SHORT TERM SCIENTIFIC MISSION (STSM) SCIENTIFIC REPORT

This report is submitted for approval by the STSM applicant to the STSM coordinator

Action number: CA16204
STSM title: Distant Reading periodicals of the Enlightenment
STSM start and end date: 24/02/2019 to 15/03/2019
Grantee name: Martina Scholger

PURPOSE OF THE STSM:

(max. 200 words)

The enlightened Spectator press of the 18th century constitutes an important cultural heritage of the world, affecting the discourse system of the Enlightenment. It complied with the democratic ideal of disseminating cultural and moral issues, techniques and practices within a non-academic audience, popularizing enlightened ideas such as cosmopolitanism, tolerance, intellectual criticism, self-reflection and social responsibility.

The purpose of the STSM has been to analyse the digitized, TEI encoded and semantically enriched texts of The Spectators in the International Context (https://gams.uni-graz.at/mws), an ongoing digital edition project which has been running since 2008, with quantitative methods. The STSM ties directly into the scope of the COST Action CA16204 and a project which was approved in December 2018 by the Austrian Academy of Sciences, as a feasibility study and preparatory activity.

Of particular interest was the application of stylometry to investigate journals and texts by anonymous authors and the investigation of collaborative works. Driven by this particular research interest, the host was chosen to benefit and learn from the host's many years of experience in the development and application of stylometric methods and tools, especially for the investigation of previously unknown authorships and suspected collaborative works.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSM

(max. 500 words)

Corpus preparation

The approximately 3700 Spectator texts from 66 different journals were downloaded as a data-dump from the website https://gams.uni-graz.at/mws. One of the major challenges has been the compilation of the text corpus according to different aspects: First of all, a sorting according to the currently six languages (French, Italian, Spanish, German, English, and Portuguese) in the corpus has been undertaken. Secondly, the individual texts have been sorted by journal. Since the individual texts are very short, it has been necessary to join together multiple texts to get a representative and computable test corpus for the analysis with "Stylometry with R" (Stylo). Furthermore, the filenames have been renamed with a Python script according to the metadata in the TEI Header, and the current TEI encoding has been removed with an XSLT-Stylesheet to extract the actual text content, omitting the metadata, etc.

Working with Stylo()
Introduction into theory and practice of stylometry in general and the use of “Stylometry with R” (Stylo) in particular and introduction into Gephi, the visualization software for graphs and networks by Dr. Jan Rybicki. Getting familiar with Stylo.

- Preparing corpus, primary_set, secondary_set, test_set, and reference_set for the analysis
- Cluster Analysis and Bootstrap Consensus Networks of Spanish, French, and Italian Spectators to compute the distances between individual journals and individual texts.
- Various analysis were conducted and methods applied to investigate journals published by anonymous authors (e.g. imposters, rolling classif(), etc.). These were conducted on both original journals and translations from the English prototype “The Spectator”.
- Analysis of specific narrative forms in the texts based on the TEI encoding, in this case the reader’s letters. For this purpose, the texts have been separated into reader’s letters and others.
- Getting familiar with rolling.classify() to investigate co-authored texts. The method has been applied to two Italian journals, Gli osservatori veneti published by Gasparo Gozzi and Il Caffé by Pietro e Alessandro Verri.
- The results have been visualized with the tools integrated into Stylo; for visualizing the graphs resulting from the Bootstrap Consensus Networks, Gephi has been applied.

Scientific exchange

- Regular meetings with Dr. Jan Rybicki.
- Participation in the weekly MA seminar in Stylometry (Translation Studies programme) taught by Dr. Jan Rybicki.
- Exchange with Prof. Maciej Eder, Joanna Byyszuk, and Magdalena Król (Institute of Polish Language, Polish Academy of Sciences and members of the Computational Stylistics Group).
- Presentation of own work at the Digital Humanities Lunch organized by the Institute of Polish Language (Polish Academy of Sciences).

DESCRIPTION OF THE MAIN RESULTS OBTAINED

Reader’s letters

There have been some interesting results regarding the reader’s letters in the Spectators texts which will now be evaluated together with experts in the field of Romance and Literary studies at the University of Graz. Since the genre presents itself to the outside world as very open and non-academic, it consists in large parts of letters to the editor. Thus, the integration and organisation of the diverse letters is given special consideration. The question is whether stylometric analyses can be used to determine the extent to which the texts actually rely on letters to the editor from the public. The question deserves special attention because the single issues were published weekly or at even shorter intervals, at the beginning of the 18th century, when communication required considerable time and effort. Therefore, two Italian journals edited by Gasparo Gozzi have been investigated, the “Gazzetta veneta” and “Gli Osservatori veneti”. Both show distinct signals of Gozzi being the only author. In comparison, another corpus (“Gazzetta urbana veneta”) now needs to be tested where evidence exists that the reader’s letters are actually from the public, i.e. a variety of different contributors.

Anonymous editors/authors

The resulting dendrograms (using different features) of the Cluster Analysis and the Bootstrap Consensus Tree of the French Spectators journals has shown which journals from anonymous authors have proximity to other journals (e.g. Anonymous 8 to Boisbeleau; Anonymous 2 to Desfournieux, etc.). For an authorship verification, the imposters() method has been tested. While the results seem decisive and are certainly promising, they will have to be evaluated in collaboration with the literary experts.

Co-Authored Texts

The Italian journal “Il Café”, published by Pietro e Alessandro Verri is known to be a co-authored text. With rolling.classify(), different author signals were identified. A sufficient amount of existing reference material from three suspected authors delivered strong signals in the course of investigation, thereby validating the method for this case. Of some additional authors, there is a lack in comparable material to apply the method successfully. At this point it will be necessary to identify additional texts from other sources by the suspected authors, to proof their authorship and identify their contributions within the journals’ corpus.
FUTURE COLLABORATIONS (if applicable)

The methods and tools learned during the STSM in Kraków will now be applied in a dedicated research project examining Distant Reading of the Spectators, funded by the Austrian Academy of Sciences. The colleagues from Kraków will be available for comments and exchange of expertise throughout the project, providing their expertise and in return gaining valuable input on the Close Reading validation of Distant Reading analysis results from the colleagues at the University of Graz (in terms of literary studies, editing and data science).

Furthermore, the TEI encoded and pre-processed texts of the Spectators will be contributed to the corpus of the COST Action CA16204.