COVER WOMEN: A MULTILINGUAL WIKIPEDIA MAIN PAGE COMPARATIVE PROPOSAL OF GENDER AND INTERSECTIONALITIES ON CONTENT, NEWSROOM GUIDELINES, AND INSIGHTS FROM THE COMMUNITY OF VOLUNTEERS

Núria Ferran-Ferrer
Universitat de Barcelona

Laura Fernández
Universitat de Barcelona

Miquel Centelles
Universitat de Barcelona

Abstract

This proposal presents a research project proposal that will look into the Wikipedia main page. The main page or front page from a communication perspective, will be analysed across the seven longest-standing Wikipedia editions. Grounded in a gender and intersectional perspective, the examination will employ communication theories like gatekeeping and agenda-setting. Beyond academic research, our goal is to actively contribute to editing communities by addressing the daily challenges and needs in crafting front-page content.

Keywords: main page, gate keeping, agenda setting, cultural diversity, gender gap, intersectional analysis.

Introduction

Despite Wikipedia being a key player in the public sphere and having a transformative impact on information dissemination, Wikipedia grapples with persistent gender bias in both editing and content (Antin et al., 2011; Bear and Collier, 2016; Wagner et al., 2016; Hinnosaar, 2019; Minguillón et al., 2021; Ferran-Ferrer, Boté-Vericad, et al., 2023) alongside additional prejudices (Redi et al., 2021; Beytía et al., 2022). Bias in contributions perpetuates imbalances in content coverage and discourages diversity, which further exacerbates the issue (Worku et al., 2020).

Scholars highlight the need for a comprehensive understanding of Wikipedia's knowledge production culture to address these biases and make Wikipedia more robust, reliable, and transparent (Menking and Erickson, 2015). Reducing the gender and other intersectional biases necessitates more than acknowledging Wikipedia as a mirror of societal biases—it involves addressing the platform's deeper logic embedded in its techno-scientific project (Ford and Wajcman, 2017).

We have selected the most popular Wikipedia page for analysis. This page, commonly referred to as the main page, or front page from a communication perspective, is accessible in all language editions of the global encyclopedia, and we will conduct our study on it. We will research into the possible gender and intersectional bias in its daily content, in its newsroom guidelines (principles and standards that govern the dissemination of information), and in the insights from the volunteer community who decide which information gets disseminated to the public on the main page. This research will utilize communication theories such as gatekeeping which examines the process by which information is filtered, selected, and ultimately presented to the public (Barzilai‑Nahon, 2009) and agenda-setting (McCombs and Shaw, 1972) which studies the effect.

Objectives and research questions

The primary objective of this research is to comprehensively examine the gender gap and intersectional biases, such as professions, ethnicity, nationality or mother tongue, within Wikipedia biographies, focusing on the daily front pages of its English, and Spanish language editions spanning a decade (2013-2023). Wikipedia's "main page" functions as the homepage, serving as the entry point with features akin to a front page, offering sections like "Today's Featured Article" and "In The News," essentially acting as the digital encyclopedia's daily cover. Through a comparative lens, this research seeks to provide insights into systemic challenges and practices. The ultimate goal is to advocate for positive changes that promote a more equitable and diverse online information environment.

The research questions that we address are:

- Research Question 1 (RQ1): What insights do interviews with volunteer gatekeepers (editors of the main page of Wikipedia) provide on decision-making, biases, and strategies affecting the visibility of gender and intersectionality-related content on Wikipedia's front page, particularly regarding how their preferences and interests, shape the topics featured?

- RQ2: How does gatekeeping impact gender gaps in content representation on digital platforms, specifically in the peer production of knowledge (decision-making system on suitable content and what is not) within newsrooms or editorial policies, and why is understanding this phenomenon crucial for addressing gender disparities?

- RQ3: How does agenda setting influence the selection of frames and sentiment adopted by Wikipedia pages concerning specific issues or events, and how does it shape the focus and intensity of user edit activity within Wikipedia?

- RQ4: How prevalent is gender and intersectional bias in the content featured on Wikipedia's front pages?

This research is necessary to draw further attention to the need for systemic change within the platform's newsroom/editorial practices to address disparities in gender and diversity representation in online knowledge and foster a more inclusive and diverse digital information landscape.

Methods

We will employ a triangulation methodology.

1. **Insights from community of volunteers**: We will conduct in-depth interviews with voluntary editors of the front page from all seven Wikipedia editions to ascertain decision-making processes, biases, and strategies that influence content visibility related to gender and other intersectionalities.
2. **Newsroom guidelines:** We will apply content analysis to main-page, or front-page editorial guidelines, for each language edition, and we will explore what leads the decision-making of the gatekeepers who determine story prominence.

3. **Main-page content analysis:** We will scrutinize the daily content selection (biographies) on the front-page in each language edition for ten years, with data wrangling using OpenRefine, enhancing data with specific Wikidata properties like P21 (sex or gender), P106 (occupation), P172 (ethnic group), and others.

Community impact plan

The project aligns with the Wikimedia Movement's 2030 strategy by focusing on delivering knowledge as a service and addressing equity in knowledge and communities overlooked by structures of power and privilege.

Furthermore, this project proposal has been designed according to the needs of various activist groups in terms of gender equality on Wikipedia, as well as with the boards of the chapters involved in each language edition. See table 1 to anticipate the impact on communities we will reach.

This project will provide:

a) Decade-long insights into gender and intersectionality content representation on Wikipedias' front pages. (RQ4)
b) Beyond descriptive stats, we'll reveal bias trends. (RQ4)
c) Editorial strategies for gatekeeping and agenda setting. (RQ1-3)
d) Guidelines for ethical content selection using AI and bots. (RQ3)
e) Technical guidance to enhance data archives on main pages. (RQ4)
f) Collaborative work with volunteers ensures inclusivity, integrating advocate perspectives for a consensus-driven approach.(RQ1)

Built on in-depth interviews and collaboration with Wikipedia chapters and user groups, this work addresses gender identity under-representation, as reflected in the Cover Women team's table 1.

Related work

To contrast the feasibility of this proposal with seven language editions of Wikipedia, we have already conducted a micro project with the sample of the English and Spanish Wikipedia to assess the viability of the global project. That is, if there are open and formalized recommendations and guidelines that determine which contents are published on the main page and if the publication criteria can be analysed. At the same time, we were interested in seeing if with data wrangling techniques we could work with the biographies

---

published on all Wikipedia main pages and analyse them from a gender and intersectional perspective using the properties of Wikidata.Results of this work previous work will be published soon (Ferran-Ferrer, Fernández, Centelles, pending publication) but we can advance that only 28% of people that appeared in the maing-page were women, in the English edition, and it was even lower in the Spanish edition at 18%, with other marginalized gender identities below 1%. And the trend is not encouraging (Ferran-Ferrer, et al., 2024) if we take into account that bias in contributions perpetuates imbalances in content coverage and discourages diversity, which further exacerbates the issue (Worku et al., 2020).

To address this, scholars stress the importance of understanding Wikipedia's knowledge production culture to tackle its gender gap (Menking & Erickson, 2015). Addressing this issue involves grappling with the underlying logics of the platform's techno-scientific project (Ford & Wajcman, 2017), necessitating the recognition and dismantling of exclusionary practices (Menking & Rosenberg, 2021).

Communication theories like gatekeeping and agenda setting provide valuable frameworks for understanding Wikipedia's potential biases. Gatekeeping theory, focusing on information filtering processes, is applied to scrutinize stories selected for the Front Page, which attracts millions of readers monthly (Barzilai-Nahon, 2009; Wikimedia, 2023b). Additionally, drawing from agenda-setting theory, we examine how Wikipedia's main page influences viewers and shapes news hierarchy, including its agenda-building power (McCombs & Shaw, 1972; Ren & Xu, 2023).

This study goes beyond affirming Wikipedia's reflection of reality to delve into its systemic challenges (Ford & Wajcman, 2017). It analyzes not only main page content selection but also newsroom guidelines, including interviews with gatekeepers, to enhance understanding and address systemic issues.

References

Barzilai‑Nahon, K. (2009). Gatekeeping: a critical review. Annual Review of Information Science and Technology, 43(1), 1–79. https://doi.org/10.1002/aris.2009.1440430117

Bear, J. B., & Collier, B. (2016). Where are the Women in Wikipedia? Understanding the Different Psychological Experiences of Men and Women in Wikipedia. Sex Roles, 74(5), 254–265. https://doi.org/10.1007/s11199-015-0573-y

Beytía, P., Agarwal, P., Redi, M., & Singh, V. K. (2022). Visual gender biases in wikipedia: a systematic evaluation across the ten most spoken languages. Proceedings of the International AAAI Conference on Web and Social Media, 16, 43–54. https://ojs.aaai.org/index.php/ICWSM/article/view/19271

Ferran-Ferrer, N., Fernández, L., & Centelles, M. (2024). Wikipedia's Front page 10 years evolution: analysis of gender and intersectionalities on content, newsroom guidelines and volunteer-driven insights in the Spanish and English editions. [Forthcoming].

Ford, H., & Wajcman, J. (2017). Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. Social Studies of Science, 47(4), 511–527. https://doi.org/10.1177/0306312717692172

Hinnosaar, M. (2019). Gender inequality in new media: Evidence from Wikipedia. Journal of Economic Behavior & Organization, 163, 262–276. https://doi.org/10.1016/j.jebo.2019.04.020

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. Public Opinion Quarterly, 36(2), 176–187.

Menking, A., & Rosenberg, J. (2021). WP:NOT, WP:NPOV, and Other Stories Wikipedia Tells Us: A Feminist Critique of Wikipedia's Epistemology. Science Technology and Human Values, 46(3), 455–479. Scopus. https://doi.org/10.1177/0162243920924783

Minguillón, J., Meneses, J., Aibar, E., Ferran-Ferrer, N., & Fàbregues, S. (2021). Exploring the gender gap in the Spanish Wikipedia: differences in engagement and editing practices. PLOS ONE, 16(2), e0246702. https://doi.org/10.1371/journal.pone.0246702

Redi, M., Gerlach, M., Johnson, I., Morgan, J., & Zia, L. (2021). A taxonomy of knowledge gaps for Wikimedia projects (Second Draft) (arXiv:2008.12314). arXiv. http://arxiv.org/abs/2008.12314

Wagner, C., Graells-Garrido, E., Garcia, D., & Menczer, F. (2016). Women through the glass ceiling: Gender asymmetries in Wikipedia. EPJ Data Science, 5(1), Article 1. https://doi.org/10.1140/epjds/s13688-016-0066-4

Wikimedia. (2023). Wikimedia statistics: all wikis. https://stats.wikimedia.org/#/all-projects

Worku, Z., Bipat, T., McDonald, D. W., & Zachry, M. (2020). Exploring Systematic Bias through Article Deletions on Wikipedia from a Behavioral Perspective. ACM International Conference Proceeding Series. Scopus. https://doi.org/10.1145/3412569.3412573

Appendix

Table 1. Research project overview

Research Questions	Methodology	Methods and techniques	Outputs
RQ1-RQ4	Systematic bibliographic review/Scoping review	SALSA Method	*Open Data set * Theoretical background for the project
RQ1&3	Quantitative and qualitative approaches	In-depth interviews with voluntary editors of the front page from all seven Wikipedia editions Content codification of the interview transcriptions and analysis with a qualitative data analysis software, and a specific codebook. Collaborative work with volunteers ensures inclusivity, integrating advocate perspectives for a consensus-driven approach	*Academic and conference papers *Editorial strategies guidelines for gatekeeping and agenda setting * Guidelines for ethical content selection using AI and bots.
RQ2	Quantitative and qualitative approaches	Preparation of the corpus of guidelines to codify for each language edition. Content analysis to front-page editorial policies (main page editing guides) through a code book designed ad hoc and qualitative data analysis software. Collaborative work with volunteers ensures inclusivity, integrating advocate perspectives for a consensus-driven approach	*Academic and conference papers *Editorial strategies guidelines for gatekeeping and agenda setting.
RQ4	Main-page content quantitative analysis.	Data from the seven language edition main pages will be wrangled.	*Open Data set *Academic and conference papers *Decade-long and bias trend report into gender and intersectionality content * Technical guidance to enhance data archives

			on main pages.

Table 2. Contributors and Collaborators Across Seven Language Editions of Wikipedia

Name	Affiliation	Country	Rol in the proposal Cover Women	Wikipedia language edition
Núria Ferran-Ferrer	* Associate Professor at UB Universitat de Barcelona *Director of EDI Office UB *Director of Information and Communication PhD Programme, UB *Board Member of Amical Wikimedia *Member Commitee WMF Grants in NW Europe	Catalonia, Spain	Co-IP	All
Miquel Centelles	* Lecturer at UB *Coordinator of Master's Degree in Digital Humanities at Universitat de Barcelona *Research collaborator at Women and Wikipedia group.	Catalonia, Spain	Co-IP	All
Laura Fernández	*Postdoctoral researcher at Universitat de Barcelona *Researcher at Women and Wikipedia research group.	Catalonia, Spain	Co-IP	All
Juan-José Boté	Lecturer professor Information Science and Media Studies	Spain	Researcher	All
Julià Minguillón	Associate professor at Computer Science, Multimedia and Telecommunication Studies Universitat Oberta de Catalunya	Spain	Researcher	All
Florencia Claes	*Lecturer at Rey Juan Carlos University. *Head of Free Culture at Rey Juan Carlos University. *President of Wikimedia España	Spain	Advisor Spanish Wikipedia	Spanish
Ester Bonet	Member WikimujeresUG	Spain	Advisor from WikiMujeres	Spanish/ Catalan

Núria Ribas	President of the Catalan Wikimedia Board	Spain	Advisor Catalan Wikipedia	Catalan
Michael David MILLER	Associate Librarian at McGill University Vice-President of Wikimedia Canada Board Member of WikiFranca	Québec-Canada	Advisor French Wikipedia	French
Natacha Rault	Director and founder of les sans pagEs	Switzerland	Advisor from Les Sans Pages	French
Daniela Brugger	Coordinator of who writes his_tory?	Germany	infodanielabrugger.ch	German
Lilli Iliev	Head of politics & public sector, Wikimedia Deutschland e. V.	Germany		German
Laurie Bridges	Founding member and leadership team for #EveryBookItsReader. Core Committee for WikiLibCon 2022. Librarian at Oregon State University, USA	United States	Advisor English Wikipedia	English
Rosie Stephenson-Goodknight	Wikimedia Foundation Board of Trustees Visiting Scholar, Northeastern University (Boston) Co-founder, Women in Red		Women in Red	English
Lilian	*Librarian at Universidade de São Paulo*Member of GLAM Bibliotecas da USP	Brazil	Advisor Portuguese Wikipedia	Portuguese
Ana Bragança	Co-founder and co-coordinator, Wiki Editoras Lx	Portugal	Advisor Wiki Editoras Lx	Portuguese
Irene	Long-time editor and has been an administrator in itwik	Italy	Advisor Italian Wikipedia	Italian
Camelia Boban	WikiDonne project in Italian Wikipedia founder WikiDonne User Group co-founder	Rome, Italy	Advisor WikiDonne	Italian