, Italy) – *Introduction to text mining for cognitive sociology research*

George Markopoulos (National and Kapodistrian University of Athens, Greece) – *Sentiment analysis and emotion detection in texts*

Maciej Eder (Pedagogical University of Kraków and Polish Academy of Sciences, Kraków, Poland) – *The role of Computational Stylistics*

Silvia Bozza (University Ca' Foscari of Venezia, Italy and Université de Lausanne, Suisse) – *The role of Statistics*

LAB SESSIONS (18h)

Sascha Diwersy (University Paul Valéry of Montpellier, France) – *Aspects of Lexical time: diachronic data, corpus linguistics and textometry*

Giuseppe Alessandro Veltri (University of Trento, Italy) – *Introduction to text mining for cognitive sociology research*

Pierre Ratinaud (University of Toulouse II, France) – *IraMuTeQ – corpus indexation, manipulation and simple description; The Reinert method in IRaMuTeQ*

George Markopoulos (National and Kapodistrian University of Athens, Greece) – *Sentiment analysis and emotion detection in texts*

Maciej Eder (Pedagogical University of Kraków and Polish Academy of Sciences, Kraków, Poland) – *Stylometry with the package 'Stylo': explanatory methods and supervised methods*

Jan Rybicki (Jagiellonian University, Kraków, Poland) – *Introduction to stylometry with Stylo*

International workshop: Quantitative Methods in Forensic Linguistics, with Georges Markopoulos, Maciej Eder, and Silvia Bozza

Organizing Committee

Francesco Gatti, Valentina Rizzoli, Ilenia Sanna, Irene Saonara, and Stefano Sbalchiero (University of Padua and Catholic University of Milan, Italy)

Quality Assessment of the School

Final Evaluation and self-assessment

Tutorship