

# Gender Gap in Natural Language Processing Research: Disparities in Authorship and Citations

Saif M. Mohammad

Senior Research Scientist, National Research Council Canada

✉ [Saif.Mohammad@nrc-cnrc.gc.ca](mailto:Saif.Mohammad@nrc-cnrc.gc.ca)    [@SaifMMohammad](https://twitter.com/SaifMMohammad)



**Tara Small**

Head of Applied Research, CSE  
Canada



**Rebecca Knowles**

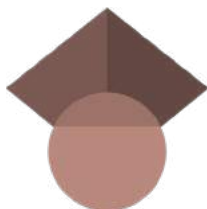
Research Scientist, National  
Research Council Canada



**Ellen Riloff**

Professor of Computer Science  
University of Utah

**Grateful  
Acknowledgments**  
for feedback and discussions



**Suzanne Cooper**

Senior Advisor to the Deputy Minister for  
Women and Gender Equality Canada



**Eric Joanis**

Research Council Officer,  
National Research Council Canada



**Isar Nejadgholi**

Research Scientist, National  
Research Council Canada

Also **Dan Jurafsky, Rada Mihalcea, Isabelle Augenstein, Shubhanshu Mishra, Svetlana Kiritchenko, Cyril Goutte, Cassidy Rae Henry, Luca Soldaini, Su Lin Blodgett, Graeme Hirst, Brendan T. O'Connor, Lucia Santamaria, Lyle Ungar, Emma Manning, and Peter Turney.**



# Gender Gaps

*Quantitative measures of the **disparities in success** (social, political, intellectual, cultural, or economic) **due to one's gender**.*

## Gender:

- People can have genders that do not fit conveniently in the male and female categories (non-binary)
- One's gender can be different from the biologically assigned sex (transgender)
- Different people can experience gender differently (or not experience gender— agender)
- Society has often viewed different gender groups differently, often imposing unequal social and power structures (Lindsey, 2015)

Thus, gender gaps are expected to differ for different groups.

# Why we should pay attention to Gender Gaps

- Inherently unfair
- Better gender balance leads to
  - higher productivity
  - better health and well-being
  - greater economic benefits
  - better decision making
- Gender gaps in AI can exacerbate other gender gaps



Our diversity impacts the breadth of technologies we create, how useful they are, and whether they reach those that need it most.

Measuring gender gaps is a crucial step towards addressing the gender gaps.

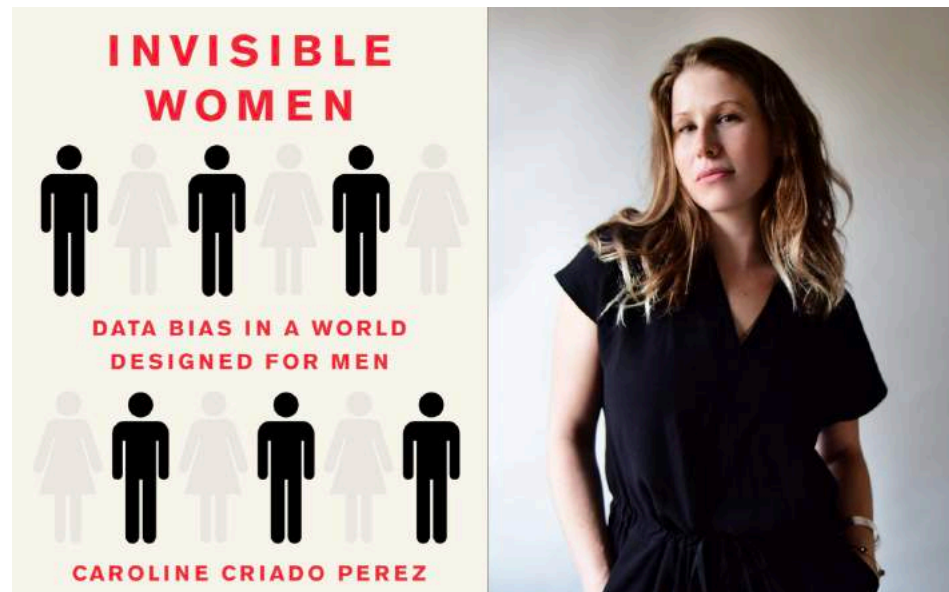
(Skjelsboek and Smith, 2001; Woetzel et al., 2015; Hakura et al., 2016; Mehta et al., 2017; Gallego and Gutierrez, 2018)

# Fighting Gender Gaps with Disaggregated Data

Considerable lack of disaggregated data for women ([Perez, 2019](#)):

- negative outcomes in all spheres of their lives
- health, income, safety, and the degree to which they succeed in their endeavors

This holds true even more for non-binary and transgender people.





# This Work

## Examining Citations of Natural Language Processing Literature. ACL 2020.

- Obtains disaggregated data for female Natural Language Processing (NLP) researchers
  - NLP: an interdisciplinary field that includes scholarly work on language and computation with influences from AI, CS, Linguistics, Psychology, Social Sciences, etc.
- Examines gender gaps between female and male researchers:
  - disparities in authorship
  - disparities in citations

Presented through a series of **eight questions** and answers.

**Future work:** explore other gender gaps, e.g., between non-binary and binary people, trans and cis people  
Other aspects of demographic diversity: nationalities, race, language, income, age, physical abilities, etc.

# Related Work: Gender disparities in research

## Outside NLP:

- Mathematics (Mihaljević-Brandt et al., 2016)
- Library and Information Science (Hakanson, 2005)
- Web of Science (for Sociology, Political Science, Economics, Cardiology and Chemistry) (Ghiasi et al. 2016)
- PubMed (life science and biomedical research) (Mishra et al. 2018)
- JSTOR (fifty disciplines) (King et al. 2017)
- Gender Gap in Science Project (<https://gender-gap-in-science.org>)

All of these found substantial gender disparities in favor of male researchers.

## NLP



- Schluter (2018): delayed attainment of mentorship status for women researchers
- Vogel and Jurafsky (2012): examined about 13,000 papers from 1990 to 2008
  - female authorship steadily increasing from the 1980s to ~27% in 2007



This work examines a much larger set of NLP papers published from 1965 to 2018, explores several new questions, especially on first author gender and disparities in citation.

# NLP Scholar Dataset

We extracted and aligned information from

- the ACL Anthology (AA) 
- Google Scholar 

to create a dataset of tens of thousands of NLP papers and their citations

**NLP Scholar: A Dataset for Examining the State of NLP Research.** Saif M. Mohammad. In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC-2020)*, May 2020, Marseille, France.

Also made use of Name-Gender Association resources:

- a. Manually curated list of ~12K AA authors and their genders provided by [\(Vogel and Jurafsky, 2012\)](#)
- b. US Social Security Administration's [\(USSA\)](#) database of names and genders of newborns
- c. 9,300,182 [PUBMED](#) authors\* and their genders [\(Torvik and Smalheiser, 2009; Smith et al., 2013\)](#)

a., b., and c. have considerable coverage from NA and Europe.

a. and c. have been shown to have considerable coverage from Asia.



# Name–Gender Association

- Strong normative tendency to assign birth names, and later keep a name or choose a new name, associated with one's gender (Pilcher, 2017; Barry III and Harper, 2014; Lieberman et al., 2000; Alford, 1987)
- Often the first name signals one's gender
- Author names in scientific literatures
  - perceived gender (from the name) can lead to unconscious citation effects (Dion et al. 2018)
  - finding only a small number of female authors in the literature is demoralizing
- Large body of work on automatically inferring gender by one's first name for:
  - scientometric analysis, public health studies, commerce, understanding language patterns

## Caution:

- Can reinforce harmful stereotypes such as the dichotomy of gender
- Can lead to *misgendering*—assuming one is of a different gender than what they are

**Our work:** Obtains disaggregated statistics for women NLP researchers (as a group)  
(does not infer individual author gender)

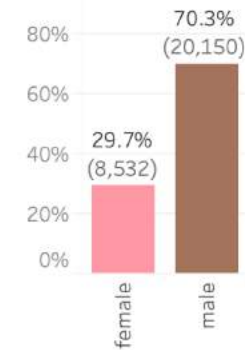
# Disparities in Authorship

Q1. What percentage of the authors in the ACL Anthology (AA) are female?  
What percentage of the AA papers have female first authors (FFA)?  
What percentage of the AA papers have female last authors (FLA)?  
How have these percentages changed since 1965?

### % female authors:

- 29.7% of the authors
- steadily increased in the early years, but stalled since mid 2000s

% authors by gender  
(among F and M authors)



### % papers by female first authors (FFA):

- 29.2% of the papers
- % female authors is close to the % papers with female first authors

### % papers by female last authors:

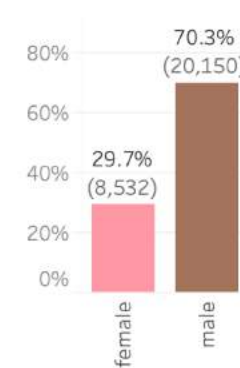
- 25.5% of the papers
- markedly lower than first author %



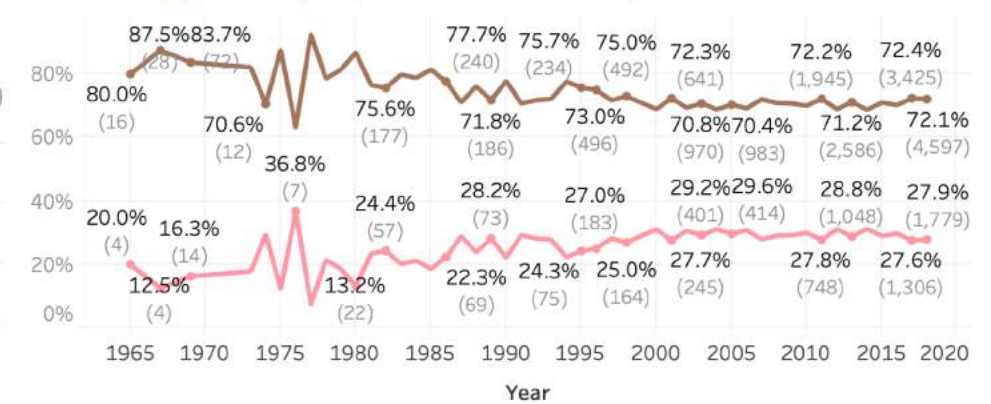
## % female authors:

- 29.7% of the authors
- steadily increased in the early years, but stalled since mid 2000s

% authors by gender  
(among F and M authors)



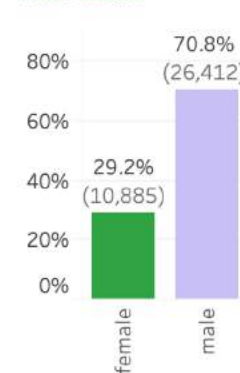
% of authors by gender each year (#authors shown in brackets)



## % papers by female first authors (FFA):

- 29.2% of the papers
- % female authors is close to the % papers with female first authors

% papers by gender of  
FIRST author



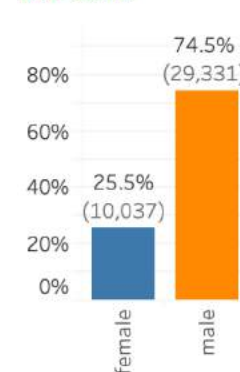
% of papers by gender of FIRST author each year (#papers shown in brackets)



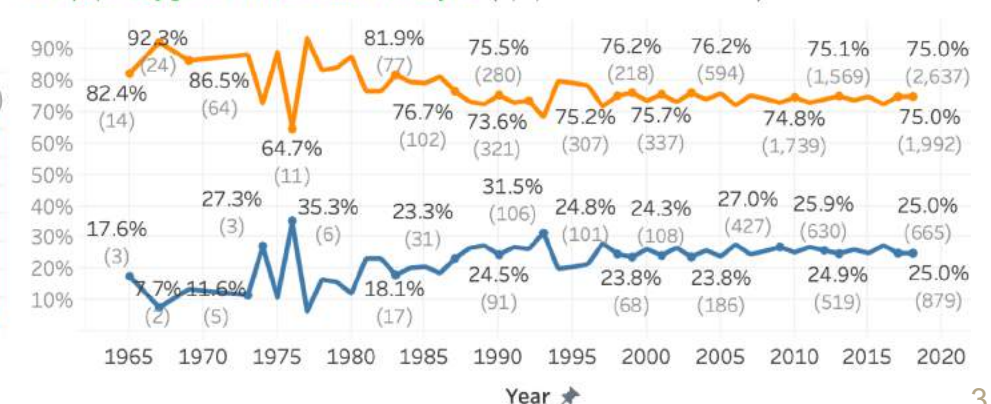
## % papers by female last authors:

- 25.5% of the papers
- markedly lower than first author %

% papers by gender of  
LAST author



% of papers by gender of LAST author each year (#papers shown in brackets)





As a community, we are far from obtaining male–female parity.

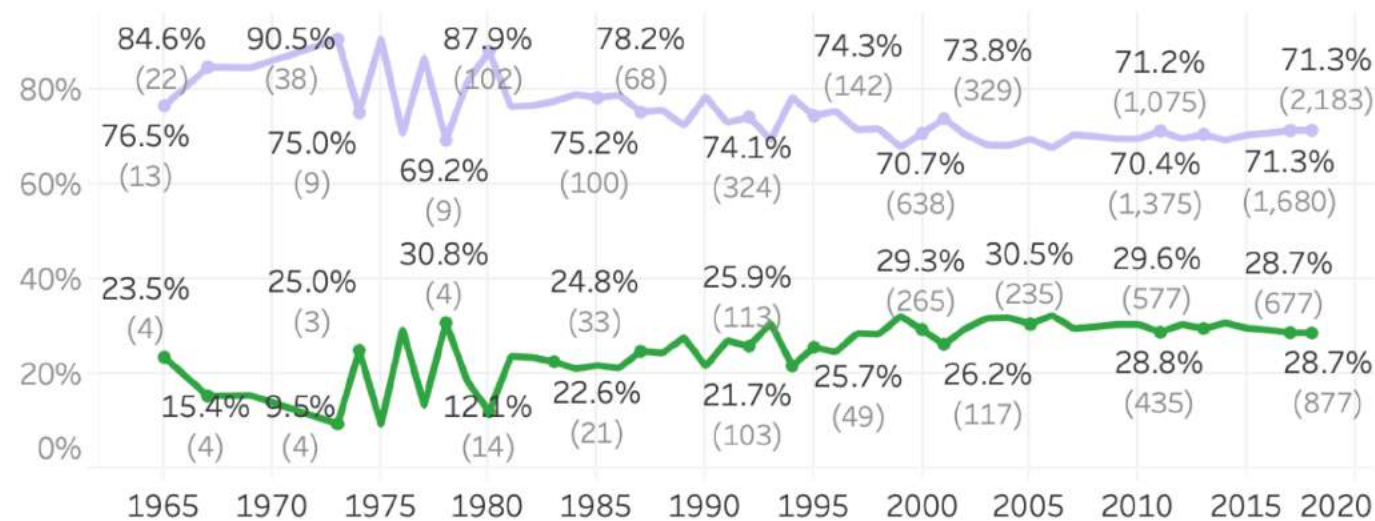
- stalled since mid 2000s

How does arXiv as a whole fare:

- female percentages are lower
- we see steady progress – even in the last decade
- yet at this rate it will take 100 years for male-female parity
  - if progress not stalled

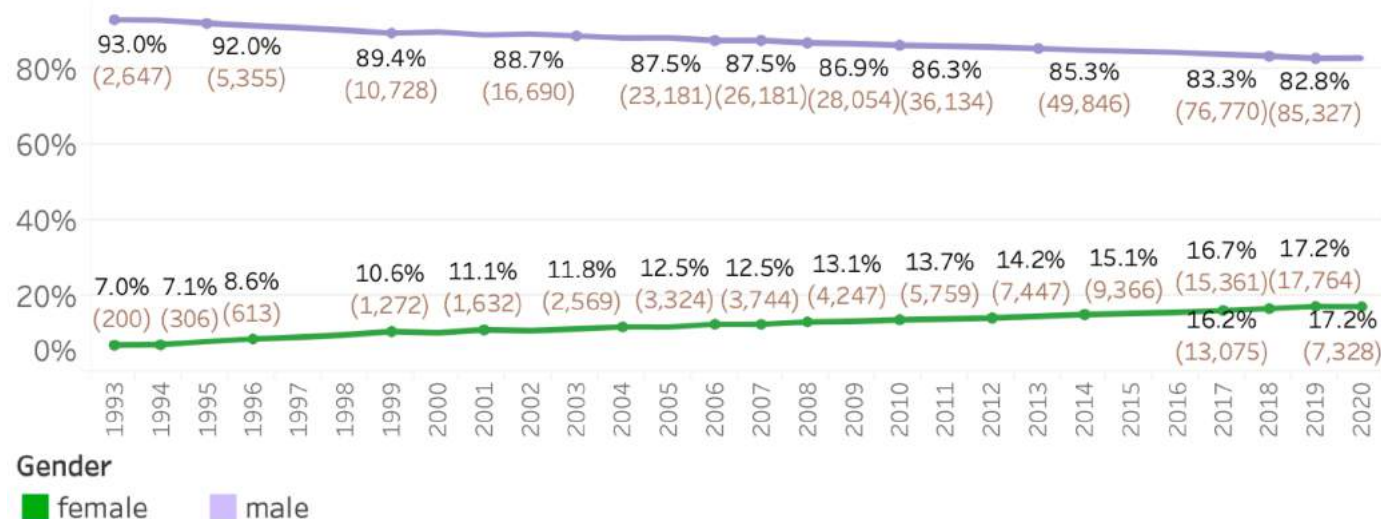
## NLP Papers

% of papers by gender of FIRST author each year (#papers shown in brackets)

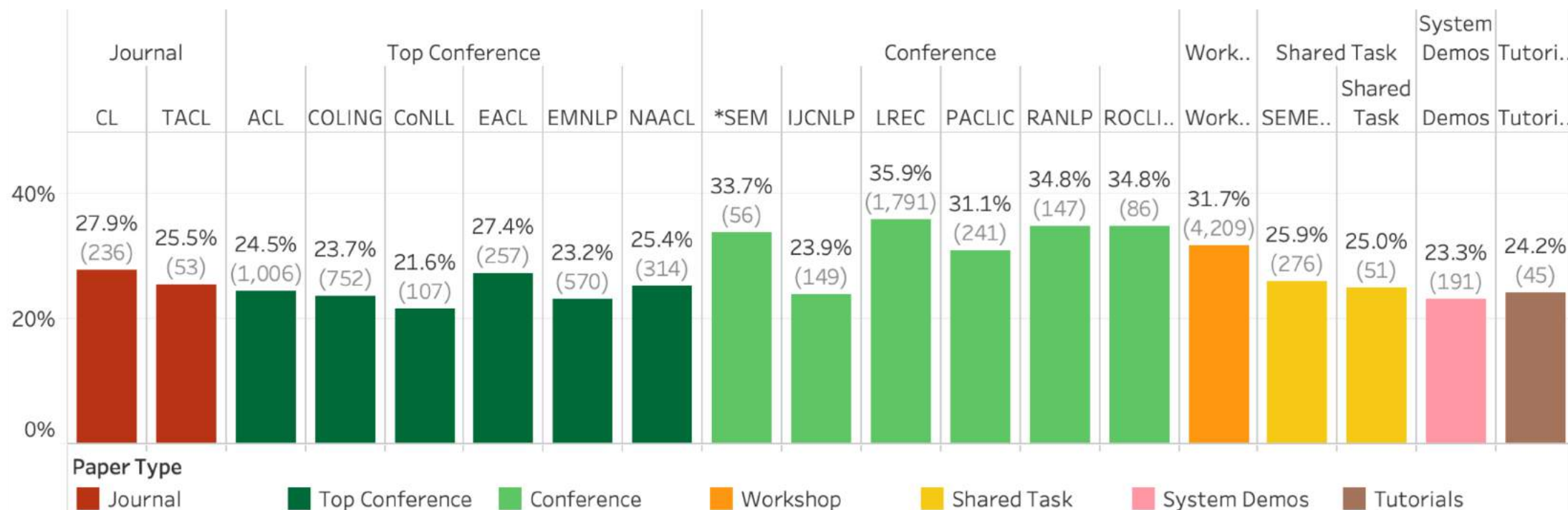


## Papers from Physics, Math, CS,...

arXiv: %papers by gender of FIRST author each year



## Q2. How does FFA% vary by paper type and venue?



FFA percentages:

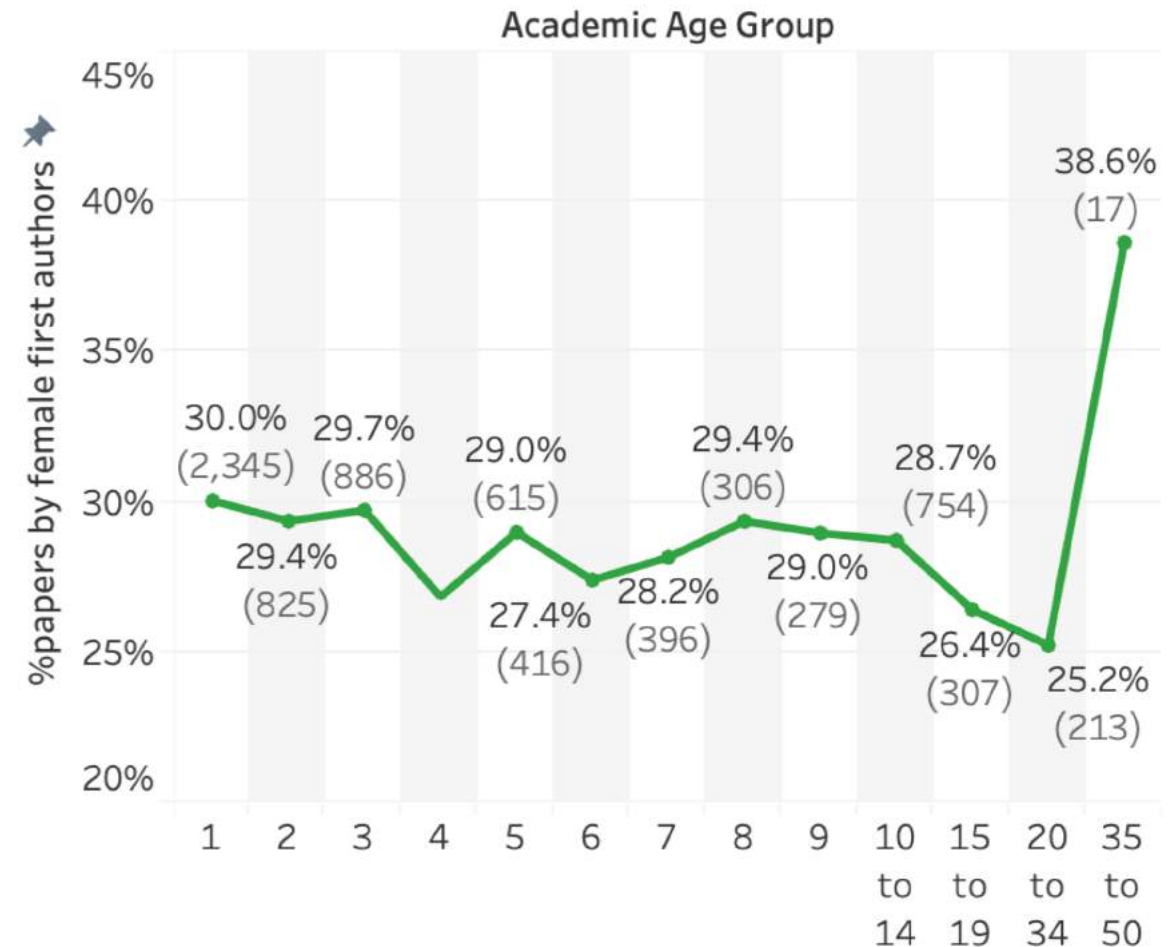
- lowest for CoNLL, EMNLP, IJCNLP, and system demo papers (21% to 24%)
- markedly higher for LREC, \*Sem, and RANLP (33% to 36%), as well as for workshops (31.7%)

### Q3. How does female first author percentage change with NLP academic age (experience)?

**NLP academic age:** number of years since first publication in AA + 1

Figure shows FFA percentage by academic age:

- the peak of year 1 is not surpassed for at least 33 years
- notable period of decline until age 6



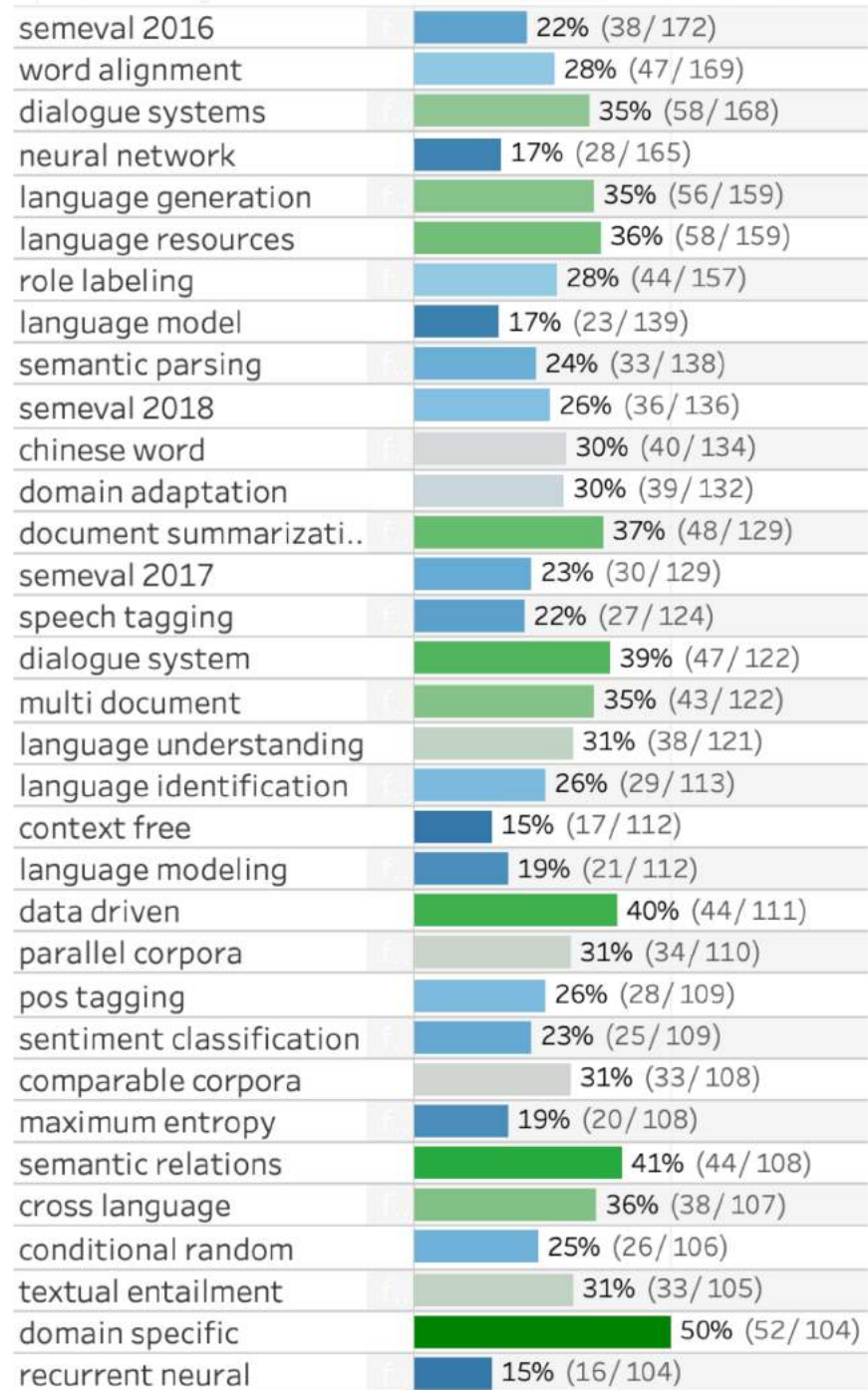
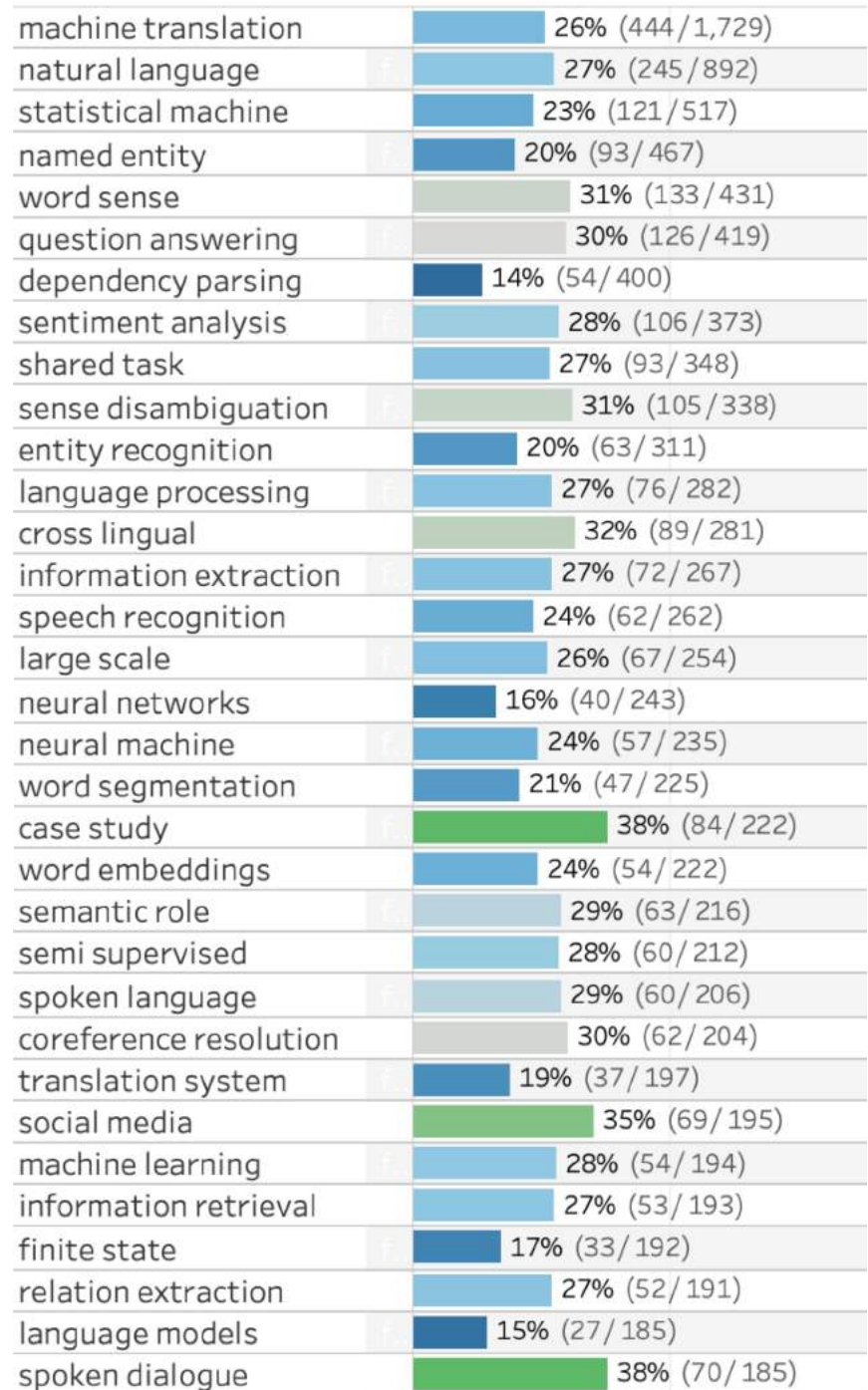
The number of FFA papers is shown in parenthesis.





## Q4. How does female first author percentage vary by area of research?

- disparities occur in most subareas
- popularity of research area and FFA%: no correlation



# Disparities in Citations





# Metrics of Research Impact (on subsequent scholarly work)

- Often derived from citations
  - number of citations, average citations, h-index, relative citation ratio, and impact factor (Bornmann and Daniel, 2009)
- However, citations do not always reflect quality or importance. Impacted by:
  - systematic biases
  - atypical contributions
  - popularity of area
  - unethical practices (e.g. egregious self citations)

Nonetheless, given the lack of other easily applicable and effective metrics, citation metrics used as an imperfect but useful window into research impact

- often a factor in funding research and hiring scientists

## Q5. *How well cited are women and men?*

Figure shows details by first author gender

Bottom: #papers

Middle: total citations

Top: box and whisker plots of citations of papers

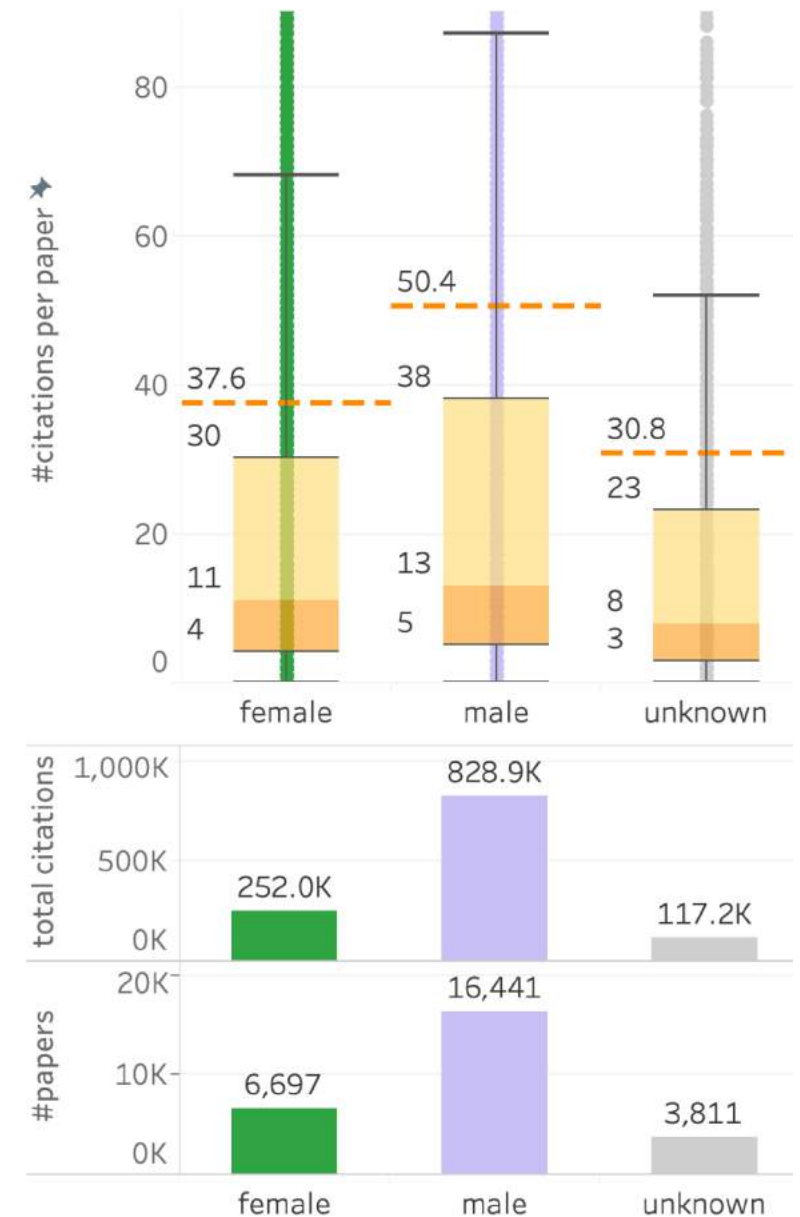
- shaded segments: quartiles on either side of the median
- upper and lower whiskers: 1.5 times the inter-quartile range (from the 25<sup>th</sup> to 75<sup>th</sup> percentile)
- orange dashed line: average citations

## Q5. How well cited are women and men?

Figure shows details by first author gender

- Average citations:
  - female first author papers: 37.6
  - male first author papers: 50.4
- Median citations:
  - female first author papers: 11
  - male first author papers: 13
- Quartile above the median: notable difference
  - more heavily cited male first-author papers

Differences are statistically significant (Kolmogorov–Smirnov,  $p < .01$ )





The differences in citations, or *citation gap*, across genders may itself vary:

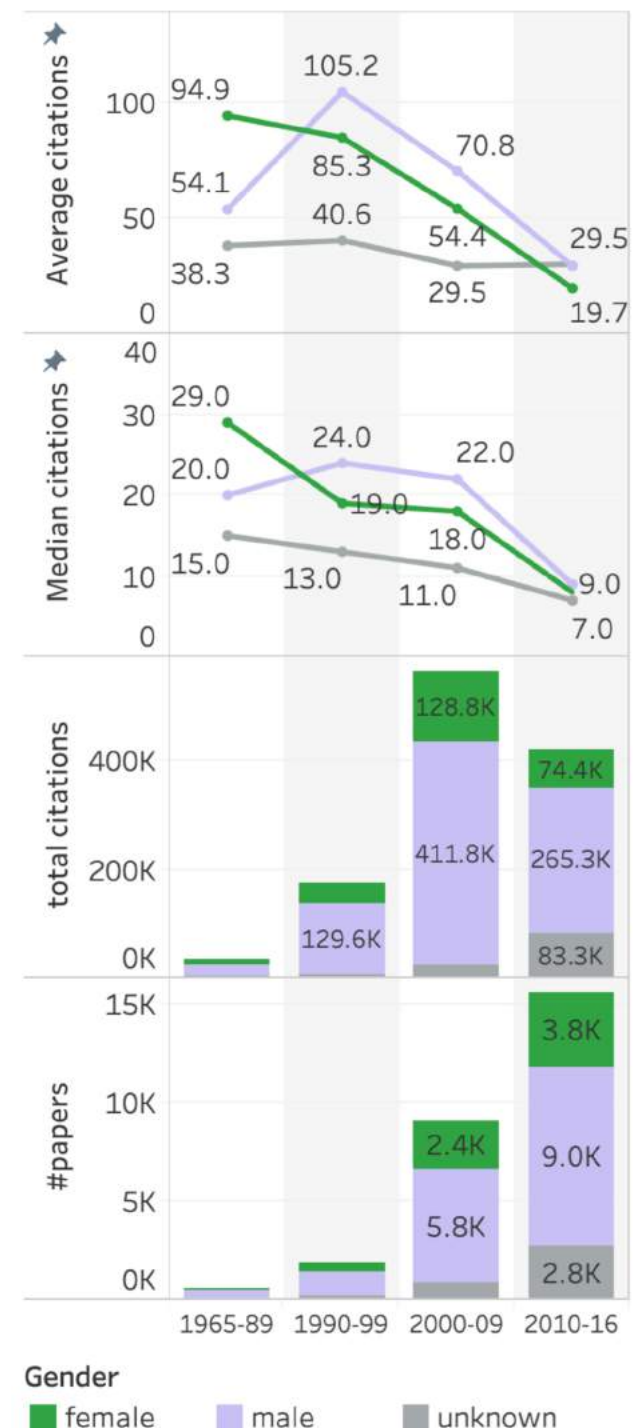
- by period of time (90s, 2000s, 2010s)
- due to confounding factors
  - academic age
  - areas of research

We explore these in the paper as well.

## Q6. How have citations varied for papers published in different time spans?

- Figure shows citation stats. for papers published in different time spans
- On average, female first author papers received:
  - more citations early on
  - fewer citations since the 1990s

Note: For the 2010s papers, the citation gap may widen as time goes on



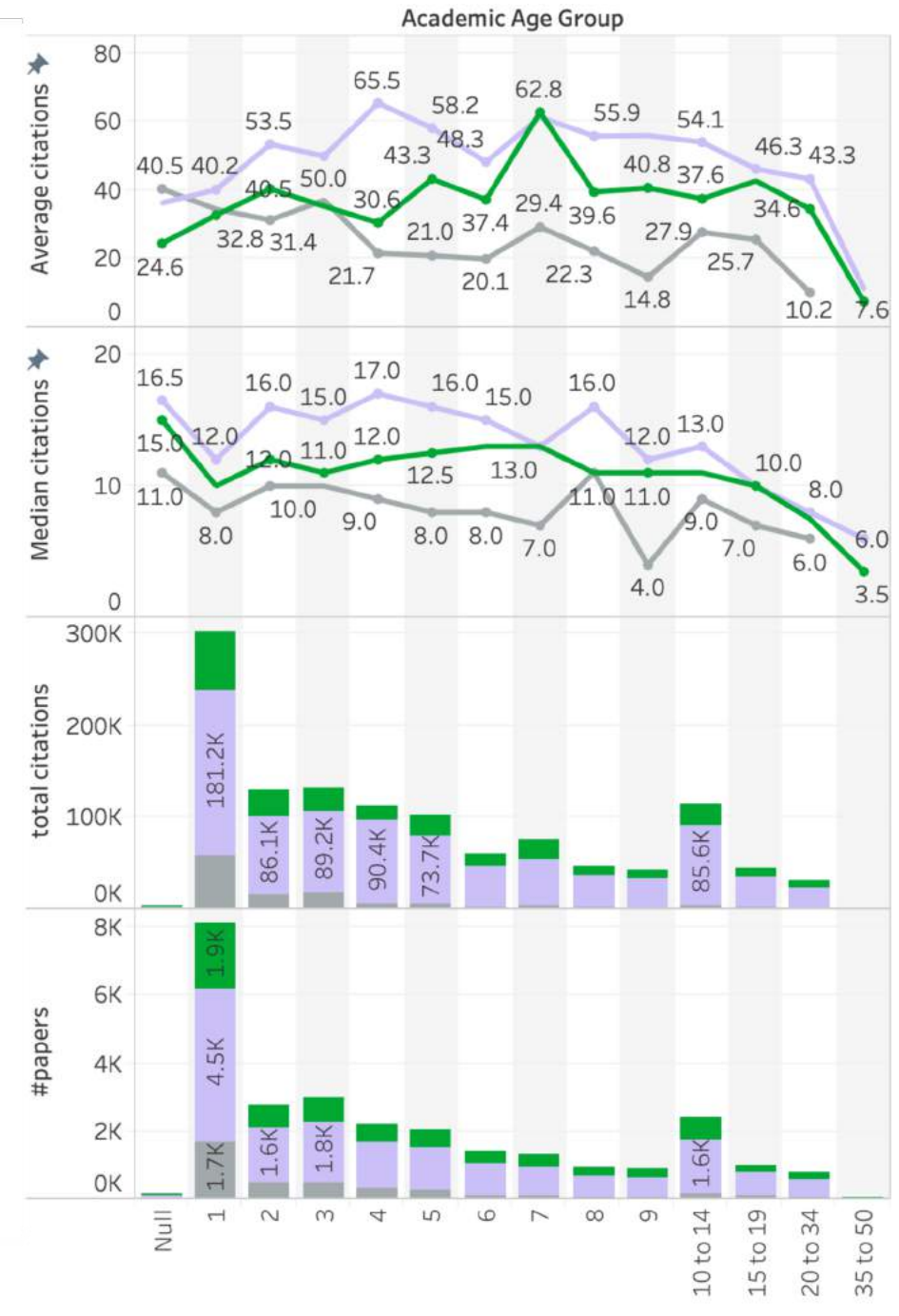


## Q7. How have citations varied by academic age?

Is the citation gap a side effect of a greater proportion of new-to-NLP female first authors than new-to-NLP male first authors?

- Figure shows citation stats. by academic age
- Female first authors consistently receive fewer citations than male first authors across the spans of their academic age

Thus, the citation gap is not simply because of a greater proportion of new-to-NLP female first authors.



## Q9. *What are the limitations and ethical considerations involved with this work?*

- Data is often a representation of people (Zook et al., 2017)
- Analysis focused on women and men leaves out non-binary people
- Excludes people and cultures that do not have gender-signalling names
- At the level of individuals: we do not want to perpetuate the idea that it is okay to infer gender from name (automatic gender detection is highly controversial)
- Self-reported data has its own challenges; and not possible for historical data
- Social category detection can lead to harm; however, one can also see the benefits:
  - public health: e.g., targeted initiatives to improve health outcomes of populations
  - social science: e.g., understanding the challenges of belonging to a social category

**NLP Scholar project webpage:** larger list of ethical considerations available through the :

<http://saifmohammad.com/WebPages/nlpscholar.html>

**Mihaljevic et al. (2019):** limitations and biases in using author names to infer gender statistics in the **Gender Gap in Science Project**: <https://gender-gap-in-science.org>

# Conclusions

In NLP, gender gaps exist both in authorship and citations:

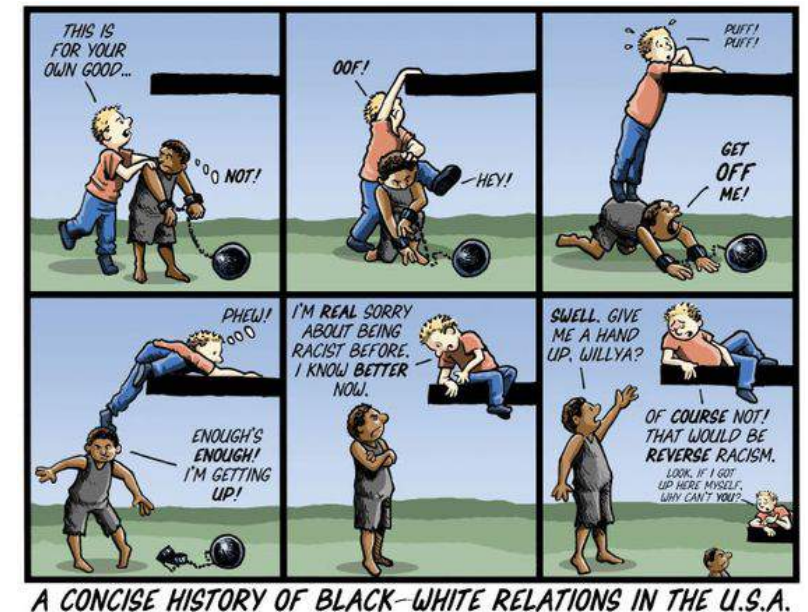
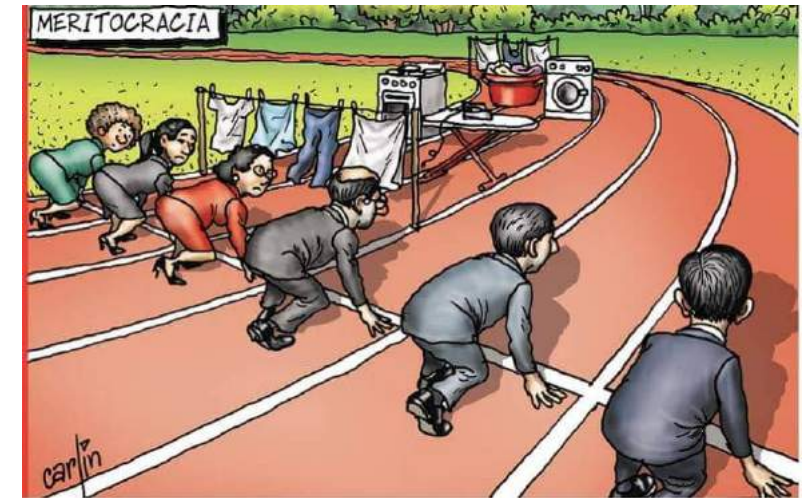
- ~29% have female first authors
- ~25% have female last authors
- female authorship percentages have not improved since the mid 2000s
- disparities occur in most subareas
- on average, male first authors are cited markedly more than female first authors
  - even when controlling for experience and area of work



# This paper did not explore...

## The reasons behind the gender gaps:

- complex, intersectional, difficult to disentangle
- sexism, inequities that impact women in scientific research (Roos, 2008; Foschi, 2004; Buchmann, 2009)
- biases that impact citation patterns unfairly (Brouns, 2007; Feller, 2004; Gupta et al., 2005)
- the myth of meritocracy (Noon, 2010)
- the social construct of gender (Schoellkopf, 2012)
  - distinct expectations and attitudes towards men and women



There are clear parallels between sexism and racism.

# This paper did not explore...

## How do we bridge the gender gap?

- no easy answers
- listen, amplify, join the voices/efforts of those affected; be an ally
- inclusiveness measures at conferences and research labs; Widening NLP; coding for girls...
- efforts will be met with feelings of unease and pushback
  - male fragility (Mitchell 2019, Diangelo2018): defensiveness of a person in discussions of sexism and patriarchy
  - can **seem** like special treatment of women (to both men and women)

the status quo is not acceptable.

📌 Pinned Tweet

Poet, Playwright, Activist



**Shailja Patel** ✓ @shailjapatel · Mar 8, 2018

Read women.  
Cite women.  
Credit women.

Teach women.  
Publish women.  
Present women.

Acknowledge women.  
Award women.  
Amplify women.

Hire women.  
Support women.  
Promote women.

Hear women.  
Believe women.  
Follow women.

Pay women.  
Pay women.  
Pay women.

#IWD2018

💬 1.5K

↻ 125.6K

❤️ 262.1K



## Future Work:

- A qualitative study on overcoming the hurdles to gender disparity:
  - interviewing women NLP researchers
  - led by expert in qualitative analysis, gender studies



Created by Symbolon  
from Noun Project

**Project page for NLP Scholar:** <http://saifmohammad.com/WebPages/nlpscholar.html>

- interactive visualizations
- limitations and ethical considerations



## In Separate Work:

Examining Citations of Natural Language Processing Literature. ACL 2020.

NLP Scholar: An Interactive Visual Explorer for Natural Language Processing Literature. ACL 2020 (Demo).

- A tool to find related work

## Contact:



Saif.Mohammad@nrc-cnrc.gc.ca



@SaifMMohammad



National Research  
Council Canada

Conseil national de  
recherches Canada

Canada