



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



IQLA-GIAT Summer School in

Quantitative Analysis of Textual Data

University of Padua, 26-30 July 2021

5th edition (online)

Final Report

The fifth edition of the IQLA-GIAT Summer School in *Quantitative analysis of textual data* took place online via the Zoom platform, hosted by the University of Padua, Department of *Philosophy, Sociology, Education and Applied Psychology* (FISPPA), from Monday 26th to Friday 30th July 2021. 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 eleven lecturers from eight countries (Austria, France, Italy, Poland, Qatar, Spain, Switzerland, US), ten different universities: Universität Wien (Austria), University of Toulouse II Jean Jaurès (France), University of Roma “Tor Vergata” and University of Trieste (Italy), Polish Academy of Sciences and Jagiellonian University of Kraków (Poland), Hamad Bin Khalifa University of Doha (Qatar), Universitat politècnica de Catalunya (Spain), University of Neuchâtel (Switzerland), and Duquesne University (Pittsburg, Pennsylvania, US). In terms of the number and quality of participants, the Summer School achieved good success. The 28 selected participants belonged to six scientific branches - linguistics, computer sciences, economics, political sciences, sociology, and psychology - and came from 9 different countries: Austria, Egypt, France, Greece, Hong Kong, Italy, Lithuania, Switzerland, US. The participants' CVs were excellent and well-suited for the Summer School objectives: in terms of their specific research interests, experiences and interdisciplinary viewpoints.

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 using 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 many 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 several 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, such as media and new media monitoring, electronic dictionaries and translators, plagiarism detection and similar.

Recent developments in digital methods are not only changing how research is conducted in the humanities and social sciences, but also how new research is planned and designed. The interest in news collections arose with the investigation of propaganda in press coverage during the first half of the 20th century. Today, large corpora of texts and text mining tools are exploited to extract hot topics, sentiment, opinions and trends.

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

1. a general part devoted to quantitative methods;
2. a special issue addressed to a relevant methodological problem that has changed over time

2021: Quality of Texts quality of News

2019: Data Science and Data scientists in Humanities and Social Sciences;

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. lab-tutorials dedicated to the computer-aided analysis of textual data.

Teaching activities at the School raised questions that can be answered thanks to quantitative methods implemented within a text analysis framework and other procedures that may be used to identify and compare text characteristics. Both lab and talking sessions took place via the Zoom platform. Each participant was sent via email materials and instructions necessary to follow the hands-on sessions.

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, expectations, satisfaction rate, suggestions etc.) and a self-assessment test 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

Despite we had to organize an online edition, the Summer School program also included enjoyable moments, as virtual tours of Padua or typical recipe sharing. All participants and teachers received a conference kit via mail including gadgets.

What's next?

Keep an eye out for the next edition...

Many thanks to...

- University of Padua for funding this fifth 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.

Programme

5th edition 2021

LECTURES (22h)

Fabio Ciotti (University of Roma “Tor Vergata”, Italy) – *The problem of quality in distant reading_ quality of data, quality in methodology*

Maciej Eder (Pedagogical University of Kraków and Polish Academy of Sciences, Kraków, Poland) – *Introduction to distributional semantics*

Ramon Ferrer – i – Cancho (Universitat politècnica de Catalunya, Spain) – *An introduction to quantitative dependency syntax with the Linear Arrangement Library (with Lluís Alemany-Puig)*

Patrick Juola (Duquesne University, Pittsburg, Pennsylvania, US) – *Theory and application of text analysis and classification*

Emmerich Kelih (Universität Wien, Austria) – *Quantitative linguistics – some particularities in modelling language data*

George Mikros (Hamad Bin Khalifa University of Doha) – *Mis/Dis Information during the pandemic. Computational stylistics to the rescue*

Stefano Ondelli (University of Trieste, Italy) – *Purpose and Design: Choosing the right corpus for our research purpose; Purpose and Design: examples of the compilation of and research on different corpora*

Jan Rybicki (Jagiellonian University, Kraków, Poland) – *Introduction to Stylometry*

Jacques Savoy (University of Neuchâtel) – *Stylistic Analysis of the US Presidential Speeches*

Floriana Sciumbata (University of Trieste, Italy) – *From EtR to FLO: creating a computer-assisted writing system using the new Italian guidelines for Easy-to-Read and text analysis tools*

LAB SESSIONS (12h)

Maciej Eder (Pedagogical University of Kraków and Polish Academy of Sciences, Kraków, Poland) – *Stylometry with the package Stylo (Hands-on session)*

Ramon Ferrer – i – Cancho (Universitat politècnica de Catalunya, Spain) – *An introduction to quantitative dependency syntax with the Linear Arrangement Library (with Lluís Alemany-Puig)*

Patrick Juola (Duquesne University, Pittsburg, Pennsylvania, US) – *Textual analysis using the JGAAP system*

George Mikros (Hamad Bin Khalifa University of Doha, Qatar) – *Using NodeXL to analyze Tweets: a practical approach*

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

Jan Rybicki (Jagiellonian University, Kraków, Poland) – *First step with Stylo (Hands-on session)*

Organizing Committee

Franco Gatti, Valentina Rizzoli, Ilenia Sanna, Irene Saonara, and Stefano Sbalchiero
(University of Bologna, Padua and Sapienza University of Rome, Italy)

Quality Assessment of the School

Final Evaluation and self-assessment

Tutorship