

GENDER GAP IN THE PROCESSES OF SOCIAL PRODUCTION OF KNOWLEDGE. A CASE OF POLISH WIKIPEDIA

KRZYSZTOF GAJEWSKI

Institute of Literary Research of the Polish Academy of Sciences
Warsaw, Poland

The notion of gender gap was used in the economy, education, and many other domains, to refer to phenomena of unequal treatment of females and males, resulting in different outcomes in terms of salary, school grades, and other material or immaterial goods [Pinker 2009, Morris 2012]. This idea has been applied to Wikipedia, an online encyclopedia that "anyone can edit." Despite its openness and the liberty of correction and contribution, Wikipedia manifests several epistemic biases, such as gender and racial bias, geographical imbalance, ideological bias, and many others. It has been noticed that the whole project is inevitably marked by systematic, epistemic bias. Gender bias in Wikipedia is closely related to the gender gap in the group of its users. According to investigations among

its editors only 10-15% is female. Another aspect of the Wikipedia gender gap is the small number of biographical entries on women in comparison with men. This phenomenon is related to the apparent demand for higher notability threshold for women. Moreover, the biographies of men and women are constructed in a different manner and of different elements: the latter concentrates more on family life, whereas the former on public activities [Wagner, Graells-Garrido, Garcia, Menczer. 2016]. In the present paper, some ways of detecting and pointing out the gender gap in Wikipedia will be presented. The sample of the material under scrutiny will be taken mostly from the Polish language version of the online encyclopedia, but also from other versions for comparison purposes.

KEYWORDS: *Wikipedia, gender gap, bias, encyclopedia, internet*

During a great part of the history of human civilization, knowledge was monopolized by a narrow group. In traditional tribe culture, the most secret medical and religious knowledge belonged to a shaman. In developed centralized big cultural formations such as ancient Egypt, only the group of priests had means to elaborate and profit out of their arithmetical, astronomical, and astrological knowledge they developed, protected, and kept in secret. The idea of equal access to knowledge and education has its origins in the 18th-century movement of Enlightenment and became one of the fundamental pillars of modernity.

The symbol of the approach inspired by Enlightenment was Encyclopedia. It aimed at embracing and publishing the entire human knowledge from the point of view of the rational philosophy of the 18th century.

The work of which we give the first volume today, has two objects: as Encyclopedia, it must expose as much as possible, the order & the sequence of human knowledge; as a Systematic Dictionary of Sciences, Arts & of Crafts, it must contain the general principles on each Science & on each Art, either liberal or mechanical, which are the basis, & the most essential details, which make the body & the substance. [Diderot 1751]

The self-definition of Wikipedia is not far from that. It defines itself as “an online free-content encyclopedia project that aims to help create a world in which every single human can freely share in the sum of all knowledge” [About n.d.]. The process of creating Wikipedia content is based on open collaboration, that in turn is explained in the corresponding entry as “any system of innovation or production that relies on goal-oriented yet loosely coordinated participants who interact to create a product [or service] of economic value, which they make available to contributors and non-contributors alike” [Open collaboration. n.d.]¹.

It may seem that “the free encyclopedia that anyone can edit” as one can read on the front page of Wikipedia, should lead to the production of a socially negotiated, democratic and balanced database of knowledge, where everyone is able to find a place to present her/his point of view. Nothing more wrong. Researchers quickly noticed that Wikipedia, similarly to other projects based on open collaboration, is characterized by several biases, which barely makes it an authoritative source of knowledge.

¹ This is actually a quote from a scientific article [Levine 2014: 2].

One of the “five pillars” of Wikipedia editing proclaims that “Wikipedia is written from a neutral point of view” [Five pillars n. d.]. Nonetheless, researches as well as Wikipedians themselves discovered plenty of examples of neutrality violations on Wikipedia, among them gender bias. In the following paper, I will try to show some examples of these phenomena in the Polish language version of Wikipedia.

Gender gap of Wikipedia

The problem of the gender gap is one aspect of gender inequality, characterizing various domains of human social life. Since 2006 the World Economic Forum issues a Global Gender Gap Index Report every year. The authors of the Report introduce Gender Gap Index

as a framework for capturing the magnitude of gender-based disparities and tracking their progress over time. The index benchmarks national gender gaps on economic, education, health and political criteria, and provides country rankings that allow for effective comparisons across regions and income groups. [Global Gender Gap Report 2020: 8]

The notion has been widely used by researchers approaching questions of gender inequalities — among other topics — in traditional societies of India [Somjee 1989], in Afrika [Njuki, Sanginga 2013], in Japan [Steel 2019], in agriculture [Quisumbing et al. 2014], in the economy [Besen-Cassino 2018], in education [Morris 2012]. The latter’s publication concerns “the “new gender gap” in education: “the significantly lower achievement of boys as compared to girls” [Morris 2012: 1]. Nevertheless, the overwhelming majority of gender gap cases are related to the discrimination of women.

The gender gap may concern different aspects of one domain, as in the case of education. On the one hand, boys have lower achievement as compared to girls, but still, the Global Gender Gap Report 2020 states that: “Based on data from the LinkedIn platform, women are under-represented in six of the eight micro-clusters with the highest employment growth rate [people and culture, content production, marketing, sales, specialized project managers, data and AI, engineering and cloud computing]” [5].

In addition to the notion of gender gap [and other gaps, such as technology gap or a communications gap, conf. Dictionary n.d.] there is also the concept of gender bias. The latter is even more popular in the literature. Dozens of diverse biases had been investigated: hindsight

bias, cognitive biases, participation bias, publication bias, media bias, gate-keeping bias, advertising bias, concision bias, corporate bias, mainstream bias, partisan bias, structural bias to mention but a few. The relationship between these approaches can be seen in part in the fact that the gap is quantitative while the bias is qualitative. The gap seems to be something more obvious, definite, and determined than the bias — a hazy, barely discernible tendency. Assuredly, these two notions inter-lapse and the phenomena they designate may have mutual consequences. I will try to clarify this vague wording with examples later. Wikipedians use mostly the term of bias, as in the Wikipedia article titled “Gender bias on Wikipedia,” devoted to the topic, but the gap is not absent either.

State of the Art

One of the first voices on the topic of gender inequality on the Internet was that of Susan Herring. The researcher opposes the notion of gender neutrality of computer-mediated communication:

In many respects, the Internet reproduces the larger societal gender status quo. Top-level control of Internet resources, infrastructure, and content is exercised mostly by men. The largest single activity on the Internet – the distribution of pornography – is not only largely controlled by men, but casts women as sexual objects for men’s use. The sexualization of women carries over into ostensibly neutral domains, such as recreational chat and personal homepages. In serious contexts, such as academic discussion groups, women participate and are responded to less than men. Moreover, it appears to be necessary for women to form their own groups to address their interests, suggesting that the default activities on the Internet address the interests of men. This evidence points to the persistence of gender disparity in on-line contexts, according to the same hierarchy that privileges males over females off-line. [Herring 2003: 218]

Already in 2003, when the English language version of Wikipedia counted less than 200 000 articles², Susan Herring pointed out several phenomena that can be perceived as different faces of the gender gap. They persisted until now, even if transformed. First of all, it is a gender gap in the sense that most technology users are men. Research-

² In May 2020 it contains more than 6 000 000 articles.

ers estimate the number of women among Wikipedia contributors to be between 13%-16% [Hill et al. 2013; Konieczny 2018]. It is actually “not bad” in comparison to the female participation in the open-source movement [FLOSS] that is between 1.1%-1.5% [Reagle 2013].

One of the early studies on the topic was based on the survey conducted among students on their usage of Wikipedia and lead to the conclusions that

Male students were more likely to discount the risks involved when using Wikipedia information compared to their female counterparts. Furthermore, male students had higher ratings than female students regarding most aspects of Wikipedia, including outcome expectations, perceptions about its information quality, belief in the Wikipedia project itself, emotional states while using Wikipedia, confidence in evaluating information quality, and further exploration. Finally, there was no gender difference regarding the number of years of Wikipedia use. However, male students reported having more positive experiences with the information quality of Wikipedia than their female counterparts. [Lim 2010: 212]

A weak point of the method of survey is that it says more about the convictions of the respondents than about the respondents themselves and their real behavior.

An analysis of the factual contribution of users is deprived of this defect. Such is the approach of a Shyong [Tony] Lam and his team. Researchers analyze edits of registered Wikipedians who declared their gender, which makes 2.8% of the general number of contributors [Lam 2011: 3]. They try to verify the hypothesis, such as “Male and female editors focus on different content areas,” “Coverage of topics with particular interest to females is inferior to topics with particular interest to males,” “Females tend to avoid controversial or contentious articles,” “Female editors are more likely to have their early edits reverted,” “Female editors are more likely to stop editing and leave,” “Wikipedia when being reverted as newcomers,” “Female editors are less likely to be blocked” [Lam 2011: 2-3]. Since the research sample of user activity was collected from a small minority of registered Wikipedians who declared their gender, there is no doubt that the data is biased because it is limited to persons exhibiting very untypical attitude towards their gender, thus who are markedly radical in their psychological features and their views.

Another team of researchers formulates hypotheses seeking gender differences “in confidence in one’s expertise, with women expressing greater lack of confidence than men,” “in comfort with editing such that women will express greater discomfort than men with editing other people’s work in Wikipedia,” in that “Women will respond more negatively than men to criticism in Wikipedia” and that “Women will report less discretionary time than men” [Bear et al. 2016: 256-257]. One could suspect that the sheer construction of such research can exhibit gender bias. This research was based on a survey conducted by the Wikimedia Foundation in 2008. The survey was completed by voluntary participants, 1/3 of whom were Wikipedia editors, either occasional or regular. This, again, is definitely not a representative sample of the community of Wikipedia editors.

The second face of the Wikipedia gender gap is the way women are represented in the medium in question. As for the Internet in general, they are usually objectified in the form of objects of male sexual desire. Herring recalls that in the mid-90s searching Internet with the word “women” brought numerous pornography sites [Herring 2003: 214]. The phenomena consisting of representing females from a certain perspective may be identified with gender bias *sensu stricto*. Researchers remark that biographies of women contains usually more facts connected to sex and marriage than men’s [Graells-Garrido et al. 2015]. They indicate also “A strong bias in the linking patterns [that] results in a network structure in which articles about men are disproportionately more central than articles about women” [Graells-Garrido et al. 2015: 165].

Somewhere in between these two forms of the gender gap is the coverage of biographical information about women and topics regarding women. One of the first interventions regarding gender imbalance of Wikipedia compared it with Britannica and found no huge differences:

We conclude that Wikipedia provides better coverage and longer articles, and that it typically has more articles on women than Britannica in absolute terms, but we also find that Wikipedia articles on women are more likely to be missing than are articles on men relative to Britannica. [...] While Wikipedia has more biographies of women than does Britannica in absolute terms [...], Wikipedia tends to be less balanced in whom it misses than is Britannica as seen in the percentages of missing articles. [Reagle 2011: 1155]

This investigation was strictly quantitative and was limited to data possible to express in numbers, such as the number of the articles or their length. One of the famous examples of coverage gaps in Wikipedia was the biography of Donna Strickland. The article on the topic was created on 2 October 2018[⊗] [Donna Strickland: Revision history n. d.], exactly on the day Donna Strickland received the Nobel Prize in Physics. Indeed, high entrance threshold for a possible object of a biographical entry in Wikipedia. Actually, her biography was created on 7 March 2014 at 6:18, but it was deleted four hours later for the reasons of “Unambiguous copyright infringement,” which satisfies the criteria for speedy deletion [conf. Deletion log n.d.]. Donna Strickland most probably has become a victim of the “Matilda Effect,” which leads to diminishing and even covering the role of women in scientific researches.

Margaret W. Rossiter coined the term “Matilda Effect” after the Robert K. Merton concept of “Matthew Effect” in science. The idea is based on the biblical quote from the Gospel of Matthew 13:12 saying that who has a lot, will be given more, but the one who has not much, will be taken away from. Merton remarked that there is a tendency to attribute to famous scientists merits of their less famous colleagues. The same goes for female collaborators, as in the case of Lise Meitner, a long-term collaborator of Otto Hahn, who alone had subsequently been awarded Nobel Prize for fast nuclear fission [Rossiter 1993: 329]. Rossiter names the effect after the name of Matilda Joslyn Gage, 19th -century feminist, free thinker, one of the authors of *Woman’s Bible* [Rossiter 1993: 336].

Other researchers try to look a bit into the content of the articles too. “Male sociologists [...] are more than twice as likely to have a page on Wikipedia [16 percent] than their female colleagues [7 percent] [...]. Similarly, white sociologists are twice as likely [14 percent] to have a page than others [7 percent]” [Adams 2019: 7] — a recent study announces. Early research on gender bias in Wikipedia concludes: “Coverage of topics with particular interest to females is inferior to topics with particular interest to males” [Lam 2011: 2].

Yet another approach is represented by Joseph Reagle. This researcher and hacker collected material from his long years of participation in online communities of Wikipedia and Open Source and drew a bitter conclusion: “Despite the values of freedom and openness, the free culture movement exhibits the same [or worse] imbalance in female participation as the computing culture from which it arose,” since “Nerd and geek identity have historically been understood as being both white and masculine” [Reagle 2013].

To summon up above reflections we can order different forms of gender gap in Wikipedia as follows:

1. participation gap
2. content bias
 1. quantity gap: biographical coverage, but also proportion of presence of women and men on the front page of Wikipedia etc.
 2. quality bias: the way of representation

One should frame the problem under consideration into the broader context of gender main-streaming and gender participation. Because of our own specialization as a literary researcher I will concentrate on the question of women and men representation rather than participation. The problem of gender imbalance of our culture started to attract the attention of researchers relatively recently. The imbalance concerns every-day or professional life, as well as fictional worlds of literary works and movies. "Male as a norm," "It's a man's world" — the idea that governs not only social, but also artistic reality — is still valid in our days.

The question of presence of women in fiction is a part of a more general, critical gender revision of the literary and art canon. A smart, quasi algorithmic tool was provided as a Bechdel-Wallace test. It was created in 1985 as a joke in a comic strip

titled 'The Rule,' in which a woman tells her friend that for her to watch a film it has to meet three conditions. It has to have [1] at least two women in it [2] who talk to each other [3] about something besides a man. This, it is implied, excludes the majority of movies to be found playing in a main-stream cinema. [van Raalte 2015: 16]

Surprisingly often in a movie, there is just one female character, which Katha Pollitt names the Smurfette Principle [Pollitt 1991]³. The phenomenon is evident, especially in movie posters. Moreover, the female character usually behaves stereo-typically and there is always some kind of erotic tension between her and the main male character of the movie.

³ Hereby I would like to thank Ms. Sophia Balicka for this clue.

Method and approach

So far, so good – however, one may observe that the position of the speaking subject is not innocent and that the present discourse by myself [K.G.] is inevitably determined – and biased – by my own conscious and unconscious presumptions as a representative of the male gender. A male researcher's opinion on women's representations will inevitably be skewed and false as representing the outer perspective of domination and oppression. A reply to this objection could be a remark that gender studies embrace all the possible gender variations and that male is a gender too. Moreover, many representations of women are connected to strictly-determined male roles of macho not every man is willing to identify with. A tool of gender oppression has usually a double blade, as we will see in the following part of the paper.

The researcher's avoidance of his own biases is an impossible challenge. This difficulty could be partly solved by a comparative approach, i.e. measuring the bias of Wikipedia content by comparing different language versions of the same article. But even a comparative analysis gives only relative results. Despite this fact one can aspire to some kind of objectivity by applying a simple principle of symmetry, presuming that the questions of gender should be presented equally from both male and female sides occupying the same amount of space and having the same visibility. Future researches should include more elaborated and detailed data including the entire spectre of possible gender identities.

My methodology is based mostly on the micro-analysis of the content of Wikipedia articles. Contrarily to the most popular line of research based on numerical data such as word frequency, I will read the text, almost like any ordinary Wikipedia reader. The difference is due to my taking notes and applying a critical perspective. I will not be able to process a lot of data, but I am supposed to encounter a lot of interesting examples of biases. They are situated on the level of discourse as it is perceived by a common Wikipedia reader. This perspective was overlooked by previous, predominantly quantitative, researches. I concentrated mostly (but not exclusively) on the best-exposed parts of the body of Wikipedia article: on the first paragraph and on the graphic material accompanying and completing textual content. I contrasted it with other language versions of the article, mostly English, but in many cases also Russian, German, French, Spanish, and Italian. I am not excluding quantitative reflection, sometimes taking into consideration such data as the length of an article and frequency of a term.

Research Sample

As the source of data for my analysis, I have chosen 22 articles from the Polish language version of Wikipedia related to the topic of gender. The version of articles I studied was that of 15 June 2020. The research sample is collected in a table, in the three columns presenting entries connected to women, men, and to no specific gender, respectively.

Table 1 Research Sample

Women	Men	No Specific Gender
Feminizm	Maskulinizm	Aborcja
Kobiecość	Męskość	BDSM
Kobieta	Mężczyzna	Cisseksualizm
Matriarchat	Patriarchalizm	Determinacja płci
	Patriarchat	Gender
	Hipoteza o wyższości mężczyzn	Gender mainstreaming
		LGBT
		Płeć
		Role płciowe
		Różnice psychologiczne między płciami
		Seks
		Seksizm

As one can see, the choice of the entries is not at all random. On the contrary, I concentrated on a few most important gender-related notions, with a special interest in topics controversial in Poland. Therefore, the results cannot under any circumstances be considered as a general diagnosis of “gender health” of Polish Wikipedia. The research is aimed at finding new, unknown forms of bias on the level of a particular entry. In further perspective, the results can serve as the basis for a future, more precise, quantitative, algorithmic, digital approach.

Results: Propositions for Classifying Wikipedia Gender Biases

Analysis of first paragraphs of the Polish Wikipedia entries above enumerated has revealed plenty of biased content — biased in various manners, hidden in different aspects of the text. The entire spectrum of heterogeneous phenomena has fallen into several groups according to criteria applied. I will limit my attention to two criteria: carrier [medium] and rhetorical strategy.

Biases by Carrier

First of all, there are a few areas in which biases may be encoded — a few separated carriers, or media. The most obvious case is a biased statement provided in the textual content of the article. There are, however, possible some other, less visible, and ostensive biases such as those hidden in the length of the article, in the percentage of space reserved for coverage of a particular topic, or in the selection of accompanying graphical materials.

As for the first category of textual content bias, further explanation is not needed; moreover, in the subsequent analysis, many examples of bias of this kind will be mentioned. As for the second category, biases incorporated in the size of the text, we shall refer to an example of the entry “Kobieta” and “Mężczyzna.”

Table 2. Length of Article

1.	Language	Pl	Ru	It	Fr	Es	De	En
2.		Kobieta	Женщина	Donna	Femme	Mujer	Frau	Woman
3.	Length [bytes]	5276	123705	15507	102600	32816	50349	72201
4.		Mężczyzna	Мужчина	Uomo	Homme	Varón	Mann	Man
5.	Length [bytes]	14057	54165	17847	26239	28802	16454	18864
6.		K/M	Ж/М	D/U	F/H	M/V	F/M	W/M
7.	Proportion	38%	228%	87%	391%	114%	306%	383%
8.		Kobieta	Женщина	Donna	Femme	Mujer	Frau	Woman
9.	Length [%]	100%	4%	34%	5%	16%	10%	7%
10.		Mężczyzna	Мужчина	Uomo	Homme	Varón	Mann	Man
11.	Length [%]	100,00%	26%	79%	54%	49%	85%	75%

The table summarizes the length of the articles on *woman* and *man* in different language versions of Wikipedia. The third and the fifth row contain the sizes of respective language versions of the articles “Kobie-

ta” [Woman] and “Mężczyzna” [Man]. The unit is a byte that corresponds approximately to one letter of the text, but this value includes also commands of Wiki syntax, which are not displayed on the screen. They are, however, quite good approximations of how long the text of the entry is. In the seventh row, the length of the articles on women compared to the length of the article on men is expressed as proportions. The value corresponding to Polish Wikipedia is the smallest in the research sample. The value, 38%, means that the size of the article on female gender is a bit more than 1/3 of the size of the entry on male. Only Polish and Italian Wikipedia articles on women are shorter than the one on men. In all the other cases, articles on women are longer than articles on men, up to three times longer in the case of French and English Wikipedia.

The ninth and tenths rows consist of the values of the lengths of Polish articles “Kobieta” [Woman] and “Mężczyzna” [Man] against the background of other language versions of Wikipedia. Both articles are more developed on non-Polish platforms, but the gap is especially elevated in the case of “Kobieta”⁴.

As for the next kind of bias, namely the proportion of the text and thus, readers’ attention, an entry “Różnice psychologiczne między płciami” [“Sex differences in psychology”] will provide a very instructive example. In the English version of the entry, the abbreviation “IQ” is cited only twice, in the context of the research conducted in 2007 which stated that the results of research on the gender difference in IQ level are ambiguous and open to interpretation. The Polish version of the entry contains a long and detailed analysis of the problem, citing a lot of sources and researches, which makes up the bigger part of the article.

Table 3 Proportion of length of text

1.		Różnice psychologiczne między płciami	Sex differences in psychology
2.	Length [bytes]	40929	102049
3.	“IQ” occurrences	22	2
4.	Proportion	0.54‰	0.02‰

⁴ These results, however, should be interpreted carefully, at best in comparison with the distribution of similar entries, such as “Kobiecość” [femininity] and “Męskość” [masculinity] that exhibit somehow contrary tendency and may contribute to this phenomenon.

The first row contains data on the size of both language versions of the article. The second row, the number of "IQ" occurrences, and the fourth row, the proportion of the latter expressed as permil, show the saturation level of the concept in the text. The resulting value is 26 times bigger in the Polish version.

Finally, the last form of bias classified as carrier would be graphical content. A very good example will be provided by the already mentioned entry "Kobieta" [Woman]. A pictorial part of this entry consists of two pictures. First one is a "Woman Montage," which contains 20 pictorial representations of "famous women," from Queen of Sheba and Venus, through Mother Teresa and Grace Hopper, computer pioneer, to Elizabeth I of Tudor and Florence Owens Thompson, a photographic icon of the Great Depression in the USA in the 30s. The second one is "Pregnancy 26 week.jpg," subtitled "Kobieta w ciąży" [A pregnant woman]. The young woman is completely nude and shown from the waist up in profile. This very same photo [in a much smaller size, though] contributes to the graphical accompaniment of the English version of the entry as one of the 26 pictures. It appears as the ninth picture in the row. The first one is "Woman mechanic working on engine.jpeg," entitled "A woman inspecting an engine." One should remark this detail. She is not an unqualified, physical worker, but an inspector, most probably an engineer. She is as young as the woman from the pregnancy photo. Unlike her, she is not only dressed but is also wearing protective clothing and a mask, has a discreet makeup, carefully arranged hair and clips. A message of the Polish version of the article, the message whose collective author is the Polish Wikipedia community, is clear: for a woman, there are two ways, either become famous, or pregnant.

Biases by Strategy

Another criterion that proved effective in ordering the collected material was the criterion of rhetorical strategy. During the analysis of the research sample, several rhetorical figures were encountered, concerning both textual, as well as graphical content. I am enumerating them according to increasing visibility and ostentation, partly related to increasing expressive power: concealment [omission], obsolete sources, camouflage [noise], presupposition/implication, open text, and drastic content.

Concealment, or Lost In Translation

This strategy has a purely negative character since it does not involve adding to the content, but simply removing it. It is ordinary censorship in which when translating an entry from a foreign language some passages are omitted.

To understand better how it works, let us take a closer look at the first four sentences of the entry “Feminizm” [“Feminism”] from the Polish language version of Wikipedia and contrast it with the English version of the article. To make it shorter, instead of the sentences I will evoke their themes⁵.

Table 5.

No	Polish	No	English
1 st	Aims of feminism	1 st	Aims of feminism
2 nd	Assumptions of the feminist worldview	2 nd	Assumptions of the feminist worldview
3 rd	Activities of feminist organisations	3 rd	Postulates for practical action
4 th	More on these activities	4 th	Activities of feminist organisations
5 th	-	5 th	More on these activities

The table above presents the topics of introductory sentences (i.e. above the box “Contents”) of both the Polish and English versions of the entry on feminism. The gap in the box for the fifth sentence of the Polish version means that the introductory part counts only four sentences. Nevertheless, as the topics and even the structure of subsequent sentences suggest, one is a translation of the other. The direction most probably was from English to Polish. If so, there is an obvious omission. The fourth sentence, in general, on the practical implications of feminism, has been lost in translation. The “satanic verse” of the entry “Feminism” goes as follows: “Efforts to change that include fighting gender stereotypes and seeking to establish educational and professional opportunities for women that are equal to those for men” (Feminism n.d.). This sentence was impossible to swallow for the editors of Polish Wikipedia.

⁵ Theme, or topic, as contrasted with the comment — rheme, according to the terminology of the Prague School.

In the exact same manner, the second sentence of the entry on sexism became its “satanic verse.” It says “Sexism can affect anyone, but it primarily affects women and girls” [Sexism n.d.] and has also been excluded by the editors of the Polish Wikipedia. I guess an anthology of such gender “satanic wikiverses” would be an interesting and informative reading.

Obsolete Sources

Another strategy used by the authors of biased content on Wikipedia is citing obsolete and no longer valid sources. That is the case of an entry “Seks” [“Human sexual activity”] and “Matriarchat” [Matriarchy].

The first one contains a chapter “Badania zachowań seksualnych” [Research on sexual behaviour]. The content of the chapter is based in great part on the Kinsey report from 1948 and 1943 and research by Masters and Johnson published in 1966 and 1970. There is no doubt that the views presented there will be more conservative than the ones based on more recent sources. In the English version of the entry, the Kinsey report is mentioned only once, in one sentence, while Masters and Johnson are absent.

Another example concerns one particular controversy, namely the hypothesis of prehistorical matriarchy, which to this day is an open and complex question on the border of science and politics. The Polish Wikipedia article on matriarchy cites as a source works of Feliks Koneczny, Polish historian and philosopher active before World War II and known for his conservatism, nationalism, and anti-Semitism.

Camouflage

Sometimes “problematic” content that contradicts Wikipedia editors’ world view is presented in a disguise and is covered by unnecessary complexity of the message or sheer informational noise.

An introductory sentence to the entry on cisgender can serve as an example of a bias hidden in a camouflage of stylistic chaos. In the English version, we find a short and clear definition: “a term for people whose gender identity matches their sex assigned at birth” [Cisgender, n.d.]. Polish version brings such a formulation: “expression regarding descriptions of gender identity, where personal experience related to it is consistent with the sex that was assigned to the person at birth”⁶.

⁶ “wyrażenie dotyczące opisów tożsamości płciowej, gdzie osobiste doświad-

Saying easy things in a complex way is most probably aimed at discouraging people to get deeper into the topic.

An example of sheer informational noise one can find in the graphical framing of an entry "Role płciowe" [Gender roles]. The first photo embedded in the article is "Bagpipe performer.jpg," a male dressed in kilt and playing a musical instrument. No need to say, that it does not appear in any other analysed language version of Wikipedia. In the English and the Russian language version of the article, the first graphical illustration of the entry is a series of photos showing "Men and women in non-traditional gendered occupations" [Gender role n.d.]. The photo of a male wearing a kilt as an introductory illustration to the question of gender role is information noise impeding access to the knowledge on this subject.

Presupposition

Textual strategies, like the ones described above – e.g. concealment, obsolete sources, etc. – are not easily visible, since obscurity is one of their main assumptions. Stylistic camouflage is more ostentatious because its function is to distract the reader's attention and direct it elsewhere. The next rhetorical strategy, presupposition, is not overt either. I understand it here loosely, as an implication of a message. It is not the main topic of the message, but a side comment, smuggled on an occasion.

An example of this phenomenon can be provided again by an entry "Kobieta" [Woman]. The second and the last sentence of the introduction to the entry said: "In biology, it is assumed that differences in the genotype of a woman and a man determine the sexual dimorphism of human"⁷ [Kobieta n.d.]. This formulation implies that there are two different genotypes: a genotype of a woman and the genotype of a man.

The English version mentions genetics in one of its four-paragraph introduction saying: "Typically, a woman has two X chromosomes and is capable of pregnancy and giving birth from puberty until menopause" [Woman n.d.]. This way of expressing the idea of genetic difference is less definite due to the word "typically." The difference is actually

czenia z nią związane są zgodne z płcią biologiczną, która danej osobie była przypisana w momencie urodzenia"

⁷ "W biologii przyjmuje się, że różnice w genotypie kobiety i mężczyzny determinują dymorfizm płciowy człowieka."

evoked [“two X chromosomes,”] which makes it less “demoniac” than the unspecified “differences in the genotype of a woman and a man” [Kobieta n.d.).

The German version sounds a bit like an amplified Polish version: “Development as a woman is genetically determined by a pair of chromosomes XX and the lack of the Y chromosome necessary for male development, which controls the development of primary and secondary sex characteristics”⁸ [Frau n.d.). However, the formulations are much less firm, because “bedingen” [en. require, determine] seems less strict than “determine” [de. bestimmen, feststellen].

The French version: “Its genetic sex or genotypic sex is determined by the presence of two X chromosomes [and / or by the absence of the SRY gene] on the twenty-third pair. The embryo is however physiologically undifferentiated until the seventh week of pregnancy”⁹ [Femme n.d.). This approach is similar to the English and German ones, focusing on the difference. Moreover, it notes that sex difference appears at some stage of fetal life.

The Spanish, Italian, and Russian versions do not mention genetics in the introductory part of the entry on *woman* at all. To summon up this example: covering genetic differences and presenting their central role in defining the term implies stressing and overestimating the importance of gender differences.

Another example involving presupposition would be a graphical accompaniment of the entry “Gender.” The first photo in the Polish language version of the entry is titled “Man and woman in formal wear” [Gender n.d.) — may very well be a wedding photo. Their height and clothes are different, displaying a play on the contrast of their black and white clothes. The English version embeds in this place “A Gender symbols intertwined,” the implication of which is close relation and connection between genders.

⁸ “Die Entwicklung als Frau ist genetisch bedingt durch ein Chromosomenpaar XX und das Fehlen des für die männliche Entwicklung notwendigen Y-Chromosoms, durch die die Entwicklung primärer und sekundärer Geschlechtsmerkmale gesteuert wird.”

⁹ “Son sexe génétique ou sexe génotypique est déterminé par la présence de deux chromosomes X [et/ou par l’absence du gène SRY] sur la vingt-troisième paire. L’embryon est cependant physiologiquement indifférencié jusqu’à la septième semaine de grossesse.”

Let us take a closer look at the logo used by the Polish Wikipedia community as a symbol for the template “Feminizm.” It is woman-power symbol — a clenched fist in Venus sign [Feminizm n.d.]. In the English version, the role of the template is played by a box leading to the whole Feminism portal. The logo is basically a Venus sign, with no fist. Here femininity does not need to show the fist.

These are only a few from many examples of more or less obvious presupposed or implied gender-biased content I encountered during the investigation of the research sample.

Open Text

A yet more ostentatious strategy is the direct formulation of a biased statement — open text. In such cases, biased content is included in the body of the article. Usually, it is placed somewhere deep in the text, but sometimes it takes a more exposed position as the introductory part of the article.

A good example of this textual phenomenon is the entry “Seks” [Human sexual activity]. The first part of the introductory sentence of the entry consists of a concise characterization of the notion in question. It goes as follows: “entirety of behaviors resulting from sex drive and aimed at satisfying individual sexual needs of a person”¹⁰ [Seks n.d.]. The English version of the article starts with such formulation: “the manner in which humans experience and express their sexuality” [Human sexual activity n.d.]. In this case, the bias is explicit in the limitation of the variety of functions human sexual life can play in human life.

Another, quite shocking example, is an entry entitled “Hipoteza o wyższości mężczyzn” [Hypothesis on the supremacy of men]. It’s beginning states: “Hypothesis on the superiority of men – an assumption about the psychological, especially intellectual advantage of men in the species *H. sapiens*, which would have a biological basis” [Hipoteza o wyższości mężczyzn n.d.]. As we can see from the lack of decision as to the choice of term, the editors are not sure whether they refer to a hypothesis or an assumption. In the following part of the entry, several sources from after 2000 documenting and supporting the above-mentioned “hypothesis” were cited. Bias is not covered here, on the contrary — it is proposed as mainstream discourse. This entry was deleted

¹⁰ “ogół zachowań wynikających z popędu płciowego i mających na celu zaspokojenie indywidualnych potrzeb seksualnych człowieka”

on 8 December 2019. Nevertheless, its content has been incorporated into the article “Różnice psychologiczne między płciami” [Sex differences in psychology].

Drastic Content

The final and most expressive way of smuggling bias into Wikipedia article I found during this investigation was a type of content that can be classified as drastic. I guess it deserves to be treated separately, even though I can provide only one example.

It is a part of the entry on abortion as the first one [out of two] pictures accompanying quite an elaborate article [96 233 bytes]. The picture is entitled “The embryo after opening the fallopian tube with an ectopic pregnancy approximately 7 weeks after fertilization” [Aborcja n.d.]. Let us take a closer look at the content of the photo. It has a quite long description. This is a part of it:

This photo of an opened oviduct with an ectopic pregnancy features a spectacularly well preserved 10-millimeter embryo. It is uncommon to see any embryo at all in an ectopic, and for one to be this well preserved [and undisturbed by the prosecutor’s knife] is quite unusual. [Uthman 2001]

What we can learn for sure is that it is not a usual image related to abortion, but a very particular case: ectopic pregnancy. Moreover, the illustration is not typical even for this particular case of ectopic pregnancy. What we see is not false, but it is not the truth either.

The second image used here is an abortion laws map, displaying the legal status of abortion in the world. Poland is a single yellow stain [“Prohibited with exceptions for maternal life and health, rape, and fetal defects,”] on a uniform blue background of Global North [“Allowed on request”]. No other reviewed language version links to the picture of ectopic pregnancy.

Conclusions

The results collected above raise many questions, doubts, but also some conclusions. Is the method applied a good tool for bias detection? For sure, I did not enumerate all the possible forms of biases. Classifications by the carrier and by the rhetorical strategy are just two

out of many possibilities. Another way is a systematic presentation according to function, such as performative, commercial, ludic, etc. Undoubtedly, there is much more to discover, that is only a matter of the number of articles analysed. As for the 22 articles analysed here, they were chosen partially from a variety of articles discussing controversial topics, in which the number of manifestations of bias seems substantial and worth further investigations. I hope I showed that not only quantitative and digital but also qualitative and traditional approaches can teach us something about Wikipedia.

Communication technology, as well as computer media, is not gender-neutral. Historical, political, and social conditions influenced strictly technical, computer-mediated communication. Human culture and human behaviour determine the way humans use technology. I wanted to demonstrate that this has a lot of consequences. One of them is a form and content of the knowledge stored and published on the internet on Wikipedia portals. Open and free knowledge happened to replicate discriminatory discourse, playing rather a performative, than a constative role. The fact that we are able to perceive inaccuracies and distortions suggests that we live in the era of breakthroughs.

References

- Adams J., Brückner H., Naslund C. 2019. Who Counts as a Notable Sociologist on Wikipedia? Gender, Race, and the “Professor Test.” *Socius: Sociological Research for a Dynamic World*, Volume 5, 1-14.
- Besen-Cassino Y. 2018. *The Cost of Being a Girl. Working Teens and the Origins of the Gender Wage Gap*. Philadelphia: Temple University Press.
- Diderot, D. 1751, Discours préliminaire des éditeurs. In: Diderot, D. et al. [eds.] *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers*. Paris <http://enccre.academie-sciences.fr/encyclopedia/volume/1?n=4> [retrieved 30.05.2020]
- Graells-Garrido, E., Lalmas, M., Menczer, F. 2015. First women, second sex: Gender bias in Wikipedia. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. New York, NY: ACM. 165-174.
- Konieczny P., Klein M. 2018. Gender gap through time and space: A journey through Wikipedia biographies via the Wikidata Human Gender Indicator. *New Media & Society*. Volume 20, Issue 12, 4608-4633.

- Lam, S. T. K., et al. 2011. WP: Clubhouse? An exploration of Wikipedia's gender imbalance. In: *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*. New York, NY: Association for Computing Machinery Press. 1-10.
- Levine, S. S., Prietula, M. J. 2014. Open Collaboration for Innovation: Principles and Performance. *Organization Science*. 25(5), 1414-1433.
- Morris E. W. 2012. *Learning the Hard Way. Masculinity, Place, and the Gender Gap in Education*. New Brunswick: Rutgers University Press.
- Njuki J., Sanginga P. C. [eds.] 2013. *Women, Livestock Ownership and Markets. Bridging the gender gap in Eastern and Southern Africa*. New York: Routledge.
- van Raalte, Ch. 2015. No Small-Talk in Paradise: Why Elysium Fails the Bechdel Test, and Why We Should Care. In: H. Savigny et al. [eds.]. *Media, Margins and Popular Culture*. Basingstoke, Hampshire: Palgrave Macmillan.
- Reagle J., Rhue L. 2011. Gender bias in Wikipedia and Britannica. *International Journal of Communication*. Volume 5. 1138-1158.
- Reagle J. 2013. Free as in sexist? Free culture and the gender gap. *First Monday*, 18(1). DOI:10.5210/fm.v18i1.4291
- Rossiter M. W. 1993. The Matthew/Matilda Effect in Science. *Social Studies of Science*. Volume 23. 325-341.
- Somjee G. 1989. *Narrowing the Gender Gap*. Basingstoke: Palgrave Macmillan.
- Steel G. 2019. *Beyond the Gender Gap in Japan*. Ann Arbor: University of Michigan Press.
- Quisumbing A. R. et al. [eds.] 2014. *Gender in Agriculture. Closing the Knowledge Gap*. Dordrecht: Springer Netherlands.

Online References

- About. n.d. In. Wikipedia. <https://en.wikipedia.org/wiki/Wikipedia:About> [retrieved 30.05.2020]
- Aborcja. n.d. In: Wikipedia. <https://pl.wikipedia.org/wiki/Aborcja> [retrieved 30.05.2020]
- Cisgender. n.d. In. Wikipedia. <https://en.wikipedia.org/wiki/Cisgender> [retrieved 30.05.2020]
- Deletion log. n.d. https://en.wikipedia.org/w/index.php?title=Special:Log/delete&page=Donna_Strickland [retrieved 30.05.2020]
- Dictionary.com. n.d. <https://www.dictionary.com/> [retrieved 30.05.2020]

Donna Strickland: Revision history. n. d. https://en.wikipedia.org/w/index.php?title=Donna_Strickland&dir=prev&action=history [retrieved 30.05.2020]

Feminism. In: Wikipedia. <https://en.wikipedia.org/wiki/Feminism> [retrieved 30.05.2020]

Feminizm. In: Wikipedia. <https://pl.wikipedia.org/wiki/Feminizm> [retrieved 30.05.2020]

Femme. In: Wikipedia. <https://fr.wikipedia.org/wiki/Femme> [retrieved 30.05.2020]

Frau. n.d. In: Wikipedia. <https://de.wikipedia.org/wiki/Frau> [retrieved 30.05.2020]

Gender. In: Wikipedia. <https://pl.wikipedia.org/wiki/Gender> [retrieved 30.05.2020]

Gender role. In: Wikipedia. https://en.wikipedia.org/wiki/Gender_role [retrieved 30.05.2020]

Global Gender Gap Report. 2020. Geneva: World Economic Forum http://www3.weforum.org/docs/WEF_GGGR_2020.pdf [retrieved 30.05.2020]

Hipoteza o wyższości mężczyzn. In: Archive.org https://web.archive.org/web/20140604032917/https://pl.wikipedia.org/wiki/Hipoteza_o_wy%C5%BCszo%C5%9Bci_m%C4%99%C5%BCczynn [retrieved 30.05.2020]

Kobieta. n.d. In: Wikipedia. <https://pl.wikipedia.org/wiki/Kobieta> [retrieved 30.05.2020]

Pollitt K. 1991. "Hers; The Smurfette Principl." The New York Times Magazine. April 7. <https://www.nytimes.com/1991/04/07/magazine/hers-the-smurfette-principle.html> [retrieved 30.05.2020]

Seks. n.d. In: Wikipedia. <https://pl.wikipedia.org/wiki/Seks> [retrieved 30.05.2020]

Sexism. n.d. In: Wikipedia. <https://en.wikipedia.org/wiki/Sexism> [retrieved 30.05.2020]

Uthman E. File:Tubal Pregnancy with embryo.jpg. 2001. https://commons.wikimedia.org/wiki/File:Tubal_Pregnancy_with_embryo.jpg [retrieved 30.05.2020]

Wikipedia: Five pillars. n.d. In: Wikipedia. https://en.wikipedia.org/wiki/Wikipedia:Five_pillars [retrieved 30.05.2020]

Woman. n.d. In: Wikipedia. <https://en.wikipedia.org/wiki/Woman> [retrieved 30.05.2020]