



Text Analytics in Gender Studies. Introduction.

Fiorenza Deriu & Domenica Fioredistella Iezzi

To cite this article: Fiorenza Deriu & Domenica Fioredistella Iezzi (2020) Text Analytics in Gender Studies. Introduction., *International Review of Sociology*, 30:1, 1-5, DOI: [10.1080/03906701.2020.1724365](https://doi.org/10.1080/03906701.2020.1724365)

To link to this article: <https://doi.org/10.1080/03906701.2020.1724365>



Published online: 04 Mar 2020.



Submit your article to this journal [↗](#)



Article views: 168



View related articles [↗](#)



View Crossmark data [↗](#)



Text Analytics in Gender Studies. Introduction.

In recent years, in all fields of knowledge, a data-driven approach has spread according to the new scenario defined by the Big Data era. The so-called *data deluge* has started a season where an impressive amount of data constitutes a valuable research material for scholars. In this new context, the data-driven approach enables academics and scientists to examine and organize data with the goal of increasing knowledge in many research areas. The deluge of data today allows us to plan new analyses on a variety of unstructured data that are produced in major part by web navigation. Recent estimates maintain that 80% of all data is textual data. Furthermore, information that comes from social networks and social media, like Facebook, Twitter, and Instagram, produces unstructured data in real time.

Unstructured data is not organized according to a predefined scheme, and information resulting from these sets of data is typically text-heavy. Anyway, it may contain data such as dates, numbers, and facts as well. This results in irregularities and ambiguities that make it difficult to understand their meaning using traditional programmes as compared to data stored in structured databases. This flow of data has allowed the development of new methods and models, i.e. sentiment analysis to measure the mood of individuals and capture gender differences in language. Sentiment analysis, which is also called opinion mining, has been one of the most active research areas in natural language processing since early 2000 (Liu, 2015), and the constant refinement of analytical tools is offering a richer array of opportunities to analyse these data for many different purposes.

These broad assets of data are nowadays available and largely accessible. They represent both a strategic opportunity and a powerful challenge for researchers enabling them to find out new paths for social life exploration, and, more specifically, for the comprehension of the complex relationships intersecting some key concepts in gender, feminist and women/men sexual identity studies.

Data is mainly textual and produced by fertile human communication and exchange activities that take place on an increasing number of social platforms with a powerful viral potential. Human relationships are shaped at different levels of abstraction, fluctuating in the virtual space of the web-net. In these virtual spaces people are actively engaged, performing their agency, and managing their voice contributing to the ‘reproduction of or the resistance to gender arrangements’ in a community (Holmes & Meyerhoff, 1999:180). For this reason, it is worth studying how the concept of gender is built in the social practice of everyday life through multiple interactions, on the one side – and how it intersects other concepts and dimensions of the same research field, on the other. The Community of Practice approach, proposed by Holmes and Meyerhoff (1999), is actually compatible with the social-constructionist theory.

This new kind of data has also changed qualitative research in gender studies. Kumsal Bayazit, Chief Executive Officer of Elsevier (2019) said ‘To make progress in gender inclusion, we need to be able to measure where we are and where we want to be’. An increasing number of scholars states the importance of carrying out research based on multiple-approaches methodologies, combining the strength points of micro–macro level analysis techniques (Bergvall, 1999). In fact, there are two main needs: on the one hand, the need to locate the speech in

the daily experience of women and men working on-site, where every day people negotiate and challenge definitions and concepts about gender issues; on the other hand, to reflect at a broader level, trying to formulate theories framing and shaping gender differences (Bergvall, 1999). The norms of speech usage can emerge from statistical textual analysis providing a powerful framework that can shed light not only on content of language, but also on how people use it, how they construct their own identities (Holmes, 1998; Holmes & Meyerhoff, 1999), and how they influence public opinion, trying sometimes to move women back to their traditional and subordinate role. In order to study how people with different sexual identities use language, we need to keep our mind free from prejudice, and, as Cameron (1998) suggests, look at the structures of language, its contents and multiple use of words, with a special attention. In fact, words significantly contribute to the definition of those concepts shaping the cultural differences and multiple identities that should be studied and understood (Ampofo, Beoku-Betts, Njambi, & Osirim, 2004).

The complexity of these questions is reflected in the fragmented and irregular scientific literature of this research field, within which the tradition of feminist scholars shows different focus and interests, when compared to their colleagues working on gender and sexual identities studies topics (Stephen, 2000). In order to plot out the multiple different concepts resulting from the proliferation of studies in this field, the focus on methodology and methods seems to represent the *fil rouge* that can balance the ongoing fragmentation of ideas and concepts. In fact, many research questions are still feeding the scientific debate in the field of gender, feminist and women's studies, and some of them are included in the contributions to this special issue of the *International Review of Sociology*.

This special issue collects four innovative papers dealing with different problems related to gender studies, but which have as their common thread the applications of text analytics techniques. These studies were discussed during the 14th International Conference on the Statistical Analysis of Textual Data, that was held in Rome (Italy), from 12 to 15 June 2018.

Text analytics has its roots in the fields of linguistics, computer science, statistics and social sciences. Computer science and social sciences have given a strong impetus to converting qualitative information into quantitative analysis. According to Hearst (1999), text analytics is the automatic discovery of new, previously unknown, information, from unstructured textual data. Its process involves three major tasks: information retrieval (gathering the relevant document), information extraction (unearthing information of interest from these documents), and data mining (discovering new associations among the extracted pieces of information). Text analytics has its origins in the 1940s. The first studies in this field involve Computational Linguistics, Natural Language Processing, and Content Analysis. This phase lasted about 30 years. In the late 1980s and early 1990s, latent semantic indexing, or latent semantic analysis, developed. In the 1990s machine learning methods gained prominence in textual data, and text mining became a popular buzzword. Since 2000, publications employing text analytics techniques have grown enormously, and topic model and sentiment analysis are widely applied in gender studies.

This themed section is doubly focused – on both methodology and techniques used to address crucial topics at the core of the discussion on gender. So, the first contribution, using a lexicometric approach, offers a highlight on different ways to approach gender issues at international level, analysing the articles published in the last two decades by the *International Review of Sociology*; the second one, analysing the semantic of Tweets, using a multi-modal content analysis, shows the growing privatization of LGBTQI+ movements and the growing strengthening of gender hierarchies and heteronormativity; the third one, sheds light on a topic that is relatively under-investigated: the hate speech populating the social media that reframes misogyny as 'acceptable' by constructing it as a form of humour (Drakett,

Rickett, Day, & Milnes, 2018). The last contribution focuses on the press's language about immigrants and refugees in Spain, adopting a gender perspective.

The above-mentioned studies use three different types of textual data: scientific texts concerning gender and women's studies, published over a long period of time by a well-known international review of the sociological community; texts concerning posts published on Twitter, a social media platform characterized by the variety of its users as well as the speed of its messages diffusion, at the core of various debates aimed at both creating consensus about social and political matters, and driving hatred and dislike, intolerance and prejudicial sentiment against those who Sumner (1906) would call 'outgroup'; finally, texts of articles published in newspapers and daily press representing the press's language.

These different texts have been analysed applying mixed methods of analysis, and using a variety of software supporting the treatment of such datasets. The first essay of this themed session, authored by Nocenzi-Mingo, analyses the texts of 67 selected scientific articles published by the *International Review of Sociology* over the past 20 years, from 1997 up to 2017, presenting a relevant focus on gender at both theoretical and empirical levels. The strategy of analysis combines lexical analysis with factorial techniques, applied to aggregated lexical tables of specific lemmas and texts, partitioned by two layering variables: year of publication (time variable) and author's gender (male, female, and mixed). Finally, the analysis is completed by applying a descending hierarchical cluster analysis to the lexical table crossing text segments and lemmas. The integration of different techniques has enabled the authors to put into evidence both the most relevant gender, and women's issues addressed in many different research fields, such as economics, politics, culture, family, work, confirming the mainstream character of these area of studies. At the same time, the diachronic analysis seems to confirm the assumption of Stephen (2000), who points out the irregular and fragmented representation of literature in this field of studies. In fact, the factorial space does not show a chronological development of the sequence of text partitions over the time period considered, drawing the attention of scholars to the need to make more efforts on trying to organize the wide knowledge accumulated in the study of gender, feminism and women's issues.

The second article, authored by La Rocca-Rinaldi, is aimed at understanding how the cloud of feelings created on Twitter by two famous LGBTQI+ icons, can be attributed to a true globally mediatized emotional exchange, and, in this case, to the world of LGBTQI+ that these icons express. In order to run this analysis, the authors use data generated on the social media platform Twitter. Even in this case, the authors apply a multimodal content analysis, that results in a merged method (Driscoll, Appiah-Yeboah, Salib, & Rupert, 2007) combining two techniques – the content analysis and the multimodal discourse analysis – allowing the decomposition and the following re-composition of polysemic communication. The focus of this work is on rebuilding the sense and meaning of hashtags concerning the different representation of the LGBTQI+ community, through two famous members, participating into two popular reality shows, *Grande Fratello Vip 2017* and *L'Isola dei famosi 2008*. The starting assumption is that sexuality, sexual behaviour and practices are of fundamental importance to understand the fast-paced transformation of identities; and the study of hashtags offers an interesting perspective on the analysis of communication expressions that can be considered speech acts. The study shows how the characters observed contribute to challenging the gender or sexual regime within a public 'arena'.

The third contribution, by Dragotto-Giomi-Melchiorre, offers an accurate discussion about the functions, social uses, and discursive mechanism of slut-shaming on the basis of the hypothesis that this phenomenon is the expression of a gendered practice of power aimed at 'putting women back in their place'. They start from the case of Asia Argento, her experience as a leader of the MeToo movement, and the story related to the sexual harassment she

experienced from Harvey Weinstein at the beginning of her career, in order to discuss and corroborate the initial hypothesis. The methodology is once again mixed; in fact, a quali-quantitative approach is used. The authors have prepared two corpora, scraping from Twitter the posts of interest marked as #AsiaArgento, over five months up to 20 March 2018. Then the authors have conducted a three-step analysis to shed light on the issue they were investigating, combining quantitative techniques (sentiment analysis) and qualitative ones, trying to identify the different narratives characterizing the corpus of data. Notwithstanding the limited dimension of the corpus, results are consistent with previous studies evidences, so it can be give a useful contribution to reflections on this specific topic, confirming, once again, that slut-shaming is a way through which the dominion and supremacy of men over women is perpetuated (Cowie & Lees, 1981).

Finally, the last contribution of this thematic issue, authored by Torvisco-Chimea, addresses the question of immigrants and refugees, with the aim of studying the news published by newspapers and online daily press, integrated with the information published on Facebook, in correspondence of two main waves of migrations in Spain in 2006 and 2015. In this research, an interesting aspect is the integrated use of two very different sources of textual data: articles and posts on social media. The authors evidence how journalistic narratives on these social phenomena are not neutral; the opinions expressed in press articles are strongly ideologically oriented. The analysis has highlighted some similarities in the narratives of migrants' and refugees' stories, as well as some differences, also related to the different regulation of fluxes of the period considered in the study.

The four contributions do not have the ambition to be exhaustive of text analytics in gender studies, but they can surely offer some useful, and also critical, insights from a double perspective – both methodological and theoretical. The growing interest in and practice of methodologies and techniques integration shows an important route to follow for future studies in this research field.



References

- Ampofo, A.A., Beoku-Betts, J., Njambi, W. N., & Osirim, M. (2004). Women's and gender studies in English-speaking sub-Saharan Africa: A review of research in the social sciences. *Gender and Society*, 18(6), 685–714. Retrieved from www.jstor.org/stable/4149390.
- Bayazit, K. (2019). Progress in improving gender diversity in science, Gender Summit Europe, Elsevier's CEO, 3-4 October 2019, GS17, Amsterdam, Netherlands.
- Bergvall, V. (1999). Toward a comprehensive theory of language and gender. *Language in Society*, 28(2), 273–293. Retrieved from www.jstor.org/stable/4168929.
- Cameron, D. (1998). Gender, language, and discourse: A review essay. *Signs: Journal of Women in Culture and Society*, 23(4), 945–973. Retrieved from www.jstor.org/stable/3175199.
- Cowie, C., & Lees, S. (1981). Slags or drags. *Feminist Review*, 9, 17–31.
- Drakett, J., Rickett, B., Day, K., & Milnes, K. (2018). Old jokes, new media – Online sexism and constructions of gender in Internet memes. *Feminism & Psychology*, 28(1), 109–127.
- Driscoll, D. L., Appiah-Yeboah, A., Salib, P., & Rupert, D. J. (2007). Merging qualitative and quantitative data in mixed methods research: How to and why not. *Ecological and Environmental Anthropology*, 3(1), 19–28.
- Hearst, M. (1999). *Untangling Text Data Mining*. Proc of ACL'99: The 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland, June 20–26, 1999.
- Holmes, J. (1998). The question of sociolinguistic universals. In J. Cotes (Ed.), *Language and gender: a Reader* (pp. 461–483). Oxford: Blackwell.
- Holmes, J., & Meyerhoff, M. (1999). The community of practice: Theories and methodologies in language and gender research. *Language in Society*, 28(2), 173–183. Retrieved from www.jstor.org/stable/4168923.
- Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge: Cambridge University Press.
- Stephen, T. (2000). Concept analysis of gender, feminist, and women's studies research in the communication literature. *Communication Monographs*, 67(2), 193–214. doi:10.1080/03637750009376504.

Sumner, W. G. (1906). *Folkways: A study of the sociological importance of usages, manners, customs, mores, and morals*. New York, NY: Dover Publications.

Fiorenza Deriu

Department of Statistics, Sapienza University of Rome, Rome, Italy

 fiorenza.deri@uniroma1.it  <http://orcid.org/0000-0003-4940-5343>

Domenica Fioredistella Iezzi

Department of Enterprise Engineering, Tor Vergata University, Rome, Italy