

5

Gender Gap in Wikipedia Editing

A Cross Language Comparison

Paolo Massa and Asta Zelenkauskaitė

INTRODUCTION

According to various surveys, the percentage of women editing Wikipedia barely reaches 10 percent.^{1,2} The issue of gender distribution on Wikipedia was first brought to public attention by an article in the *New York Times* on January 2011. The article, titled “Define Gender Gap? Look Up Wikipedia’s Contributor List,”³ started by highlighting how, in just ten years, the Wikipedia community accomplished some remarkable goals, such as reaching more than 3.5 million articles in English and starting an online encyclopedia in more than 250 languages. Yet Wikipedia failed to reach at least a minimal gender balance: according to the United Nations and Maastricht University 2010 reported survey, less than 13 percent of contributors were female. A more recent survey carried out in 2011 by the Wikimedia Foundation, the nonprofit organization that coordinates the various Wikipedia projects,⁴ reported even a steeper gender gap: women account for just 9 percent of editors.

The importance of Wikipedia editors’ diversity is relevant since Wikipedia is increasingly becoming one of the most accessed Web sources for information needs. Some 53 percent of American Internet users searched for information on Wikipedia as of May 2010; 88 percent of 2,318 university students used Wikipedia during a course-related research process, and finally, Wikipedia is the sixth most visited site on the entire Web.⁵ Thus, because so many people read the content of Wikipedia pages, it is important to become aware that these pages reflect the point of view of a predominantly male population. This issue has also been addressed by the Wikimedia Foundation, and in fact they have devoted attention to gender balance on Wikipedia as one of the key points of their five-year strategic plan of 2010: “Doubling the percentage of female editors to 25 percent by 2015.”⁶

In light of this pronounced gender gap, since 2011 a number of studies have tried to understand the reasons and the extent of this phenomenon in Wikipedia.^{7, 8, 9, 10} Yet the challenge of comprehending and improving the situation is still lingering and unsolved.

This study focuses on the extent of the gender gap across various language editions of Wikipedia. Unlike the majority of studies on Wikipedia that predominantly focused on the English Wikipedia, this study uses a cross-cultural and crosslinguistic approach to investigate a mosaic of Wikipedias and sociolinguistic cultural practices that developed over the years, given that as of 2013 there are more than 280 Wikipedia language editions. Yet in this study we consider the English Wikipedia as a reference point, given that it is the oldest, the largest, and the most researched one.

The motivating questions are hence: Is the gender gap in other language editions of Wikipedia as pronounced as it is in the English one? Are there Wikipedia language editions with a narrower or wider gender gap?

The goal of this study is to make a contribution to a better understanding of cross-cultural and crosslinguistic varieties of Wikipedia as continuing the seminal research in this area.^{11, 12} Wikipedia provides us a unique context to study the sociocultural differences across language editions, given that they all share a constant element—the MediaWiki open source platform, which serves as a sort of controlled variable. Therefore, the differences found among the different editions of Wikipedia should be due to the different communities of practice inhabiting them and their cultural and historical differences. From this perspective, Wikipedia exhibits advantages to analyze differences among communities. Compared to an analysis of user behaviors across different platforms such as Facebook and Twitter, it keeps the platform as a constant variable.

This study is guided by two research questions:

RQ1: What is the percentage of users who set their gender in different language editions of Wikipedia?

RQ2: Among those who express gender, what percentages comprise female and male contributors?

Our analytic approach is based on data about gender explicitly entered by registered users of Wikipedia. In fact, in the MediaWiki software since 2010 it is possible to express gender in the preferences panel. Yet gender in English Wikipedia is referred to as an optional field, and the template notifies that it is “used for gender-corrected addressing by the software. This information will be public.”

METHOD

Wikipedia, as a sociotechnical system,¹³ is based on a unified technological platform—MediaWiki—that serves as a basis for all languages editions. Moreover, all

Wikipedias share the same set of established rules and core policies such as the Neutral Point of View.¹⁴ However, each platform in its language represents a singular, independent case, characterized by specific sociotechnical variables: cultural settings, year of creation, number of registered and active users, and customization of the settings of the platform.

To answer the research questions of this study, we created an automated script querying the replicated databases of Wikipedia that are made available on Toolserver (<http://www.toolserver.org>) and released it as open source.¹⁵ For every Wikipedia language edition, the script computes the number of registered users, the number of users who expressed their gender, and the number of users who indicated male or female. The percentage of users expressing their gender and the percentage of males and females over those who expressed their gender is computed as well.

We considered only users who registered after January 1, 2010, by setting a parameter of the script. The reason behind this choice is that the gender-setting functionality—the core phenomenon that we are addressing—had been introduced in most Wikipedias in early 2010.

The results refer to the situation as of March 16, 2013. The script produced a list of 289 language editions of Wikipedia. However, many of them had very few users, and in order to conduct reasonable quantitative investigations based on the number of users who expressed their gender, we arbitrarily decided to consider only editions with at least twenty thousand registered users, and this filtering step reduced our sample to seventy-six Wikipedias.

In this contribution we consider the English Wikipedia as a baseline because it was the first to start, and it has the largest number of users, pages, and edits.¹⁶ Moreover, according to the editor survey carried out by the Wikimedia Foundation in 2011,¹⁷

An overwhelming majority of Wikipedia editors read and edit English Wikipedia—which also has the largest and most diverse pool of editors, as editors from other projects contribute regularly. In total, 76% of Wikipedia editors contribute to English Wikipedia, although only 40% primarily contribute to English Wikipedia. In other words, in addition to the 40% of editors who primarily edit English Wikipedia, 36% of editors from other language projects contribute to English Wikipedia. An impressive 93% of Wikipedia editors read English Wikipedia, and about half of them (49%) primarily read English Wikipedia. We can clearly see that editors who work mainly in other language projects help English Wikipedia grow.

USERS SPECIFYING THEIR GENDER IN DIFFERENT LANGUAGE EDITIONS OF WIKIPEDIA

The first research question refers to the percentage of users who set their gender in different language editions of Wikipedia. Table 5.1 shows the language editions that have a larger percentage of users expressing their gender.

Table 5.1. Wikipedias with the Largest Percentages of Users Setting Their Gender

<i>Language (Wikipedia web address)</i>	<i>Percentage of users setting their gender</i>	<i>Number of users who set their gender</i>	<i>Number of users</i>
Russian (ru.wikipedia.org)	22.58%	139,038	615,690
Vietnamese (vi.wikipedia.org)	10.54%	18,090	171,568
Ukrainian (uk.wikipedia.org)	9.43%	10,969	116,378
Thai (th.wikipedia.org)	9.12%	7,869	86,321
Portuguese (pt.wikipedia.org)	8.72%	39,499	453,229
Turkish (tr.wikipedia.org)	8.40%	17,370	206,750
Arabic (ar.wikipedia.org)	8.16%	29,143	357,296
Marathi (mr.wikipedia.org)	7.53%	1,820	24,163
Japanese (ja.wikipedia.org)	7.49%	25,428	339,637
Chinese (zh.wikipedia.org)	7.14%	44,796	627,058
Farsi (fa.wikipedia.org)	6.59%	11,917	180,908
Tamil (ta.wikipedia.org)	6.43%	2,258	35,105
Indonesian (id.wikipedia.org)	6.43%	19,319	300,673
....			
English (en.wikipedia.org)	4.85%	353,056	7,283,226

At first glance, the overall small percentage in relation to gender specification stand out. Russian Wikipedia (22.58 percent) in table 5.1 is an outlier, compared to the rest of the language editions. The second in the ranking in table 5.1, the Vietnamese Wikipedia, drops the rate of gender identification to 10.54 percent, maintaining the decreasing pattern in gender self-identification throughout the rest of the table 5.1. These results are contrastive, compared to social networking sites such as Facebook on which almost everybody (93.8 percent according to Lampe et al.¹⁸ and 97 percent according to Stecher et al.¹⁹) set the gender. On the other hand, in Wikipedia this practice is less common. On the English Wikipedia, just 4.84 percent of users who registered since 2010 expressed their gender, precisely 353,056 out of 7,283,226.

One possible reason behind a small percentage of users expressing their gender when compared to, for example, Facebook is that the link to the page for changing the gender is not very visible in the interface. The user should first click on “Preferences” and then set their gender on this page.

More importantly there are no clear and visible effects or benefits of setting your gender on Wikipedia. In the interface, it is written that the gender information is optional, and it is “used for gender-correct addressing by the software.” Note also the additional text, “this information will be public,” has been present since 2010 in the interface (at least in the English Wikipedia). It is not precise since it is unclear when this information will become public, and it is actually incorrect since there are, and have been, different ways to obtain gender-setting information; for example, using the official Wikipedia API, as the WikiTrip web tool does.²⁰ Thus, there are no clear benefits of setting your gender on Wikipedia, and, contrary to social networking

sites such as Facebook, the social aspect of Wikipedia is not predominant at all. For example, the policy page “What Wikipedia is not” clearly states: “Wikipedia is not a social networking service like Facebook or Twitter. [. . .] Wikipedians have their own user pages, but they should be used primarily to present information relevant to working on the encyclopedia. Limited biographical information is allowed, but user pages should not function as personal webpages.”²¹

The question that remains unanswered is why on some Wikipedias there are larger percentages of users expressing their gender. The reasons are many, but the main ones have to do with the interface. Some languages (due to grammatical rules) highlight gender, referring to users as “she-user” or “he-user.” This is, for example, the case in the Russian Wikipedia, which also happens to be the first in table 5.1 with 22.58 percent of users expressing their gender. Moreover, in some of the Wikipedias in table 5.1, the user page has different addresses based on the self-declared gender; for example, in the Portuguese Wikipedia (fifth in table 5.1), male users have their page prefixed by *Usuário* while female users by *Usuária*. In this way, it is possible to hypothesize that users who see these differences become aware of this option and have more incentives in setting their gender.

Interestingly enough, the blog post “Nine Reasons Women Don’t Edit Wikipedia (in their own words)” by Wikimedia director Sue Gardner reports that “some women whose primary language has grammatical gender find being addressed by Wikipedia as male off-putting.” From a female Portuguese Wikipedian: “I have no problem with the male ‘*Usuário*’ (in portuguese). And sincerely, I don’t think the fact of see a male word will push me out Wikippedia. We are quite used to use a male word in portuguese when we don’t know the gender of someone, but yes, would be nice to see a ‘*Usuária*’ in my page :D.”²²

On the other hand, in English there isn’t a male and female version of the term User, adjectives referred to the active user do not differ based on gender, and “there’s only one message in English which is gender-customized.”²³

From the interface point of view, the platform is customizable by authorized users, specifically administrators of a given Wikipedia language edition. For most of the Wikipedias in table 5.1, for which a larger portion of registered users expressed their gender, we note that the “welcome message” interface page they receive after they register is very simple (see figure 5.2). It merely contains an invitation and a link to change their preferences and leads to the preference page in which they can also set their gender. On the other hand, the welcome page on the English Wikipedia (figure 5.1) does not present a link to the preferences page but is mainly focused on suggesting content pages to the user that he or she can start contributing to.

Therefore, based on differences that are exposed in figure 5.1 and figure 5.2, we argue, it would be possible to predict that configuration choices made by individual Wikipedias having an impact on the interface might influence the ways in which users perceive the importance of the gender setting. It is during the new user registration process that it is most likely for users to access this setting, if it is clearly visible.

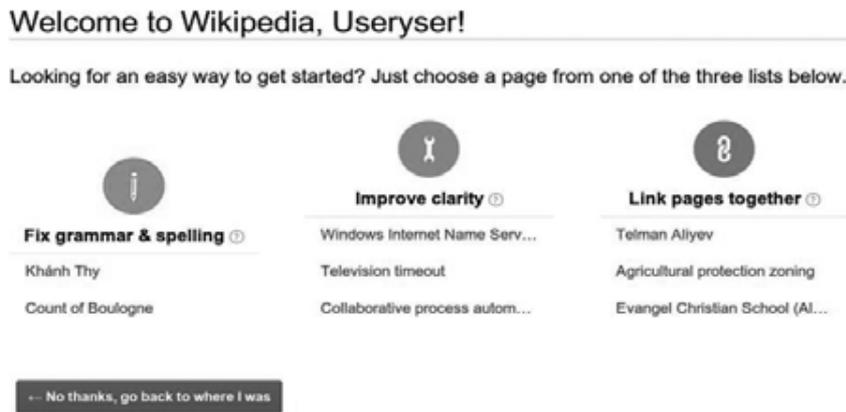


Figure 5.1. Screenshot of the page after creating a new user on English Wikipedia.

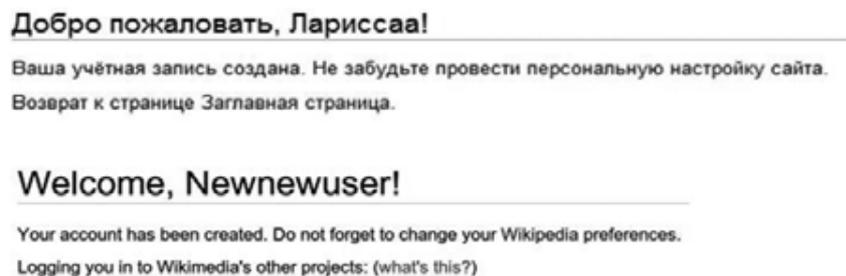


Figure 5.2. Screenshot of the page received after creating a new user on Russian Wikipedia.

PERCENTAGE OF MALE AND FEMALE USERS IN DIFFERENT LANGUAGE EDITIONS OF WIKIPEDIA

The second level of analysis focuses on users who expressed their gender. In this subset, what is the percentage of male and female users (RQ2)?

The sample comprised again solely the users registered since 2010 in all language editions of Wikipedia. In this case we additionally excluded Wikipedias in which the number of users who expressed gender is less than one thousand to obtain a meaningful sample to compare the percentage of females and males, and this additional filter left forty-six language editions of Wikipedia.

Table 5.2 shows the percentage of users who indicated their gender as female. In the English Wikipedia, taken as a comparative baseline, women are 17.36 percent of the 353,056 users who expressed their gender. Respectively, 82.64 percent have

indicated themselves as males. These numbers are a bit higher compared to previously conducted surveys^{24,25} which identified 10 percent of female contributors, but this number is still far away from the desired percentage of 25 percent expressed by Wikimedia Foundation in their strategic plan.²⁶

The question that we can now focus on is if, among the forty-six language editions of Wikipedia with at least twenty thousand users registered since 2010 and at least one thousand of them setting their gender, there are editions in which the percentage of females is greater than that of males, or at least in which the gender gap is more reduced. The results are summarized in table 5.2.

As shown in table 5.2, the Slovenian Wikipedia is the one in which the gender gap is narrower: even if males are still the majority, females account to 39.93 percent of the registered users who expressed their gender. Percentages reported in table 5.2 clearly show a pattern: nine of thirteen of the Wikipedias with relatively more females are related to languages spoken in Eastern Europe while two of them are spoken in East Asia.

Looking at the other side of the list, the Wikipedias with the smallest percentage of females are reported in table 5.3. The smallest shares of females are all present in Wikipedias in languages spoken in India that comprise Hindi, Bengali, Malayalam, Tamil, and Marathi. Then there is Persian or Farsi, spoken predominantly in Iran, preceded by Chinese, Turkish, Korean, and German.

Comparing table 5.1 with tables 5.2 and 5.3, we observe there is no definitive overlap between language editions in which many users express their gender and in

Table 5.2. Wikipedias with the Largest Percentages of Users Setting Their Gender as Female

<i>Language (Wikipedia web address)</i>	<i>Percentage of females among users with gender</i>	<i>Number of users who expressed female as gender</i>	<i>Number of users who expressed their gender</i>
Slovenian (sl.wikipedia.org)	39.93%	849	2,126
Estonian (et.wikipedia.org)	38.12%	390	1,023
Lithuanian (lt.wikipedia.org)	36.20%	640	1,768
Malay (ms.wikipedia.org)	31.27%	661	2,114
Czech (cs.wikipedia.org)	30.49%	2,159	7,082
Ukrainian (uk.wikipedia.org)	30.05%	3,296	10,969
Bulgarian (bg.wikipedia.org)	30.04%	825	2,746
Thai (th.wikipedia.org)	29.42%	2,315	7,869
Georgian (ka.wikipedia.org)	29.13%	416	1,428
Hungarian (hu.wikipedia.org)	27.72%	1,459	5,264
Catalan (ca.wikipedia.org)	26.34%	343	1,302
Portuguese (pt.wikipedia.org)	25.86%	10,214	39,499
Russian (ru.wikipedia.org)	25.67%	35,694	139,038
...			
English (en.wikipedia.org)	17.36 %	61,288	353,056

Table 5.3. Wikipedias with the Smallest Percentages of Users Setting Gender in Preference Page

<i>Language (Wikipedia web address)</i>	<i>Percentage of females among users with gender</i>	<i>Number of users who expressed female as gender</i>	<i>Number of users who expressed their gender</i>
French (fr.wikipedia.org)	16.94%	4,103	24,225
Indonesian (id.wikipedia.org)	16.83%	3,252	19,319
Japanese (ja.wikipedia.org)	16.82%	4,278	25,428
Arabic (ar.wikipedia.org)	16.22%	4,726	29,143
Azerbaijani (az.wikipedia.org)	16.21%	285	1,758
German (de.wikipedia.org)	15.16%	5,052	33,334
Korean (ko.wikipedia.org)	13.89%	496	3,572
Turkish (tr.wikipedia.org)	13.55%	2,353	17,370
Chinese (zh.wikipedia.org)	12.74%	5,709	44,796
Persian (fa.wikipedia.org)	11.14%	1,328	11,917
Marathi (mr.wikipedia.org)	7.75%	141	1,820
Tamil (ta.wikipedia.org)	6.07%	137	2,258
Malayalam (ml.wikipedia.org)	5.34%	71	1,330
Bengali (bn.wikipedia.org)	4.09%	65	1,589
Hindi (hi.wikipedia.org)	3.75%	76	2,024

which there are more females. The Russian, Thai, and Portuguese Wikipedia editions rank high both in terms of expressing gender and females but, for example, the Turkish Wikipedia is in fifth position according to the percentage of users expressing their gender but in a low position for the percentage of women. Similarly, on the Wikipedias in Marathi, Tamil, Japanese, Chinese, and Indonesian, more users expressed their gender than on average, but the percentage of women is among the smallest. These differences might suggest diverse reasons behind the fact that users express their gender and the fact that this gender is female.

The first part of this study, referring to gender expression on Wikipedia, could be explained, as we already did, most in terms of user interface choices. On the other hand, the relative percentage of females and males inhabiting the different instances of the sociotechnical platforms call for a different explanation.

In fact, the percentage of females in different language editions of Wikipedia can be associated with many gender-related indexes published by United Nations, the United Nations Development Programme, and World Bank with regard to the participation of women to societal life. In this contribution we concentrate on a specific report published by UNESCO in 2011 about “Women in Science.”²⁷ The report focuses on researchers, defined as “professionals engaged in the conception or creation of new knowledge, products, processes, methods and systems, as well as in the management of these projects.” This is a definition we believe many Wikipedians would find suitable for themselves as well, and it presents a list of countries with the

Table 5.4. Comparison between Percentage of Women among Researchers in a Country Based on the UNESCO Report and Percentage of Users of a Language Edition of Wikipedia That Specified Their Gender as Female

<i>Country</i>	<i>Percentage of female among researchers</i>	<i>Official language and percentage of female among users of that language edition of Wikipedia</i>
Eastern Europe		
Georgia	52.7%	Georgian — 29.13%
Lithuania	51.4%	Lithuanian — 36.20%
Bulgaria	47.0%	Bulgarian — 30.04%
Ukraine	44.8%	Ukrainian — 30.05%
Russian Federation	41.9%	Russian — 25.67%
Estonia	41.7%	Estonian — 38.12%
Slovenia	35.1%	Slovenian — 39.93%
Hungary	33.0%	Hungarian — 27.72%
Czech Republic	28.9%	Czech — 30.49%
East Asia		
Thailand	51.2%	Thai — 29.42%
Malaysia	37.7%	Malay — 31.27%
China	32.2%	Chinese — 12.74%
Indonesia	30.6%	Indonesian — 16.83%
Korea	15.6%	Korean — 13.89%
India	14.8%	Marathi — 7.75%
		Tamil — 6.07%
		Malayalam — 5.34%
		Bengali — 4.09%
		Hindi — 3.75%
Japan	13.0%	Japanese — 16.82%
Portuguese-speaking countries		
Brazil	48.0%	Portuguese — 25.86%
Portugal	43.0%	Portuguese — 25.86%
Western Europe		
Germany	23.2%	German — 15.16%
Austria	26.4%	German — 15.16%
France	27.4%	French — 16.94%
Turkey	36.3%	Turkish — 13.55%
Iran	26.6%	Persian — 11.14%
Arabic-speaking countries		
Tunisia	47.4%	Arabic — 16.22%
Algeria	34.8%	Arabic — 16.22%
Morocco	27.6%	Arabic — 16.22%
Libya	24.8%	Arabic — 16.22%
Saudi Arabia	1.4%	Arabic — 16.22%

percentage of women among researchers, a sort of gender gap in science. Relevant data are reported in table 5.4.

Data about female researchers for some countries are missing in the UNESCO report, such as, for example, the United States and Canada.²⁸

While in the previous tables about Wikipedia each line referred to a language, in table 5.4 data refer to countries, so it is necessary to keep this important difference in mind and treat comparisons with caution. In particular, the comparison between a country and the official language of that country should be taken only for explanatory reasons. Nevertheless, it is interesting to note that the two variables seem to have a linear relationship. Repeating again the caution about a comparison between countries and languages, the Pearson correlation coefficient of the two variables is 0.7 and represents a strong linear relationship (for languages spoken in India, their average was considered, and Arabic was excluded given the large variability of countries in which that language is spoken, as we comment later).

In particular, languages spoken in countries in which the science gap is more reduced mirror the languages whose Wikipedia editions show a smaller gender gap (table 5.2). This is especially true for many Eastern European countries and for Thailand. A similar point can be made for Brazil and Portugal, countries in which Portuguese is spoken.

On the other side of the spectrum, there are Western European countries such as Germany and Austria or France, whose relative language editions of Wikipedia tend to also have fewer women contributors. Similar arguments hold for Japan, Korea, Iran, China, and Turkey—all of them ranking low both on women in science and women in the relative Wikipedias. In India the share of women in science is among the smallest (14.8 percent). Analogously, the bottom five Wikipedias, in terms of proportion of women contributors, are all in fact languages spoken in India (Marathi, Tamil, Malayalam, Bengali, Hindi). Undoubtedly, we are not claiming a perfect mapping of the two datasets, especially considering the constraints in mapping languages with countries, especially for languages such as Arabic, spoken in many diverse countries, or of Malay, spoken both in Malaysia and Indonesia, and Indonesian.

CONCLUSION

This study compared gender across 289 language editions of Wikipedia. First, we analyzed the extent to which expressing gender is a diffused practice in various Wikipedias. We conclude that the differences in the amount of users expressing their gender can be explained by the differences in the interfaces, both the visibility of gender and the incentive to express it, especially during the process of the new user-profile creation.

The second research question focused on the cross-Wikipedia evaluation of the gender gap. Overall results show that there is not a single sociotechnical system in

which women constitute the majority, thus confirming that the gender gap is not just present in the English Wikipedia but it is diffused across all language editions of Wikipedia. However, there are notable differences: in some Wikipedias (Slovenian, Estonian, Lithuanian) the percentage of women is close to 40 percent, in others (Bengali, Hindi) it is around 4 percent, while on the English Wikipedia, the chosen baseline given its international nature reaches 17 percent.

Notably, languages whose editions of Wikipedia have larger shares of women tend to be spoken in countries with a larger participation of women in science. In conclusion, we observe that, even if Wikipedia is an online system, it reflects the real-world societies that inhabit the different language versions of it, and across languages and countries there are differences in women participation in public life. In particular, given that the context of Wikipedia is about creating knowledge, the best explanatory factor is the participation of women in knowledge-creation activities: the gender gap in different language editions of Wikipedia reflects the gender gap in science across the different countries of the real world.

Future research should conduct interviews with Wikipedians to identify benefits and drawbacks of visible gender settings as well as possible techniques that would encourage more diverse populations of these sociotechnical systems.

NOTES

1. Noam Cohen, "Define Gender Gap? Look up Wikipedia's Contributor List," *New York Times*, January 30, 2011, http://www.nytimes.com/2011/01/31/business/media/31link.html?_r=1&hpw.
2. Wikimedia Foundation, Editor Survey 2011, http://meta.wikimedia.org/wiki/Editor_Survey_2011.
3. Cohen, "Define Gender Gap?"
4. Wikimedia Foundation. Editor Survey 2011.
5. Paolo Massa, and Federico Scrinzi, "Exploring Linguistic Points of View of Wikipedia," in *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (Mountain View, CA: ACM, 2011, 213–4).
6. Wikimedia Foundation, *Wikimedia Strategic Plan: A Collaborative Vision for the Movement through 2015*, February 2011, http://upload.wikimedia.org/wikipedia/commons/c/c0/WMF_StrategicPlan2011_spreads.pdf.
7. Judd Antin, Raymond Yee, Coye Cheshire, and Oded Nov, "Gender Differences in Wikipedia Editing," in *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (Mountain View, CA: ACM, 2011, 11–14).
8. Shyong K. Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R. Musicant, Loren Terveen, and John Riedl, "WP:Clubhouse?: An Exploration of Wikipedia's Gender Imbalance," in *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (Mountain View, CA: ACM, 2011, 1–10).
9. Benjamin Collier and Julia Bear, "Conflict, Criticism, or Confidence: An Empirical Examination of the Gender Gap in Wikipedia Contributions," in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, WA: ACM, 2012, 383–92).

10. Joseph Reagle and Lauren Rhue, "Gender Bias in Wikipedia and Britannica," *International Journal of Communication* 5 (2011). Published electronically August 8, 2011, <http://ijoc.org/ojs/index.php/ijoc/article/view/777/631>.
11. Noriko Hara, Pnina Shachaf, and Khe Foon Hew, "Cross-Cultural Analysis of the Wikipedia Community," *Journal of the American Society for Information Science and Technology* 61, no. 10 (2010): 2097–2108.
12. Ulrike Pfeil, Panayiotis Zaphiris, and Chee Siang Ang, "Cultural Differences in Collaborative Authoring of Wikipedia," *Journal of Computer-Mediated Communication* 12, no. 1 (2006): 88–113.
13. Sabine Niederer, and José van Dijck, "Wisdom of the Crowd or Technicity of Content? Wikipedia as a Sociotechnical System," *New Media & Society* 12, no. 8 (2010): 1368–87.
14. Massa and Scrinzi, "Exploring Linguistic Points of View of Wikipedia," 213–4.
15. Our script is released as open source so that other researchers can replicate our analysis and also improve it. The Python script is at https://github.com/volpino/toolserver-scripts/blob/master/python_scripts/get_gender_table.py.
16. Brent Hecht and Darren Gergle, "The Tower of Babel Meets Web 2.0: User-Generated Content and Its Applications in a Multilingual Context," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, GA: ACM, 2010, 291–300).
17. Wikimedia Foundation, Editor Survey 2011.
18. Cliff Lampe, Nicole Ellison, and Charles Steinfield, "A Familiar Face(Book): Profile Elements as Signals in an Online Social Network," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, CA: ACM, 2007, 435–44).
19. Kristin Stecher and Scott Counts, "Thin Slices of Online Profile Attributes," *Proceedings of ICWSM* (January 2008).
20. Paolo Massa, Maurizio Napolitano, Federico Scrinzi, and Michela Ferron, "WikiTrip: Animated Visualization over Time of Geo-Location and Gender of Wikipedians Who Edited a Page," in *Proceedings of Wikisym 2012: The International Symposium on Wikis and Open Collaboration*, Linz, Austria, 2012.
21. Wikipedia:What Wikipedia is not, http://en.wikipedia.org/wiki/Wikipedia:What_Wikipedia_is_not.
22. Sue Gardner, "Nine Reasons Women Don't Edit Wikipedia (in their own words)," February 19, 2011, <http://suegardner.org/2011/02/19/nine-reasons-why-women-dont-edit-wikipedia-in-their-own-words/>.
23. Email sent by Erik Möller, deputy director of the Wikimedia Foundation, to the Gendergap MediaWiki mailing list, February 14, 2011, <http://lists.wikimedia.org/pipermail/gendergap/2011-February/000449.html>.
24. Cohen, "Define Gender Gap?"
25. Wikimedia Foundation, Editor Survey 2011.
26. Wikimedia Foundation, *Wikimedia Strategic Plan*.
27. UNESCO, "Women in Science by Region, Female Researchers in 2009," <http://www.uis.unesco.org/FactSheets/Documents/fs14-women-science-2011-en.pdf>.
28. UNESCO, "Women in Science by Region, Female Researchers in 2009."