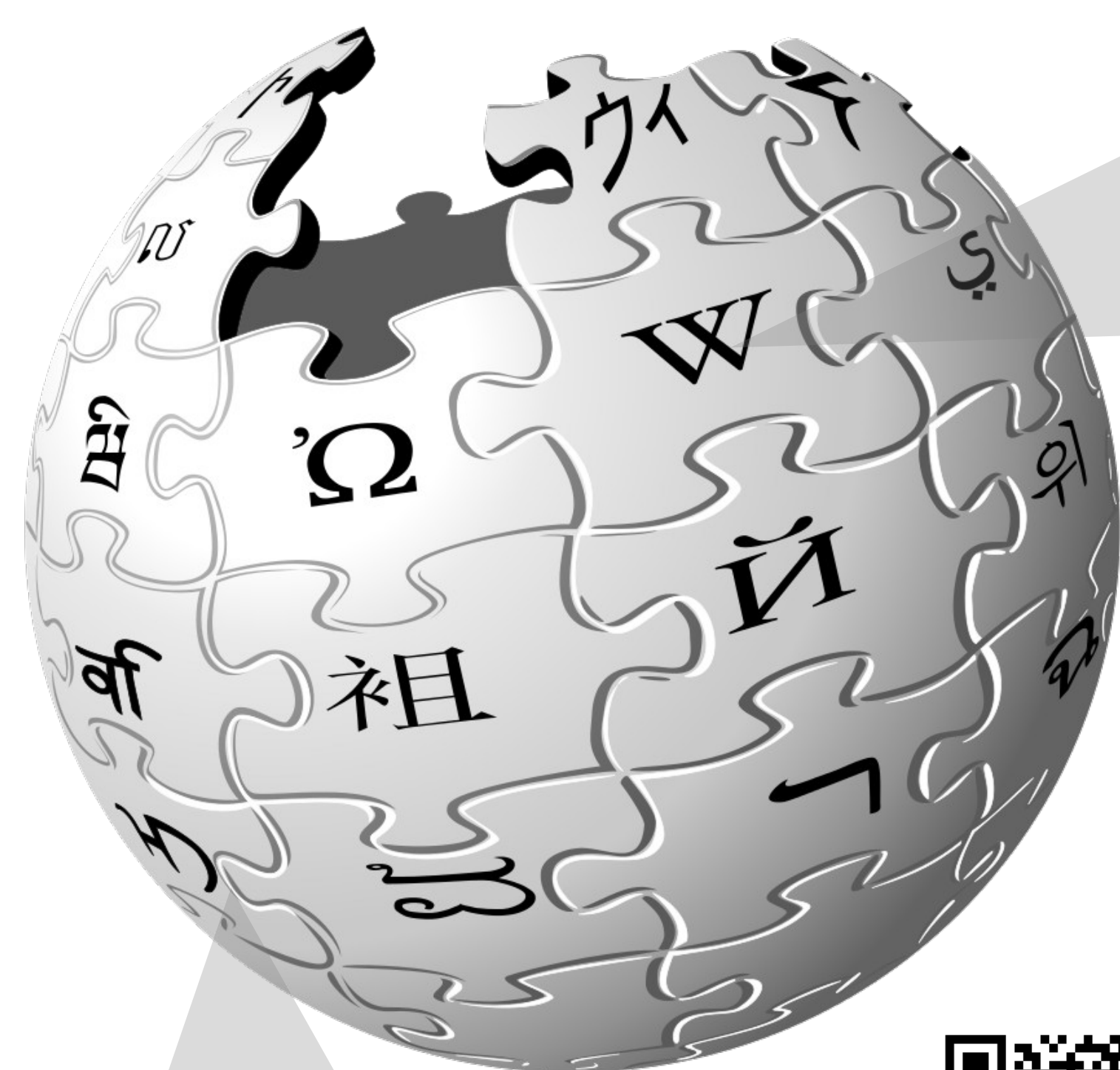


Are Women Present, Absent or in Disguise? Analyzing Gender Bias in the Spanish WP

J. Minguillón, J. Meneses, S. Fàbregues, E. Aibar, N. Ferran-Ferrer
Universitat Oberta de Catalunya, Barcelona, Spain



Wikipedia

"The sum of all knowledge"

5th site in Alexa ranking

300 different languages

Nov. 2018	English (1 st)	Spanish
Articles	5,7 M	1,5 M (8 th)
Edits	864 M	111 M (4 th)
Editors	35 M	5,2 M (2 nd)
Active	135 K	18 K (4 th)



Only 12-16% of editors are known to be females

Failed to reach 25% goal

Higher dropout rate

But...

Gender bias

Hostile editing Environment

Content quality and diversity

Wikipedia decline

Project goals

- 1) Analyze women presence in Spanish Wikipedia
- 2) Early dropout reasons
- 3) In disguise? Why?

Sample

October the 1st, 2017 dump

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

13,210 candidates to be coded

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

Instrument (manual)

User page content, but no username

Use of infoboxes

Explicit saying of gender or real name

Use of gendered language

+

API (automatic)

User profile

(default)
Usuario:
Usuaría:

	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

Further research

Gender differences in editing practices, participation, interests, dropout, ...

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

Extend / update manual coding

Discussion

4,747 coded, few discrepancies manually solved
Krippendorff's $\alpha = 0.9635$ (2 coders, 104 r. s.)

Men are more likely to include their gender in profile (59.6% vs 51.4%)

Only 354 (7.5%) can be identified as **women**, 2,775 (58.4%) **men**, 34.1% remain unknown

Our projection: **11.3% are women**