Are Women Present, Absent or in Disguise? Analyzing Gender Bias in the Spanish Wikipedia

J. Minguillón, J. Meneses, S. Fàbregues, E. Aibar, N. Ferran-Ferrer

*Universitat Oberta de Catalunya
Barcelona, Spain*

Wikipedia PROs:

- *"The sum of all knowledge"*
- Ranked 5th in Alexa
- 300 language versions

English (1st) vs Spanish:

- Articles: 5,7M / 1,5M (8th)
- Edits: 864M / 111M (4th)
- Editors: 35M / 5,2M (2nd)
- Active: 135K / 18K (4th)

CONs:

- Hostile editing environment
- Several known bias: racial, **gender**
- Content quality and diversity
- Wikipedia decline

Despite Wikimedia Foundation efforts:

- Only 12% - 16% of editors are women
- Failed to reach 25% goal
- Higher dropout rate among women

Project goals:

1) Analyze women presence in Spanish Wikipedia

2) Explain early dropout

3) In disguise? Why?

Are Women Present, Absent or in Disguise?
Analyzing Gender Bias in the Spanish Wikipedia

Sample:

- October the 1st, 2017 dump

- Registered users (no IPs, no bots) that have edited in the last 5 years with at least 50 edits and personal user page

- 13,210 candidates to be coded

- Stratified random sampling (4 x 4) by number of edits and user page size

Coding instrument:

- User page content only, no username

- Coded as men / women if:
  - Use of infoboxes
  - Explicit saying of gender of real name
  - Use of gendered language
  - "Unknown" otherwise

- Krippendorff's α = 0.9635 (2 c., 104 s.)

Automatic coding (API):

- User profile: Usuario: (+ default) / Usuaria:

Are Women Present, Absent or in Disguise?
Analyzing Gender Bias in the Spanish Wikipedia

Results: Manual x API

	Not coded	Men	Women	Total
Default	1,618	**1,122**	**172**	2,912
Men	**746**	**907**	0	1,653
Women	**59**	0	**123**	182
Total	2,423	2,029	295	4,747

Discussion:

- 4,747 user profiles coded (35.9%), few discrepancies manually solved:
  - 354 women (7.5%)
  - 2,775 men (58.4%)
  - 1,618 unknown (34.1%)

- Men are more likely to include their gender in their profile (59.6% vs 51.4%), but no differences in exposure through user page

- Our projection: **11.3% are women**

Are Women Present, Absent or in Disguise?
Analyzing Gender Bias in the Spanish Wikipedia

Further research:

- Gender differences in editing practices:
  - Content / interests
  - Activity (intensity)
  - Participation
  - Exposure through user page
  - Dropout

- Content analysis (16 interviews to current and former women editors)

- Extend / update manual coding

Further information

Project website
http://waww.blogs.uoc.edu/

Twitter: **@jminguillona**
email: **jminguillona(at)uoc(dot)edu**