

Impact of COVID-19 in Natural Language Processing Publications: a Disaggregated Study in Gender, Contribution and Experience

Christine Basta and Marta R. Costa-jussà

TALP Research Center, Universitat Politècnica de Catalunya, 08034 Barcelona
{christine.raouf.saad.basta,marta.ruiz}@upc.edu

Abstract

This study sheds light on the effects of COVID-19 in the particular field of Computational Linguistics and Natural Language Processing within Artificial Intelligence. We provide an inter-sectional study on gender, contribution, and experience that considers one school year (from August 2019 to August 2020) as a pandemic year. August is included twice for the purpose of an inter-annual comparison.

While there has been an increasing trend in publications during the crisis, the results show that the ratio between female authors' publications and male authors' publications decreased. This only reduces the importance of the female role in the scientific contributions of computational linguistics (it is now far below its peak of 0.24). The pandemic has a significantly negative effect on female senior researchers' production in the first position of authors (maximum work), followed by the female junior researchers in the last position of authors (supervision or collaborative work).

1 Introduction

The pandemic situation is a burden for everyone. Given the uncertainty and pressure of the situation, it is expected to have a significant impact on the lives of researchers. However, it appears that women are bearing considerable burdens due to the social impact of the virus. According to the Scientist article ¹, the pandemic exacerbates existing inequalities, with female researchers' output increasing less during the pandemic. This is due to several factors: housework, children staying at home, and many other factors. In our paper, we aim to investigate the effects on the productivity of female researchers during the pandemic in Natural

Processing Field (NLP). Women have long been underrepresented in NLP research. This fact leads to the question of whether the pandemic affects female NLP researchers more. We focus on testing the following hypothesis:

1. Although the pandemic was a stressful period, it has increased the scientific production, since distance-based working favours writing.
2. As first authors, production has decreased among the senior women, considering that they are more susceptible to doing housework and taking care of children (or elder people) at home.
3. In a collaborative work, while working remotely, the contributions of junior women seems to be disregarded and they are not included unless their contribution is predominant. This decreased of scientific production of the junior women as last authors.

2 Impact Statement

This work is doing a binary gender study. In that sense, we are simplifying the real gender spectrum (D'Ignazio and Klein, 2018). Our motivation is because we want to study the effect of COVID-19 in women, in the line of previous works (Schluter, 2018).

By highlighting the role of women in scientific contributions, we emphasize the importance of strengthening the educational programs, scientific grants and support that motivate woman in NLP and in STEM (science, technology, engineering and mathematics) in general.

¹<https://www.the-scientist.com/news-opinion/gender-gap-in-research-output-widens-during-pandemic-67665>

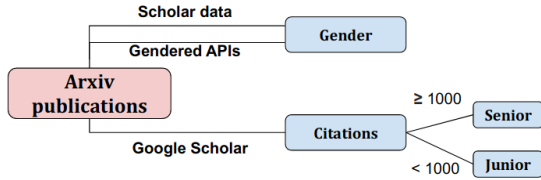


Figure 1: Data diagram

3 Related work

Several studies that show the imbalance between females/males emerge with different objectives in NLP. We can assume three main categories that show imbalances: in our data, in the automatic representations from our systems and in the professional level. Our study falls into this third category.

Data. Similar to the study of the number of video frames in computer vision, e.g., (Madaan et al., 2018b), which analyses the number of dedicated frames for women and men, there is a study in NLP that examines the gender-specific references in the literature (Madaan et al., 2018a). This study shows a higher disproportion in mentioning men more than women. More than the disproportion, it is the role that men are usually associated with "high-level" occupations compared to women who are associated with more "care-giving" occupations.

Automatic Representations. Our systems have shown that they perpetuate and amplify biases (Bolukbasi et al., 2016; Caliskan et al., 2017) as examples in word representation. We have also seen biases in applications such as machine translation (Prates et al., 2020), sentiment analysis (Kiritchenko and Mohammad, 2018) and others.

Professional level. A deep study of the *glass ceiling* (Schluter, 2018) shows profound evidence of the existence of such limitations. It is worth mentioning that while doing our study, there were parallel studies in several fields which have led to similar conclusions (Büyüm et al., 2020; Coleman et al., 2021; Ancochea et al., 2020; Ruiz-Frutos et al., 2021; Spagnolo et al., 2020; Squazzoni et al., 2020; Sohrabi et al., 2021; King and Frederickson, 2020).

4 Data

In this section, we describe the procedure for extracting the data that allows us to discuss the impact of COVID-19 on the scientific production of NLP papers in terms of gender, contribution and experience. This data is available to other researchers².

²<https://github.com/ChristineBasta/Cov19PubStats>

We select a source of papers, extract them, and classify the authors according to gender and experience. The contribution is measured according to the criteria of the first author and the last author: The first authors are the maximum contributors and the last authors are the collaborators or supervisors. We show a block diagram of the entire data extraction process in Figure 1 and explain the details in the following sections.

4.1 Paper selection

We have decided to get a source of publications that can be easily updated daily and overall months. Therefore, Arxiv³, an open access archive for scientific articles, was chosen as a source with such criteria. We have extracted from the Arxiv NLP publications of the years 2019 and 2020 from August 2019 to August 2020. In order to extract those of the NLP category, we specified the category *cs.cl*, which corresponds to the field of Computer Science and Computation and Language. This category corresponds to the category ACM i.2.7 (Natural Language Processing). We found about 6760 unique articles, starting in August 2019. Given this data, we now have to infer gender and experience to the authors of the papers. The following sections describe how to infer the gender of the names and how to classify them on the experience level.

4.2 Inferring gender of names

The study (Mohammad, 2020b) examines the citations and authors of the publications of the ACL anthology. In this dataset, called NLP-scholar, we can find the gender of the first and last authors. Thus, this was the first direction to find out the gender of the authors. To investigate the gender of authors, not available in this dataset, we used two APIs^{4,5}, which allow extracting the gender of the name to ensure that the gender of the name is correctly recognized. We use the former to predict the names. However, if the name is predicted with less accuracy, we use the latter for verification. It is a difficult task to guess the gender from Chinese names (van de Weijer et al., 2020). While there are open tools that can make this guess⁶, we are not aware of any tools for the case of Chinese names written in English. Therefore, we have decided to

³<https://arxiv.org/>

⁴<https://genderize.io/>

⁵<https://pypi.org/project/gender-guesser/>

⁶<http://www.chinese-tools.com>

neglect them when they do not exist in the NLP-scholar dataset.

4.3 Senior and Junior classification

Google Scholar is a free web search engine for academic literature to access metadata for authors and publications. Google Scholar does not provide an official API to extract information about authors, but we have used scholarly⁷ to extract citations of the authors from their Google Scholar profile pages. A Google Scholar Profile page is the page where authors can include their papers (citations are included) and their affiliations (Mohammad, 2020a). Seniors are classified as authors with more than 1000 citations, while juniors are those with less than 1000 citations. Sometimes the same author name can have several profile pages. Accordingly, we have performed a manual inspection to ensure that the interests of the extracted author include a related field, e.g. 'computer science', 'Natural Language Processing', etc.

5 Discussion

Our discussion will target solving the hypothesis raised in section 1 together with additional ones.

What is the effect of pandemic on NLP publications? Our study examines the total production of NLP research during the year and the gender distribution. In the figure 2, we observe an increase in the number of both female and male authors during the period from March to July 2020, which confirms *hypothesis 1* from section 1. This may be explained by the fact that scientific work has gained popularity during the pandemic, together with the fact that working from a distance tends to leave more time for writing articles.

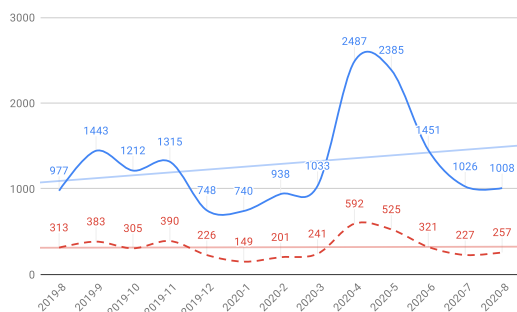


Figure 2: Total number of male (blue) and female (dashed-red) authors per month.

⁷<https://pypi.org/project/scholarly/>

What is the ratio of females/males contributions? There is a real gap in the contributions of females/males in NLP, which varies between 0.16 and 0.24 in monthly terms (figure 3). Within the period under study, this gap gets larger, which brings us to the next question.

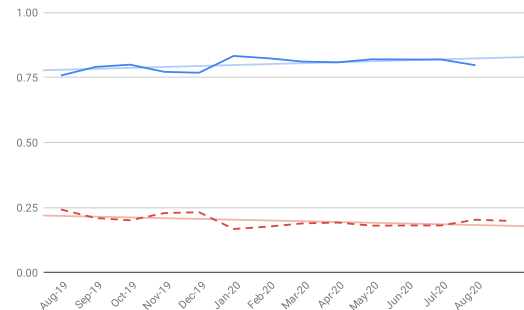


Figure 3: Ratio of men (blue) and women (dashed-red) in total authors

What is the effect of the pandemic in the ratio of females/males contributions? The tendency during this period is the reduction of the females/males ratio. We can analyze the effect in this period in two different ways, but with similar conclusions. First, if we compare production in August 2019 and 2020, the females/males ratio has decreased from 0.24 to 0.20 (figure 3). Second, a comparison of the average production in the months before the first lockdowns⁸ (August 2019-February 2020) and for the months of 2020 with intermittent lockdowns (April-August). In the former case, the ratio is 0.21, whereas, in the latter, the ratio is reduced to 0.18.

What is the effect of the pandemic in the ratio of females/males first and last authors contribution? The ratio of females/males first authors is stable (figure 4 (top)), whereas the ratio of females/males last authors decreases slightly (figure 4 (bottom)).

What is the effect of the pandemic on the ratio of females/males first authors contribution related to experience? The ratio of females/males first authors during the studied period increases among juniors (figure 5 (top)) and decreases significantly among seniors (figure 5 (bottom)), confirming *hypothesis 2*. The influence of the pandemic is less relevant for junior female scientists than for senior female researchers. This reinforces the hypothesis that the productivity of senior female researchers declined. This could be explained due to the combination of "care-giving"/family and professional re-

⁸https://en.wikipedia.org/wiki/COVID-19_pandemic_lockdowns

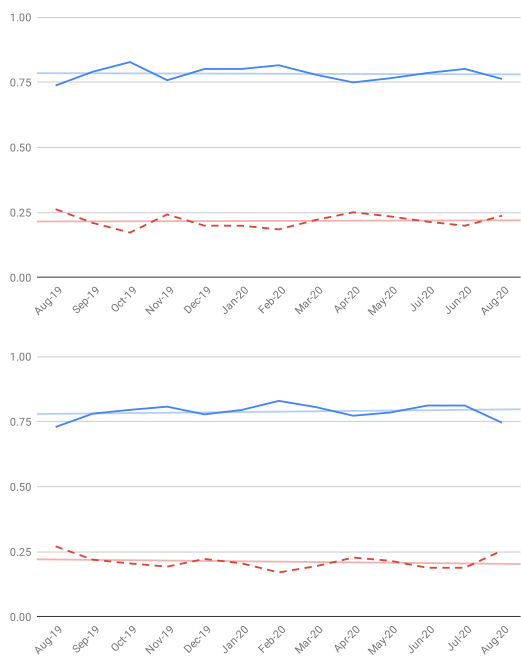


Figure 4: First (top) and last (bottom) authors.

sponsibilities (Zimmer, 2020). However, we want to make clear that there is no proven correlation between care-giving responsibilities and the level of seniority based on number of citations provided.

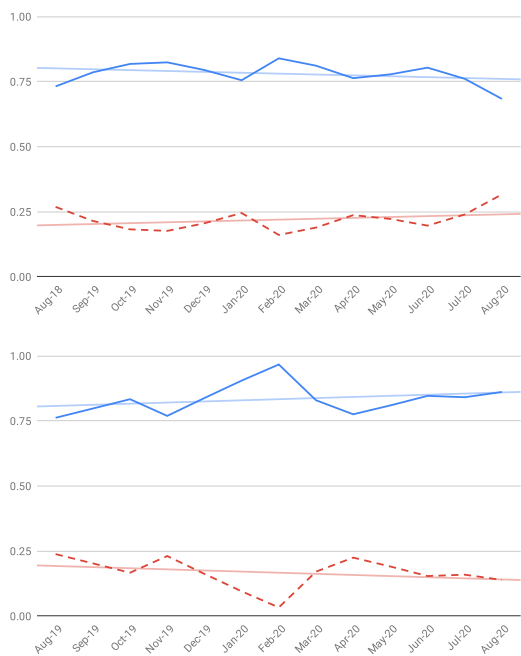


Figure 5: First authors: juniors (top); seniors (bottom)

What is the effect of the pandemic on the ratio of females/males last authors contribution related to experience? Here we see an interesting effect, which is that the ratio females/males ratio decreases

in the juniors (figure 6 (top)), which confirms *hypothesis 3*, while in the seniors, the ratio is stable (figure 6 (bottom)). It is noted that there are opposite effects on junior female scientists, depending on whether they occupy the last or the first position in the publication. This is not straightforward to explain, but we can speculate about the following. In the case of leading the work as first authors (figure 5 (top)), the proportion increases compared to men. In contrast, the proportion decreases when women are the last authors. From our point of view, this could show that female junior researchers are less likely to be involved if they do not lead the work, at least during the pandemic. This could be a form of discrimination against women in the sense that some of their collaborations could be neglected. This is pure speculation, as studying the factors which support this hypothesis is out of the scope of the paper. In the case of the senior women, their production is stable as last authors (as opposed to the decline as first authors), possibly due to the fact that they supervise the work at a higher level of detail and this is less demanding in terms of dedication. Therefore, in the case of the last authors and the experience level, we do not see a decline in production during the period under study.

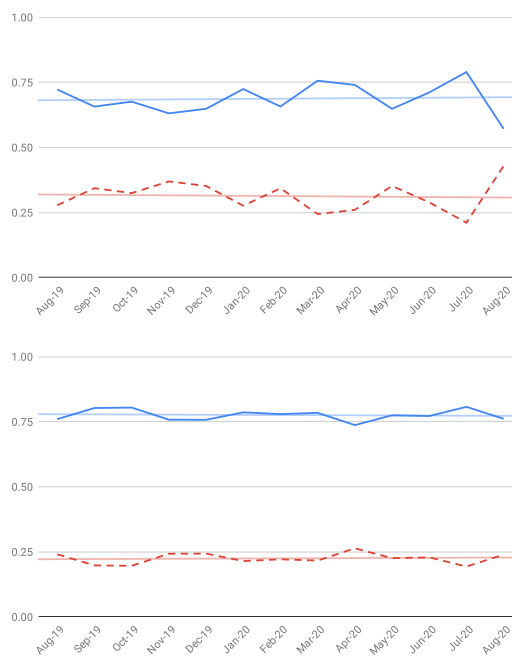


Figure 6: Last authors: juniors (top); seniors (bottom)

6 Conclusions

The main goal here is to estimate the pandemic effect on the NLP production. The impact was studied in terms of gender, contribution and experience. When combining the three criteria, different effects are noted. In the first position (maximum contribution), the females/males proportion increases for juniors and decreases for seniors. This can be explained by the fact that the senior researchers had to take on more responsibility (family, care-giving) during the crisis, which prevented them from keeping up with scientific production. In the last position (supervision or collaboration contribution) the females/males proportion decreases for juniors and is kept stable for seniors. It could be speculated that female junior researchers were more neglected than male junior researchers in terms of small contributions when working at a distance.

7 Room for improvement

Since the data is just for a year, we do not have any previous trends to show seasonal variation patterns. Without countering for seasonal variation we may at least question whether the patterns are for pandemic or part of the natural variation/fluctuation in the system. To give an example, in Figure 2, we see a spike in the number of papers during April-May-June. This could be because we have external factors just as that the camera-ready deadlines are at that time, or anonymity period ends or begins just after this period, etc. To answer these questions, we need to look at the graphs for previous few years.

We are simplifying scientific productivity by assuming that it is equal or proportional to the number of publications. We are aware that publication count is just a proxy of productivity.

However, even taking into account these limitations, our study shows the pervasive gap between women and men (also disaggregated in contribution and experience) in scientific NLP publications during the COVID-19 first year.

Acknowledgements

This work is supported in part by the Catalan Agency for Management of University and Research Grants (AGAUR) through the FI PhD Scholarship.

This work also is supported in part by the project PID2019-107579RB-I00 / AEI / 10.13039/501100011033.

References

- Julio Ancochea, Jose L Izquierdo, Joan B Soriano, Savana COVID-19 Research Group, et al. 2020. Evidence of gender bias in the diagnosis and management of covid-19 patients: A big data analysis of electronic health records. *medRxiv*.
- Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Proceedings of Advances in Neural Information Processing Systems 29*, pages 4349–4357.
- Ali Murad Büyüm, Cordelia Kenney, Andrea Koris, Laura Mkumba, and Yadurshini Raveendran. 2020. Decolonising global health: if not now, when? *BMJ Global Health*, 5(8):e003394.
- Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora necessarily contain human biases. *Science*, 356:183–186.
- Julia R Coleman, Jad M Abdelsattar, Roan J Glocker, Navin G Vigneshwar, Randi Ryan, Qiong Qiu, Apoorve Nayyar, Michael R Visenio, Cheyenne C Sonntag, Pranit Chotai, et al. 2021. Covid-19 pandemic and the lived experience of surgical residents, fellows, and early-career surgeons in the american college of surgeons. *Journal of the American College of Surgeons*, 232(2):119–135.
- Catherine D’Ignazio and Lauren Klein. 2018. *Data feminism*. MIT Press.
- Molly M King and Megan Frederickson. 2020. The pandemic penalty: The gendered effects of covid-19 on scientific productivity.
- Svetlana Kiritchenko and Saif M. Mohammad. 2018. Examining gender and race bias in two hundred sentiment analysis systems. In **SEM@NAACL-HLT*.
- N. Madaan, Sameep Mehta, S. Mittal, and Ashima Suvarna. 2018a. Judging a book by its description : Analyzing gender stereotypes in the man bookers prize winning fiction. *ArXiv*, abs/1807.10615.
- Nishtha Madaan, Sameep Mehta, Tanea Agrawaal, Vrinda Malhotra, Aditi Aggarwal, Yatin Gupta, and Mayank Saxena. 2018b. [Analyze, detect and remove gender stereotyping from bollywood movies](#). volume 81 of *Proceedings of Machine Learning Research*, pages 92–105, New York, NY, USA. PMLR.
- Saif M. Mohammad. 2020a. [Examining citations of natural language processing literature](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5199–5209, Online. Association for Computational Linguistics.
- Saif M. Mohammad. 2020b. [NLP scholar: A dataset for examining the state of NLP research](#). In *Proceedings of The 12th Language Resources and Evaluation Conference*, Marseille, France. European Language Resources Association.

- Marcelo O. R. Prates, Pedro H. C. Avelar, and Luis Lamb. 2020. Assessing gender bias in machine translation – a case study with google translate. *Neural Computing and Applications*, 32, page 6363–6381.
- Carlos Ruiz-Frutos, Mónica Ortega-Moreno, Regina Allande-Cussó, Sara Domínguez-Salas, Adriano Dias, and Juan Gómez-Salgado. 2021. Health-related factors of psychological distress during the covid-19 pandemic among non-health workers in spain. *Safety Science*, 133:104996.
- Natalie Schluter. 2018. [The glass ceiling in NLP](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2793–2798, Brussels, Belgium. Association for Computational Linguistics.
- Catrin Sohrabi, Ginimol Mathew, Thomas Franchi, Ahmed Kerwan, Michelle Griffin, Jennick Soleil C Del Mundo, Syed Ahsan Ali, Maliha Agha, and Riaz Agha. 2021. Impact of the coronavirus (covid-19) pandemic on scientific research and implications for clinical academic training—a review. *International Journal of Surgery*.
- Jessica Spagnolo, Lara Gautier, Mathieu Seppey, and Nicole Anne D’souza. 2020. Re-thinking global and public health projects during the covid-19 pandemic context: Considerations and recommendations for early-and not-so-early-career researchers. *Social Sciences & Humanities Open*, 2(1):100075.
- Flaminio Squazzoni, Giangiacomo Bravo, Francisco Grimaldo, Daniel Garcia-Costa, Mike Farjam, and Bahar Mehmani. 2020. No tickets for women in the covid-19 race? a study on manuscript submissions and reviews in 2347 elsevier journals during the pandemic. *SSRN Electronic Journal*.
- Jeroen van de Weijer, Guangyuan Ren, Joost van de Weijer, Weiyun Wei, and Yumeng Wang. 2020. [Gender identification in chinese names](#). *Lingua*, 234:102759.
- Katarina Zimmer. 2020. Gender gap in research output widens during pandemic. *The Scientist*.