DE: Redirected from “female journalist” 
"(Weitergeleitet von Journalistin)”: 
The Gendered Presentation of Professions on Wikipedia

Authors:
Olga Zagovora
Fabian Flöck
Claudia Wagner
Gender stereotypes

Watch video from Project “Inspiring the future” from here
http://www.inspiringthefuture.org/redraw-the-balance/
Motivation & Objective

- What do readers see when they want to inform themselves about professions on Wikipedia?

Objective:
- Reveal how (im)balanced the gender presentation is on the profession pages of the German Wikipedia
- Test if any imbalance can be explained by underlying labor market data or other facts
Profession article. Example:

DE: female journalist

DE: male journalist

E: Male journalist

http://de.wikipedia.org/wiki/Journalistin

DE: female journalist

DE: Male journalist

DE: Redirected from female journalist
dehmale journalist
Profession article. Example:

DE: female journalist
DE: male journalist

Selbstverständnis der Journalisten


Im Unterschied zu vielen anderen Ländern hat man seit Bestehen der Bundesrepublik vermieden, die Journalisten aktiv in die jeweilige Regierungspolitik einzubinden, da die Gefahr einer wiederholten Instrumentalisierung der Presse als propagandistisches Erfüllungsorgan aus der NS-Propaganda befürchtet wird. Deutschland ist seitdem das einzige Land, dessen höchste Organisationsform der Journalisten, die Bundespressekonferenz, die Regierungssprecher zu den Pressekonferenzen einlädt.11[ Embedded Journalism, wie ihn die USA während des Irak-Krieges praktizierten, war in Deutschland bislang nicht vorgesehen.

In vielen Ländern werden Bild und Selbstverständnis der Journalisten durch zahlreiche Romane, Kurzgeschichten, Theaterstücke und Filme dokumentiert. So taucht in den USA etwa The Front Page, das 1928 uraufgeführte Standardwerk vor Ben Hecht und Charles MacArthur, in immer wieder neuen Adaptionen sowohl auf dem Broadway als
3 Analyses

1. **Redirection analysis** → Which professions/gender forms exist as articles and which are just redirects to the other gender or don’t have a lemma?

2. **Image analysis** -> What is the distribution of female and male persons depicted in the articles’ images?

3. **Textual analysis** -> What is the distribution of female and male persons mentioned by name in the articles?
1. Redirection analysis

- **Data: Seed list of professions** [based on profession list from “Federal Employment Agency”]:
  - "Lehrer": "Lehrerin", male teacher, female teacher
  - "Krankenpfleger": "Krankenschwester", male nurse, female nurse
  - "PR-Fachkraft", "Fotomodell" PR Specialist, photo model

  - Male-female pairs n = 4274
  - Neutral title n = 183
1. Redirection analysis

- (manually verified) matching of professions & gender forms to existing German WP article titles → **885 lemmata matched**

![Bar chart showing redirection analysis](chart.png)

- most articles have **male title** [94%]
- most redirects are from **female** to **male title** [96%]
2. Images analysis

Data: Images from profession articles
CrowdFlower task

How many persons are depicted in this image?
- Image is not shown
- No Person
- One Person
- Several Persons, “but one” person’s depiction is dominant
- Several Persons, “no single” person’s depiction is dominant

What is the gender of the person depicted?
- Female
- Male
- Gender is not recognizable

How many persons are depicted in this image?
- Image is not shown
- No Person
- One Person
- Several Persons, “but one” person’s depiction is dominant
- Several Persons, “no single” person’s depiction is dominant

What is the gender of the persons depicted?
- Only female
- Only male
- Mixed, but predominantly male persons
- Mixed, but predominantly female persons
- Mixed (equal amount of male and female persons)
- Gender is not recognizable

How many persons are depicted in this image?
- Image is not shown
- No Person
- One Person
- Several Persons, “but one” person’s depiction is dominant
- Several Persons, “no single” person’s depiction is dominant

What is the gender of the person depicted in a dominant way?
- Female
- Male
- Gender is not recognizable
Images analysis. Results

- 906 images from 885 Wikipedia articles
- 3 judges per photo -> response of majority [reliability of agreement $\kappa = 0.75$]
Do Wikipedia images reflect on labor market statistics?

Professions with female majority:
- Women in image (n=59): 38.4%
- Gender is not recognizable (n=8): 2.8%
- Men in image (n=111): 30.1%
- Mixed, equal amount of men and women in image (n=24): 8.3%
- No person (n=87): 20.4%

Professions with male majority:
- Women in image (n=52): 47.4%
- Gender is not recognizable (n=23): 3.7%
- Men in image (n=292): 36.9%
- Mixed, equal amount of men and women in image (n=22): 8.4%
- No person (n=227): 3.6%
3. Textual analysis

Data: Mentioned people from profession articles
- gender identification according to the first name (accuracy=0.97)
- 5085 (4272 men and 813 women) persons from 885 articles
- 411 articles with at least one person

Distribution of percentage of male names in an article

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.83</td>
</tr>
<tr>
<td>median</td>
<td>0.98</td>
</tr>
<tr>
<td>25%</td>
<td>0.8</td>
</tr>
<tr>
<td>75%</td>
<td>1.0</td>
</tr>
<tr>
<td>avg. number of persons</td>
<td>10.4 m 1.9 f</td>
</tr>
</tbody>
</table>
Gender equality/neutrality?

Hypothetical case

- In Wiki articles, the percentage of men is higher compared to the German labour market.
- In the neutral case, the distribution is more even, indicating gender equality or neutrality.

Olga Zagovora
The Gendered Presentation of Professions on Wikipedia, WebSci’17
## Relation to labor market statistics

<table>
<thead>
<tr>
<th>Mentioned people in an article</th>
<th>German labor market statistics</th>
<th>Correlation</th>
<th>The higher the percentage of / The more …</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature 1</td>
<td>Feature 2</td>
<td>Correlation</td>
<td>The higher the percentage of / The more …</td>
</tr>
<tr>
<td>percentage of mentioned men</td>
<td>percentage of men in the labor market</td>
<td>0.27***</td>
<td>mentioned men is in the article, the lower the percentage of women is in the profession</td>
</tr>
<tr>
<td>number of mentioned men</td>
<td>number of people in the labor market</td>
<td>-0.2***</td>
<td>men are mentioned in the article, the fewer people are employed in the profession</td>
</tr>
<tr>
<td>number of mentioned men</td>
<td>number of men in the labor market</td>
<td>-0.23***</td>
<td>men are mentioned in the article, the fewer men are employed in the profession</td>
</tr>
<tr>
<td>number of mentioned men</td>
<td>number of women in the labor market</td>
<td>-0.15**</td>
<td>men are mentioned in the article, the fewer women are employed in the profession</td>
</tr>
</tbody>
</table>
Time as confounder

- dbpedia -> birthDate
- divide people on those who were born before & after 1960

<table>
<thead>
<tr>
<th>born after 1960</th>
<th>percentage of men in the labor market</th>
<th>cor</th>
</tr>
</thead>
<tbody>
<tr>
<td>percentage of mentioned men</td>
<td>percentage of men in the labor market</td>
<td>0.23*</td>
</tr>
<tr>
<td># mentioned men</td>
<td># people in labor market</td>
<td>-0.12*</td>
</tr>
<tr>
<td># mentioned men</td>
<td># men in labor market</td>
<td>-0.12</td>
</tr>
<tr>
<td># mentioned men</td>
<td># women in labor market</td>
<td>-0.11</td>
</tr>
</tbody>
</table>
Summary

• Male bias over all dimensions:
  • Redirects
  • Images
  • Mentioned people

• High female bias for some professions
  • Examples: “Model” (mentioned people), “Midwife” (images)
Discussion & Outlook

Why does the male bias exist on Wikipedia?
• Implicit stereotypes of each individual
• Male bias in sources (Web, search engine, …)
• Historical confounder

What can be done to reduce it?
• Implementation of Wikipedia equality writing guidelines for profession articles
  • define gender equality, think about target reader groups
• Redirection re-design (applied to German & similar Wikipedias)
  • neutral profession titles as target lemmata
• Mentioned people & images
  • apply equality rule for other sections despite historical one

Future directions:
• Modify method in order to better control for historical confounder
• Cross-language analysis of gender inequalities for different Wiki editions
• Study actual effects on readers
Questions?

Me (Olga Zagovora)
olga.zagovora@gesis.org
zagovora@uni-koblenz.de

Fabian Flöck
fabian.floeck@gesis.org

Claudia Wagner
claudia.wagner@gesis.org
clwagner@uni-koblenz.de

You can access our paper from https://arxiv.org/abs/1706.03848