



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



IQLA-GIAT Summer School in

## ***Quantitative Analysis of Textual Data***

University of Padua, 4-8 September 2023

*6th edition*

### **Final Report**

The sixth edition of the IQLA-GIAT Summer School in *Quantitative analysis of textual data* took place at the University of Padua, Complesso Polo Beato Pellegrino, from Monday 4th to Friday 8th September 2023. 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.fisppa.unipd.it/giat-en/home](http://www.fisppa.unipd.it/giat-en/home)) in collaboration with IQLA – *International Quantitative Linguistics Association* ([www.iqla.org](http://www.iqla.org)).

The teaching staff included eleven lecturers from five countries (Austria, Italy, Poland, Qatar, US), eight different affiliations: University of Vienna (Austria), CNR (Pisa), University of Padua, and University of Trieste (Italy), Polish Academy of Sciences of Kraków and University of Wrocław (Poland), Hamad Bin Khalifa University (Doha, Qatar), Duquesne University (Pittsburg, Pennsylvania, US). In terms of the number and quality of participants, the Summer School achieved good success. The 19 selected participants belonged to five scientific branches - linguistics, computer sciences, nursing sciences, pedagogical sciences, and psychology - and came from 8 different countries: China, Czech Republic, Germany, Greece, Italy, Philippines, Spain, US. The participants' CVs were excellent and wellsuited for the Summer School objectives: in terms of their specific research interests, experiences and interdisciplinary viewpoints.

### **Teaching activities**

Distant reading, text data science, digital humanities, and computational text analysis are umbrella terms that refer to a wide range of methods that have a common aim: retrieving information from texts using quantitative methods and computer-aided tools. Today, text analysis refers to a number of qualitative, quantitative and mixed-methods approaches. It is an object of research in many sectors of linguistics, computer sciences, social sciences, mathematics and statistics. Furthermore, it represents a research tool within many disciplines such as psychology, philosophy, sociology, sociolinguistics, education, history, political studies, literary studies, communication and media studies.

Digital sources that provide archives of books, documents, journals or disseminate information, such as newspapers and social media, now constitute an indispensable basis for many research studies. The increasing availability of large text corpora requires new tools to address new research questions.

Recent developments in computational methods have rapidly changed not only the way research is conducted in the humanities and social sciences but also the way research is conceived and designed. The demand for data scientists is growing in all disciplines, both in academia and in many productive and commercial sectors.

A new generation of researchers must be trained to correctly apply data science methods in many application domains of humanities and social sciences.

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

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

**2023: Language Variation and Change from a Computational Perspective**

2021: Quality of Texts quality of News (online edition)

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 will raise 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 in room 7, Complesso Polo Beato Pellegrino (Via E. Vendramini, 13). Students were asked to bring their own laptop. One afternoon was dedicated to a R Lab.

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**

The Summer School program also included lunch for each day of the school, social dinner, and coffee breaks. Overnight stay was included in the participant fees. All participants and teachers received a conference kit including gadgets (bag, folder, maps, pen drive, bookmark, block-notes), brochures and flyers about the hosting town. Besides this, the Summer School staff was in contact through social platforms with participants and lecturers to provide them with all the necessary information and details for their stay.



### **What's next?**

*Keep an eye out for the next edition...*

### **Many thanks to...**

- University of Padua for funding this sixth 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**

6th edition 2023

### **LECTURES (18 h)**

**Dominique Brunato** (CNR, Pisa) - *Characterizing language variation through a computational linguistics lens*

**Joanna Byszuk** (Polish Academy of Sciences of Kraków) - *Examining language variation with stylometry*

**Silvia Crafa** (University of Padua) - *The non-neutrality of data science*

**Emmerich Kelih** (University of Vienna) - *Quantitative linguistics: current state of the art & Modelling language and text data: the quantitative linguistics perspective*

**Patrick Juola** (Duquesne University, Pittsburgh) - *Stylometry and the problem of Coauthorship*

**George K. Mikros** (Hamad Bin Khalifa University, Doha) - *Quantitative Profiling of AI writing*

**Stefano Ondelli** (University of Trieste) - *Representativeness and Balance: Corpora as Samples of Language*

**Adam Pawloski** (University of Wrocław) - *Bibliodata in quantitative research – data, methods, case-studies*

### **LAB SESSIONS (14 h)**

**Joanna Byszuk** (Polish Academy of Sciences of Kraków) - *Examining language variation with stylometry*

**Patrick Juola** (Duquesne University, Pittsburgh) - *Stylometry and the Problem of Coauthorship*

**George K. Mikros** (Hamad Bin Khalifa University, Doha) - *Text Classification using the simple transformers Python library*

**Andrea Sciandra** (University of Padua) - *Exploring textual data with R (part 1 & 2)*

**Matilde Trevisani** (University of Trieste) - *Language variation and change in diachronic corpora: sparsity, asymmetry and peak-and-valley course of word frequencies over time & Life-cycle of words: temporal pattern recognition and curve clustering (with a lab in R)*

### **ORGANIZING COMMITTEE**

**Franco Gatti, Valentina Rizzoli, Ilenia Sanna, Andrea Sciandra, and Stefano Sbalchiero** (CNR- Florence, University of Padua and Sapienza University of Rome, Italy)

*Working Group Lab*

*Quality Assessment of the School*

*Final Evaluation and self-assessment*

*Tutorship*