IQLA-GIAT Summer School in
Quantitative Analysis of Textual Data
University of Padua, 4-8 September 2017
3rd edition

Final Report

The third 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 4th to Friday 8th September 2017. 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) and the Department of *Computational Linguistics* of the University of Trier (www.uni-trier.de).

The teaching staff included nine lecturers from seven European universities of six different countries and a lecturer from the United States: University of Toulouse II (France), University of Trier (Germany), National and Kapodistrian University of Athens (Greece), University of Neuchâtel (Switzerland), Polish Academy of Science and Jagiellonian University of Kraców (Poland), University of Trieste (Italy), and Duquesne University (Pittsburg, Pennsylvania). Moreover, three teaching assistants have been involved in tutorials and lab sessions. In terms of the number and quality of participants, the Summer School achieved a good success. The 19 selected participants (out of 32 applications) belonged to seven scientific branches - linguistics, computer sciences, mathematics, economics, political sciences, and psychology - and came from 10 different countries: China, Croatia, Germany, Greece, Italy, Lebanon, Poland, Russia, Switzerland, and United Kingdom. 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

Digital methods have been utilized by a variety of disciplines and the growing availability of large corpora and large databases (the BIG DATA era) calls for new methods to deal with new problems, open the door to new questions and develop new knowledge.

"Quantitative analysis of textual data", "Digital Methods" and "Distant Reading" are general terms that refer to a wide range of methods sharing a common aim: retrieving and summarizing 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 and it is used as a research tool within a number of 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 quantitative methods has led to a number of distinct but interrelated sectors (e.g. computational linguistics, information retrieval, natural language processing, text mining, text analytics, sentiment analysis, opinion mining, topic extraction, etc.) with interesting industrial applications (e.g. electronic dictionaries, artificial intelligence, computer-aided translation, plagiarism detection, web reputation).

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

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

1. a general part devoted to quantitative linguistics;
2. a special issue addressing a relevant methodological problem (2017: topic detection and authorship attribution in Elena Ferrante's case-study; 2015: measuring style and computational stylistics; 2013 measures and methods in authorship attribution);
3. several lab-sessions dedicated to 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 packages were available (the limit of 20 participants in the Summer School was determined by the number of PCs in the lab).

For its third edition the IQLA-GIAT Summer School has included a Workshop on Elena Ferrante's case-study. It was an open event also available in live streaming. The speakers were Jacques Savoy (University of Neuchâtel), Jan Rybicki (Jagiellonian University of Krakow), Maciej Eder (Polish Academy of Sciences), Patrick Juola (Duquesne University, Pittsburg), George Mikros (National and Kapodistrian University of Athens), Pierre Ratinaud (University of Toulouse II), and Arjuna Tuzzi (University of Padua). Rocco Coronato and Luca Zuliani (University of Padua) played the role of chairs. The international group of experts applied different methods to a corpus of 150 novels of 40 contemporary Italian authors, gathered by Arjuna Tuzzi and Michele Cortelazzo (University of Padua). The workshop achieved resounding success, and arouse the attention of national and international media.

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 next editions.

Enjoyable details

The Summer School program also included welcome and goodbye lunch, social dinner, coffee breaks and social events (a visit to the Botanical garden, and a night visit to the Baptistery of the Padua Cathedral).

All participants and teachers received a conference kit including gadgets (bag, folder, maps, pen drive, stationery), brochures and flyers about the hosting town. Beside 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, IQLA association, and Department of Computational linguistics of the University of Trier for their support (and friendship);
- 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.

Programme
3rd edition 2017

LECTURES (14h)

Stefano Ondelli (University of Trieste, Italy)
> *The Impact of Sociolinguistic and Morphological Factors on Corpus Design*

Reinhard Köhler (Uniersity of Trier, Germany)
> *General Methodology in Empirical Linguistics. Evaluation of Data and Hypothesis Testing*
> *A Crash Course in the Central Terms and Concepts of Science*

Sven Naumann (University of Trier, Germany)
> *Syntactic Complexity*
> *Topic Modeling*

George K. Mikros (National and Kapodistrian University of Athens, Greece & University of Massachusetts, Boston, Massachusetts)
> *Author profiling: Detecting author's gender in social media*
> *Computational stylistic analysis of translated texts. Detecting the stylometric identity of both the author and the translator*

Maciej Eder (Polish Academy of Sciences, Kraków, Poland)
> *Introduction to distributional semantics: topic modelling and word vector representations*

Jacques Savoy (University of Neuchâtel, Switzerland)
> *Other Applications of Authorship Attribution Methods*

LAB SESSIONS (7h)

Maciej Eder (Polish Academy of Sciences, Kraków, Poland)
> *Stylometry with the package 'Stylo': Explanatory methods and Supervised methods*

Jan Rybicki (Jagiellonian University, Kraków, Poland)
> *Stylometry with the package 'Stylo': Explanatory methods and Supervised methods*

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

Patrick Juola (Duquesne University, Pittsburgh, Pennsylvania)
> *Feature Sets in Authorship Attribution: A Software-Based Case Study*

OTHER ACTIVITIES (6h)

Workshop: Drawing Elena Ferrante's profile

Valentina Rizzoli, Stefano Sbalchiero, and Irene Saonara (University of Padua, and Catholic University of Milan, Italy)
> *Lab tutorial*
> *Basics in QL*
> *Quality Assessment of the School*
> *Final Evaluation and self-assessment*
> *Free Lab time*
> *Lab sessions*
> *Tutorship*