

# Assessing Gender Bias in Wikipedia: Inequalities in Article Titles

Agnieszka Falenska and Özlem Çetinöglü

Institute for Natural Language Processing

University of Stuttgart

{falenska, ozlem}@ims.uni-stuttgart.de

## Abstract

Potential gender biases existing in Wikipedia’s content can contribute to biased behaviors in a variety of downstream NLP systems. Yet, efforts in understanding what inequalities in portraying women and men occur in Wikipedia focused so far only on *biographies*, leaving open the question of how often such harmful patterns occur in other topics. In this paper, we investigate gender-related asymmetries in Wikipedia titles from *all domains*. We assess that for only half of gender-related articles, i.e., articles with words such as *women* or *male* in their titles, symmetrical counterparts describing the same concept for the other gender (and clearly stating it in their titles) exist. Among the remaining imbalanced cases, the vast majority of articles concern sports- and social-related issues. We provide insights on how such asymmetries can influence other Wikipedia components and propose steps towards reducing the frequency of observed patterns.

**Bias statement** Inequalities in how men and women are represented in Wikipedia titles can be captured by NLP models and translate into biased behaviors creating *representational harms* (Blodgett et al., 2020). For example, if by default Wikipedia articles about national sports teams are about male teams, then a search engine might assume that a prototypical sportsperson is a man. Such a system when asked about *famous volleyball players* might exhibit recognition bias and return no women. Similarly, if Wikipedia provides special articles listing *women photographers* next to *photographers*, then an automatic knowledge extractor trained on such data might learn and propagate stereotypical generalizations that women within these occupations are an exception and should hold special qualities. Our work limits itself to binary gender values in extracting gender bias patterns from Wikipedia titles due to data scarcity. We acknow-

ledge that not incorporating other values into our analysis indirectly causes recognition bias against non-binary people.

## 1 Introduction

Gender bias can be defined as a systematic preference or discrimination against people of a particular gender (Friedman and Nissenbaum, 1996). NLP systems that exhibit such biased behavior can perform better for the favored gender, e.g., speech recognizers that achieve higher accuracy for male voices (Tatman, 2017), or reinforce harmful stereotypes, e.g., social media platforms that misgender LGBT+ community members (Villaronga et al., 2021).<sup>1</sup>

Biased behaviors have (in-)direct origins in statistical patterns occurring in data that the NLP models are trained on (Sun et al., 2019). Such patterns contribute to different types of biases, such as selection bias that comes with the gold-standard annotated resources, or semantic bias related to the information encoded in pre-trained word embeddings. As noted by Shah et al. (2020), understanding the origins of observed biases is crucial in developing countermeasures for them. As a consequence, diagnosing what type of biases occur in the primary training data is an important initial step in this process.

Wikipedia is one of the largest and most commonly used sources of training data for NLP models: it serves as unannotated data for pre-training word representations (Devlin et al., 2019), a textual source for annotated corpora (Webster et al., 2019), or treebanks (de Kok, 2014; Zeldes, 2017). Wikipedia is also a community-based effort created by a predominantly male group of editors (Lam et al., 2011; Collier and Bear, 2012). This homogene-

<sup>1</sup>We refer to Sun et al. (2019) for fine-grained categorization and more examples of harms and biases in NLP tasks.

ity of Wikipedia authors constantly raises the question of possible *inequalities in ways that people of different genders are represented in Wikipedia articles* (Callahan and Herring, 2011; Reagle and Rhue, 2011; Wagner et al., 2015; Konieczny and Klein, 2018; Schmahl et al., 2020, among others). For example, Wagner et al. (2015) analyzed articles about notable men and women and observed systemic lexical and structural asymmetries, i.e., articles about women contain more family- and gender-related words, and more hyperlinks to articles about men than the other way around. In their following analysis, Wagner et al. (2016) showed that gender inequalities can be observed also on other dimensions, e.g., women have to be more notable to have their biographies than men. These patterns occurring in Wikipedia have been shown to directly influence NLP models, such as word embeddings (Schmahl et al., 2020) or relation extraction (Gaut et al., 2020).

All the above-mentioned studies focus on diagnosing gender bias in *biographies*. Biographies are also the main focus of the Wikipedia community when it comes to mitigating the content gender gap. For example, the project *Women in Red*<sup>2</sup> aims at increasing the number of female biographies. However, as pointed out by Criado Perez (2019, p. 13), inequalities in the way that women and men are portrayed in Wikipedia are present also in other domains. For example, articles about national sports teams such as *England women’s national football team* and *England national football team* commonly omit the word *men’s* in titles of men’s teams, presenting them as if they are the default concept. Yet, a computational method to find such gender-related asymmetries on a bigger scale and understanding of how often they occur in Wikipedia articles are missing.

In this paper, we make an initial step in assessing *gender-related inequalities in Wikipedia titles*. We design a simple three-step heuristic for filtering articles that describe specific concepts and topics, i.e., the above-mentioned *English football team*, in an unbalanced way (Section 2). We apply the proposed method to four Wikipedia editions: Turkish, English, German, and Polish, and find coherent patterns across all four languages (Section 3). Only half of the articles that use gender-indicating words such as *men’s* or *female* in their titles have

their symmetrical counterparts describing the same concept for the other gender. Among the remaining imbalanced cases, the vast majority either describe male-related concepts as generic or represent women as an exception within a more general topic. We discuss the possible harmful effect of the diagnosed title inequalities on other Wikipedia components, such as cross-lingual hyperlinks and NLP models that use them (Section 4). Finally, we propose steps towards reducing the frequency of discovered patterns (Section 5).

## 2 Methodology

Our main goal is to recognize topics and concepts that are described in Wikipedia asymmetrically, i.e., with respect to only one gender. Since finding all possible Wikipedia inequalities is an immense challenge, in this paper, we focus on *article titles*. In Section 4, we provide additional insights on other Wikipedia components that have a direct connection to titles thus should be investigated further in the future.

We design a simple three-step methodology for spotting Wikipedia inequalities in titles, which is presented in Figure 1. Below we describe each of the steps in more detail, but first, explain the selection of languages used for evaluation.

**Languages** We process four languages: Turkish, English, German, Polish (in the order of gender marking in the language) (Stahlberg et al., 2007; Hellinger and Bussmann, 2001, 2003).

German and Polish are Germanic and Slavic languages respectively and both have grammatical gender, though gender marking is more prominent in Polish. In these languages, nouns and (third-person) pronouns have grammatical gender. Determiners and attributive adjectives agree with nouns. Reflexive pronouns, however, are not gender-marked. Polish, in addition, has gender agreement in predicative adjectives, and verbs in past and future tenses agree with the gender of the subject. English is a Germanic language that falls under natural gender languages. Almost all nouns are gender-neutral, but third-person personal pronouns are gendered. Turkish is a genderless language from the Turkic language family. There are no gendered nouns or pronouns in principle, exceptions are mostly loanwords (e.g. *aktör – aktris*). Having no explicit gender markers makes it harder to observe gender bias linguistically, e.g., approaches using pronouns (Rudinger et al., 2018; Zhao

---

<sup>2</sup>[en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Women\\_in\\_Red](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red)

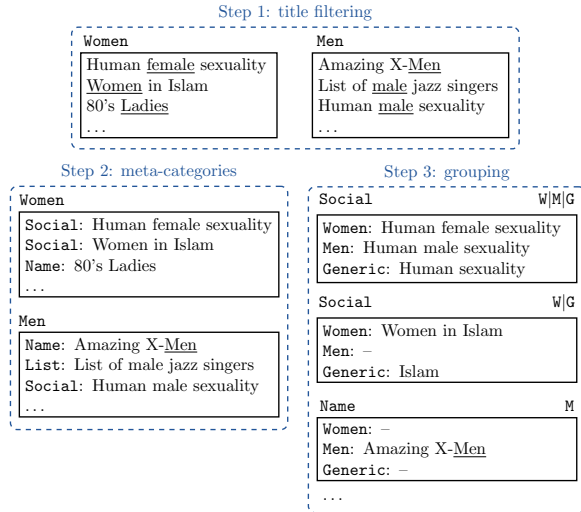


Figure 1: Three-step method used for finding concept-related article tuples.

et al., 2018; Webster et al., 2018) are not applicable to Turkish. Braun (2001) utilizes sociolinguistic tests to demonstrate that Turkish indeed exhibits inherent gender bias that she refers to as *covert gender*.

**Step 1: title filtering** The first step in Figure 1 consists of filtering all Wikipedia titles that describe a particular topic or concept from a gender-related perspective and dividing them into Men and Women groups. For this, we manually select a list of word indicators, i.e., words such as *men* or *female*, and search for titles that contain at least one of them (underlined words in Figure 1).

Table 1 lists word indicators used for the filtering. We intentionally use plural nouns *women* and *men* and not singular *woman* and *man* for English, German and Polish, since we found that the latter predominantly occur in proper names, such as the movie title *Scent of a Woman* or the island *Isle of Man*.<sup>3</sup> Turkish, on the contrary, has both singular and plural forms, and heavily uses the singular form *kadın* which translates to *women's* and *female* in context, as well as the standard *woman* meaning. Moreover, for German, Polish, and Turkish the list in Table 1 is extended with all inflected forms of the presented words.

It is important to note that by filtering the articles only through Men and Women indicators we leave out non-binary genders (Richards et al., 2016) and do not address the non-binary gender biases in this

<sup>3</sup>For example, in the English Wikipedia 81% of titles containing the word *woman* and 71% with the word *man* received the meta-category Names.

	Women	Men
English	women, ladies, female, feminine	men, gentlemen male, masculine
German	frauen-, damen-, weiblich, -innen	männer-, herren- männlich
Polish	kobiety, żeński	kobiecey, mężczyźni, męski
Turkish	kadın, kadınlar, bayan, bayanlar	erkek, erkekler

Table 1: Indicators used for filtering gender-related articles. Indicators are mostly words, but also prefixes and suffixes for German.

paper. This decision is based on the current Wikipedia coverage that provides very little content on other genders. For example, we find 125 articles with the word *transgender* in the title, compared to 38385 and 34240 for the words *men* and *women*, respectively. Similarly, according to the *Denelezh tool*<sup>4</sup> only 0.066% biographies in the English Wikipedia are about people of other genders, compared to 18.6% of female biographies.

**Step 2: meta-categories** In the next step, we assign one meta-category to all the filtered articles. We use five meta-categories that represent the best the majority of gender-related articles: *Sports*, for articles about sports teams or events, *Lists*, for listings of particular people or organizations, *Social*, for articles related to history, awards, gender issues, etc., *Names*, for proper names, i.e., articles about titles of movies, books, names of universities, etc., and *Other*, for articles that did not fit into any other meta-category.

The assignment of meta-categories is based on a list of manually selected keywords<sup>5</sup> and takes into consideration the titles of the articles as well as their Wikipedia categories (see Figure 3 and Section 4 for an example Wikipedia page, categories, and redirections).

**Step 3: grouping** In the final step, we group articles into *concept-related tuples*. Each tuple can be built from three articles: Women, Men, and Generic (notice in Figure 1 that some articles in tuples might be missing).

First, we search for pairs of Men and Women articles that describe the same concept. We determine this by simply removing the gender-related

<sup>4</sup>denelezh.wmcloud.org/

<sup>5</sup>All the manually-designed filters and source code are available for download on the first author's website and under the address <https://github.com/AgneszkaFalenska/GeBNLP2021>.

indicators and pairing articles with the same remaining titles, e.g., *Human female sexuality* and *Human male sexuality*. To limit the number of incorrect tuples, we pair only articles that belong to the same meta-category and completely leave out the ones from the `Names` group.

Second, we fill the collected tuples with `Generic` articles using the same approach as in the previous step, i.e., removing gender-related indicators and pairing titles with the same remaining words (e.g., title *Human sexuality* for the example above). However, to increase the coverage at this stage we apply few additional heuristics. Firstly, we search for `Generic` articles across all Wikipedia titles as well as redirections. Secondly, for languages with inflection, we lemmatize titles to be able to find mappings between such titles as *Kobiety w Islamie* (*Women in Islam* [`case:loc`]) and *Islam* (*Islam* [`case:nom`]).<sup>6</sup> Finally, we apply language-specific and manually designed rules to find mappings between titles with specific noun phrases. For example, in the Polish title above we additionally remove the word *w* (*in*) to be able to find the pairing with the `Generic` title *Islam*.

**Result** The methodology described above provides a list of concept-related tuples of a maximum of three articles. Each of them has assigned one meta-category (`Sports`, `Lists`, `Social`, `Names`, or `Other`) and falls into one of six groups: `W|M|G` (if all three articles are present), `W|M` (if no `Generic` article was found), `W|G` and `M|G` (if no `Men` or `Women` article was found), and `W` and `M` (for tuples with only either `Women` or `Men` present).

### 3 Inequalities in Titles

In this section, we apply the described methodology to four Wikipedia editions: English, German, Polish, and Turkish.<sup>7</sup> First, we look at statistics of filtered articles and tuples to provide insights on the coverage of the proposed method (Section 3.1). Then, we detect types of inequalities occurring in titles (Section 3.2) and finally assess which concepts and topics they relate to the most (Section 3.3).

#### 3.1 Statistics

**Gender-related articles** Table 2 provides the frequency of filtered gender-related articles, i.e.,

	English	German	Polish	Turkish
Women	36241	8679	5568	2400
Men	39069	6356	7994	1585
TOTAL	75310	15035	13562	3985
%WIKIPEDIA	1.23%	0.67%	0.98%	1.04%

Table 2: The frequency of gender-related articles.

	English	German	Polish	Turkish
W M G	129	333	128	15
W M	18033	3480	3712	896
W G	4438	1895	407	161
M G	155	52	10	4
W	13641	2971	1321	1328
M	20752	2491	4144	670
TOTAL	57148	11222	9722	3074

Table 3: The frequency of concept-related tuples.

the result of the first step of the method presented in Figure 1. Our simple filtering heuristic finds between 75k and 4k articles, depending on the size of the Wikipedia that it starts from. In three out of four cases this constitutes around 1% of all the articles. German is an outlier here, with only 0.67% of the whole Wikipedia. One of the reasons might be the high frequency of compounds in German. Although we introduce additional heuristics for this language that look at prefixes of words, such as *männer-* in *Männerorchester* (*Men’s orchestra*), we do not find cases where the gender-related indicator appears in the middle of the word, e.g., *frauen-* in *Deutscher Landfrauenverband* (*German Rural Women’s Association*) or *männer* in *Weltmännertag* (*Men’s World Day*).

Another reason for German being an outlier might be the fact that it is a gender-marking language. Female-related titles, such as *Liste von Dramatikerinnen* (*List of female dramatists*) do not contain a separate gender-indicator such as the word *female* in the English translation. Instead, the gender of the subject is indicated by the suffix *-innen*. To cover such cases, we add this suffix to our gender-indicators, however, we accompany it with additional strong filtering rules to not take titles such as *Webspinnen* (*Spiders*) or *Drei Finnen* (*Three Finns*) as instances of the `Women` group.<sup>8</sup>

<sup>6</sup>Lemmatization with spacy v.3.0 ([spacy.io/](https://spacy.io/)).

<sup>7</sup>Wikipedia IDs: English (2021-03-20, 6144966 non-disambiguation articles), German (2021-03-01, 2241506 articles), Polish (2021-03-01, 1376989 articles), Turkish (2021-03-20, 381908 articles).

<sup>8</sup>Polish is also a gender-marking language. The difference might be that in Polish it is still common to use masculine professions when referring to women, even if feminine equivalents exist (Sosnowski and Satoła-Staśkowiak, 2019).

**Article tuples** Table 3 provides statistics for the final step of our exploration method, i.e. frequency of concept-related article tuples. The total number of discovered tuples is proportional to the initial Wikipedia sizes, ranging from 3k for Turkish up to 57k instances for English. Interestingly, only around one-third of them belongs to the symmetrical groups  $\bar{W}|M|G$  and  $\bar{W}|M$ , i.e., tuples that cover both Women and Men perspectives on the concept or topic in question. Moreover, these tuples cover only around half of all the filtered gender-related articles counted in Table 2 (between 46% for Turkish and 57% for Polish). When it comes to the asymmetrical  $\bar{W}|G$  and  $M|G$  groups the pattern is clear across all the languages – these tuples are much more frequent for Women, e.g., 4438 vs. only 155 cases for English. Finally, for the last two groups  $M$  and  $\bar{W}$  no clear pattern can be noticed at this stage; for two out of four languages  $M$  tuples are more frequent than  $\bar{W}$ .

### 3.2 Types of Title Inequalities

Table 4 presents examples of English articles that fell into each of the six groups and their meta-categories. Interestingly, our filtering method was able to find examples not only belonging to all the groups but also all the meta-categories (at this stage we leave out `Other` articles since they concern a variety of unrelated topics). We now investigate deeper each of the examples to see if we can notice any recurring patterns among them. We mark all discovered inequalities in red in Table 4.

**$\bar{W}|M|G$  and  $\bar{W}|M$**  All articles that belong to these two groups depict how symmetrical gender-related content looks like. In both cases, there are separate Wikipedia articles that describe women and men-related issues regarding the topic under question, such as *human sexuality* or *detective characters*. When necessary, an additional general article exists that describes the topic from a broader perspective. We mark all these examples in green.

**$\bar{W}|G$  and  $M|G$**  The next two groups show examples of inequalities among Wikipedia titles. We can notice that the source of the problem across the three meta-categories is different.

In the `Sports` and `Lists` meta-categories, the observed inequalities are a direct result of the decisions of Wikipedia editors. The `Sports` article *Finland national football team* is in fact an article about men’s team. Similarly, the article *Town Challenge Cup* refers to a women’s event. In both cases,

the general titles are missing information about the gender of the participants, and as a result, suggest that the article is about a general concept. In the previous section we found that the male generalization  $\bar{W}|G$  is much more frequent, which can be attributed to the *male generic bias*, i.e., the phenomenon in which the prototypical human is commonly assumed to be male (Silveira, 1980).

When it comes to the `Lists` meta-category, a slightly different type of inequality can be noticed. The article *List of Albanian writers* is in fact a proper general article, that covers both female and male writers. However, since the male counterpart does not exist, it makes the concept of *female writer* an exception in the general topic of *writers*. This pattern can be seen as a case of a more implicit gender bias, in which men and women are commonly presented with different levels of *linguistic abstraction* (Menegatti and Rubini, 2017).

Finally, the examples in `Social group` are strongly related to the *societal and biological asymmetries* that apply to people of different genders, such as history or health-related factors.

**$\bar{W}$  and  $M$**  Not all tuples belonging to these two groups are examples of title inequalities. `Names` consists of proper names that we treat from the beginning as symmetrical on their own and do not pair with other articles. `Sports` contains articles that similarly to `Names` refer to specific concepts, i.e., names of teams (*Ulster Senior League (men’s hockey)*) or events that were held only for men or women (*Danish Ladies Masters*). Therefore, we mark examples that belong to these two meta-categories in green in Table 4.

On the contrary, articles from `Lists` and `Social` meta-categories represent examples of asymmetrical Wikipedia content. Similar to the described above articles from  $\bar{W}|G$  and  $M|G$ , lack of male counterpart for the title *List of Danish women photographers* and female one for *List of male jazz singers* can be attributed to the decisions of the editors. Likewise, the inequalities in `Social` meta-category can be explained by historical and societal aspects.

### 3.3 Title Inequalities and Meta-Categories

Now that we have established what types of inequalities we can find in Wikipedia titles we investigate how frequent they are. We take only instances of asymmetrical tuples, i.e., only tuples that belong to the groups marked in red in Table 4, and plot their

		Women	Men	Generic
W M G	Sports	U Sports women’s soccer	U Sports men’s soccer	U Sports soccer
	Lists	List of women’s magazines	List of men’s magazines	List of magazines
	Social	Human female sexuality	Human male sexuality	Human sexuality
W M	Sports	Argentina women’s national softball team	Argentina men’s national softball team	–
	Lists	List of female detective characters	List of male detective characters	–
	Social	Bollywood Movie Award – Best Female Debut	Bollywood Movie Award – Best Male Debut	–
W G	Sports	Finland women’s national football team	–	Finland national football team
	Lists	List of Albanian women writers	–	List of Albanian writers
	Social	Women in Islam	–	Islam
M G	Sports	–	Town Challenge Cup (men)	Town Challenge Cup
	Lists	–	List of Thai representatives at international male beauty pageants	List of Thai representatives at international beauty pageants
	Social	–	Men’s health in Australia	Health in Australia
W	Sports	Danish Ladies Masters	–	–
	Lists	List of Danish women photographers	–	–
	Social	Violence against women in Guatemala	–	–
	Names	Four Ladies (TV series)	–	–
M	Sports	–	Ulster Senior League (men’s hockey)	–
	Lists	–	List of male jazz singers	–
	Social	–	Male Studies in the Caribbean	–
	Names	–	Anding Men station	–

Table 4: Examples of content-related tuples. We mark in green and red symmetrical and asymmetrical groups, respectively. Names appear only in the last two groups, because proper names are not paired with other articles.

frequency in Figure 2.

Interestingly, the picture is very similar across all the languages (we note that the scales on all four plots are different). Among the asymmetrical cases, the vast majority of tuples include `Women` articles, i.e., belongs to `W|G` or `W`. The largest group across all the languages is `W|G` and most of the articles that it contains are about sports. These are the titles such as *England women’s national football team* and *England national football team* that Criado Perez (2019) pointed out as potential inequalities in how women and men are portrayed in Wikipedia, that inspired our investigation. However, Figure 2 demonstrates that `Sports` is not the only topic that is asymmetrically covered in Wikipedia. The second most frequent meta-category is `Social`, especially in the `W` group, followed by `Lists`, that although can be noticed for all the groups, are much more frequent for women than for men (cf. `W|G` vs. `M|G` and `W` vs. `M`). The asymmetrical `M` titles within the Turkish Wikipedia seem prominent with respect to other languages’ distributions (although

note that it is only 25 instances). A closer inspection revealed that most cases are award pages that either are worded differently than women’s award pages or, although women’s awards exist, they are not represented in the Turkish Wikipedia.

#### 4 Beyond Captured Inequalities

Our methodology clearly illustrates that gender bias exists in Wikipedia titles in various patterns. Nevertheless, these patterns do not constitute an exhaustive list; more fine-grained mismatches could be captured if we extend our observations towards asymmetrically named titles and other Wikipedia components.

Before we move on with our observations, we define the structure of a Wikipedia article, exemplified in Figure 3. Aside from the self-explanatory title and content parts, there is a hatnote in some Wikipedia pages that is always placed at the very top of a page and that helps readers distinguish the page they are at, especially after a redirection or visiting a disambiguation page. On the left column

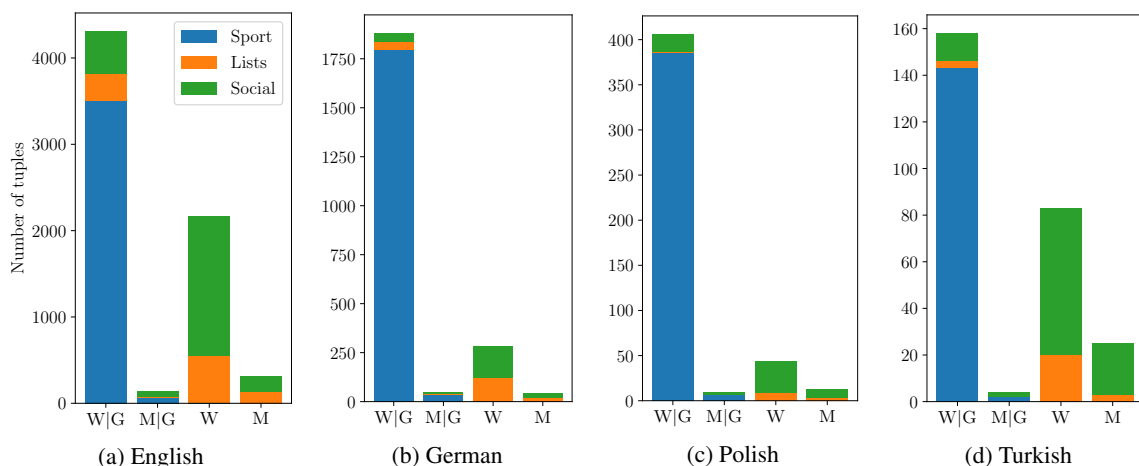


Figure 2: The frequency of asymmetrical article tuples, i.e., tuples marked in red in Table 4.

of a page, there are interwiki links that connect the page to its versions in other languages. Last but not least, categories are groupings of similar topics. Each page is part of at least one category. To keep category sizes and the number of categories per article manageable, the category structures are hierarchical.

**Lexically asymmetric tuples** Sometimes it is hard to find concept-related articles because the naming of symmetric concepts hence their Wikipedia titles are not aligned. Articles about the Spanish football league in the English Wikipedia represent such a case. The article about the male league is titled *La Liga*, which is the more commonly known version of the longer name *Campeonato Nacional de Liga de Primera División (Division National League Championship)*. The women’s article, on the other hand, is titled *Primera División (women)*, which is a shorter version of *Primera División de la Liga de Fútbol Femenino*. Our method recognizes the second title as an instance of *Women*, but it is not able to pair it with *La Liga* and the tuple falls into the symmetrical *W (Sports)* group instead of the asymmetrical *W|G* (see Table 4). It is not possible to match these two titles as long as there is a specific hatnote explaining the relationship between these two pages and linking them, neither for human readers nor for automatic systems.

**Semantically asymmetric tuples** In American collegiate athletics, a common word choice for representing women’s teams is adding *lady* to the name of men’s teams, e.g. *Statesmen vs. Lady Statesmen* of Delta State University (Eitzen and Zinn, 1993; Pelak, 2008). This phenomenon extends to professional sports such as *Professional Golfers’*



Figure 3: Structure of a Wikipedia article. 1: title, 2: hatnote, 3: content, 4: interwiki links, 5: categories. Example redirections to this article are *England football team* and *English football team*. The page [en.wikipedia.org/wiki/England\\_national\\_football\\_team](https://en.wikipedia.org/wiki/England_national_football_team) was accessed on 2021-04-26.

*Association vs. Ladies Professional Golf Association*. While choosing the word *lady* over *woman* is semantically charged (Lakoff, 1973), both words serve equally for our methodology and capture such article tuples as *W|G* type inequalities.

In Turkish, on the other hand, *lady* is used in the *Sports* category in a different way. The corresponding word *bayan* is used as an euphemism for *kadın (woman)*, as the latter is considered impolite

or has a sexual connotation from a sexist perspective (Arpınar-Avşar et al., 2016). This general use in language is carried over to national federations of Olympic sports. Arpınar-Avşar et al. (2016) report that 27 out of 34 federations use *bayan* to refer to female athletes as of 2016. It is not only in result sections or in regulations, but also in league and team names where male counterparts employ *erkek* (*man*). In terms of automatic processing, the problem is two-way. If we consider, for instance, *Erkekler Ligi* vs *Bayanlar Ligi* (*Men’s League* vs. *Ladies’ League*) as a concept-related pair, then seemingly there is symmetry; such tuples fall into the symmetric  $W|M$  group. If we do not pair these two titles, then they belong in  $M$  and  $W$  groups respectively, and as *Sports* articles are not treated as inequalities. In both cases, we do not capture the implicit semantic bias behind these two titles, i.e., bias related to historical and societal factors.

**Title-content mismatch** To remove inequalities in Wikipedia, editors commonly edit *Generic* titles so that they become *Men* titles (e.g. *National Team* → *Men’s National Team*). An additional approach is to redirect *Generic* titles to *Men* titles to facilitate search. Yet, if this procedure is done uncarefully and the content is not edited accordingly, it can create pages with title-content mismatch. For instance, searching for *Türkiye Millî Voleybol Takımı* (*Turkey National Volleyball Team*) redirects to *Türkiye Erkek Millî Voleybol Takımı* (*Turkey Men’s National Volleyball Team*). Since the *Women* article also exists, the tuple belongs to the  $W|M$  group thanks to editors; the titles are symmetrical. However, the content still uses the *Generic* reference: “*Türkiye Millî Voleybol Takımı ... Türkiye’yi uluslararası erkek voleybol karşılaşmalarında temsil eden takımdır.*” (“*Turkey National Volleyball Team ... is the team representing Turkey in the international men’s volleyball matches.*”). Therefore there is a mismatch between the title and content of the page.

**Cross-lingual mismatch** Interwiki links are beneficial in NLP tasks such as machine translation (Labaka et al., 2016), bilingual dictionary extraction (Tyers and Pienaar, 2008), or multilingual named entity recognition (Kim et al., 2012). When a title in one language is edited after interwiki links are established between languages, it might cause a cross-lingual mismatch. At the time of writing, the pages for the Turkish national basketball teams are

TR: Türkiye Millî Basketbol Takımı	
DE: Türkische Basketballnationalmannschaft	G+h
EN: Turkey Men’s National Basketball Team	
FR: Équipe de Turquie de basket-ball	
AZ: Türkiyə milli basketbol komandası	
DU: Turks basketbalteam	G
ES: Selección de baloncesto de Turquía	
PL: Reprezentacja Turcji w koszykówce mężczyzn	
RU: Мужская сборная Турции по баскетболу	M

Table 5: The titles and content properties of the Turkey men’s national basketball team in several languages. G+h: no *Men* in title, a hatnote explaining this is a *Men* page; G: no *Men* in title, no hatnote; M: *Men* in title.

a  $W|G$  tuple: *Türkiye kadın millî basketbol takımı – Türkiye millî basketbol takımı*. The *Generic* title also has the hatnote explaining that although the title is *Generic*, the page is about *Men*, and there is a separate *Women* article. However, interwiki links connect the page to its English counterpart which is titled *Turkey men’s national basketball team*, that is, *men* indicator is explicit only on the English side. The *Generic* to *Men* mapping in titles depicts a cross-lingual mismatch. NLP tasks that would use this pair would fail to extract an exact translation. Following multiple interwiki links shows that approaches vary across languages, as summarized in Table 5. Moreover, cross-lingual asymmetries in this article can be observed not only in its title but also in its content. The English page reads *The Turkey men’s national basketball team* (*Turkish: Türkiye Millî Basketbol Takımı*) . . . , causing a *Generic* vs. *Men* inconsistency between English and Turkish names.

Finally, cross-lingual mismatches come also with semantic asymmetries, such as the use of *bayan* (*lady*) instead of *kadın* (*woman*). Interwiki still links them to non-Turkish titles with the word *woman* or its equivalent, causing a translation mistake.

## 5 Discussion: Towards Debiasing Titles

The types of inequalities that we presented in this paper call for different debiasing measures. Lexically asymmetric tuples can only be paired via mutual hatnotes. In the specific case of *La Liga*, adding (*men*) to the title in parallel to its female counterpart would at least prevent the male generic bias. Title-content mismatches could be avoided if the content could be updated together with the title. Cross-lingual mismatches are the easiest to catch



as there are Wiki bots designed for that purpose.<sup>9</sup>

Among meta-categories, debiasing the `Sports` category is seemingly the simplest one. For instance, the Turkish Basketball Federation<sup>10</sup> symmetrically names national basketball teams by explicitly using *kadın* and *erkek* in their titles. The `W|G` inequality comes from Wikipedia editors. Inserting *erkek* to the `Generic Türkiye millî basketbol takımı` (and updating the content consequently) would bring balance to this article tuple. However, sometimes the imbalance is actually in the *real world*. The International Cycling Union (UCI)<sup>11</sup> defines both cups and teams as `W|G` tuples, e.g., *UCI Road World Cup* vs. *UCI Women's Road World Cup*, and Wikipedia titles follow suit. After all, it is not *wrong* to take the actual name of a concrete entity. The next step is subject to a debate: Should Wikipedia editors continue to reflect the real world or should they already convert the title to `Men` and keep the existing `Generic` naming, perhaps with a redirection for search purposes?

The `Lists` category with a `W|G` or `M|G` inequality can be balanced by introducing the missing gender-specific articles or going into the complete opposite direction and keeping only the `Generic` page. If the answer to this question is to create gender-specific pages, the non-binary gender values come into the picture to be represented in separate `Lists` pages. Such lists exist, e.g. *List of transgender political office-holders*, yet, they are few. Wikipedia's metadata system facilitates such a listing by providing non-binary values for annotating biographies; currently, it is a set of seven: male, female, non-binary, intersex, transgender female, transgender male, agender.<sup>12</sup> However, page specification raises more questions: How about the people who do not want to be identified with one of the values, where should they be placed?

A substantial amount of the `W` only tuples in the `Social` category consists of women's rights movements. 10% of the `Social` titles in the English Wikipedia contains the words *suffrage* or *rights*. While `W` seems to be an imbalanced category by definition, the reason behind it is not that women are dominating the scene. On the contrary, it shows the reaction to suppression and that women had to fight over rights that should be the default. Simi-

<sup>9</sup>[meta.wikimedia.org/wiki/Interwiki\\_bot](https://meta.wikimedia.org/wiki/Interwiki_bot)

<sup>10</sup>[www.tbf.org.tr](http://www.tbf.org.tr)

<sup>11</sup>[www.uci.org](http://www.uci.org)

<sup>12</sup>P21 'sex or gender' property, [www.wikidata.org/wiki/Property:P21](http://www.wikidata.org/wiki/Property:P21)

larly, `W|G` tuples such as *Women in Engineering* vs. *Engineering* seem to be biased towards women by definition, yet the urge to discuss such topics separately comes from the complete opposite reason. Men have been considered the 'norm' such that women are a deviating subset.

It is hard to develop an umbrella strategy for debiasing the `Social` category. For some of these titles, the question is whether they should be *debaised* rather than how they could be debaised. In our opinion, for instance suffrage pages, e.g., *Women's suffrage in Alabama* should remain as is, but the `W|G` tuple *Women's Suffrage* vs. *Suffrage* can be extended to a `W|M|G` tuple. At the time of writing, the title *Men's Suffrage* redirects to *Suffrage*. Instead, a separate `Men` page that very briefly explains the historical developments (i.e., why there has been no need to coin an explicit term for men) would be more informative and more in line with Wikipedia's encyclopedic nature.

## 6 Conclusion

Inequality in male and female Wikipedia biographies has driven a lot of attention in recent years, both from the research community as well as the press.<sup>13</sup> The editors' community is consequently tackling this problem and fighting the gender gap. As recently shown by [Schmahl et al. \(2020\)](#), these efforts pay off and not only more biographies about women are being added, but also NLP models such as word embeddings trained on Wikipedia articles are exhibiting less stereotypical biases.

In this paper, we aimed at raising awareness that gender inequalities in Wikipedia extend beyond biographies. Women and men are systematically represented in article titles in a different way, especially in such domains as sports and social issues. We showed that such inequalities can be computationally assessed by investigating asymmetrical tuples, i.e., titles that describe a particular topic or concept for only one gender.

The topics we argue here by no means cover all possible inequalities in Wikipedia titles, let alone inequalities in its overall components. Yet, even a simple but competent, systematic approach is strong enough to demonstrate the existing gender bias. With this paper, we hope to draw attention to such bias and open possible solutions to discussion.

<sup>13</sup>See [en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Women\\_in\\_Red/About\\_us](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Women_in_Red/About_us) for a list of recent news articles regarding this subject.

## Acknowledgements

We thank Pınar Arpınar-Avşar for pointing out real-world sport title inequalities. The second author is funded by DFG via project CE 326/1-1 “Computational Structural Analysis of German-Turkish Code-Switching” (SAGT).

## References

- Pınar Arpınar-Avşar, Serkan Girgin, and Nefise Bulgu. 2016. [Lady or woman? The debate on lexical choice for describing females in sport in the Turkish language](#). *International Review for the Sociology of Sport*, 51(2):178–200.
- Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. [Language \(Technology\) is Power: A Critical Survey of “Bias” in NLP](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online. Association for Computational Linguistics.
- Friederike Braun. 2001. [The communication of gender in Turkish](#). *Gender across languages: The linguistic representation of women and men*, 1:283–310.
- Ewa S. Callahan and Susan C. Herring. 2011. [Cultural Bias in Wikipedia Content on Famous Persons](#). *Journal of the American society for information science and technology*, 62(10):1899–1915.
- Benjamin Collier and Julia Bear. 2012. [Conflict, Confidence, or Criticism: An Empirical Examination of the Gender Gap in Wikipedia Contributions](#). In *CSCW ’12 Computer Supported Cooperative Work, Seattle, WA, USA, February 11-15, 2012*, pages 383–392. ACM.
- Caroline Criado Perez. 2019. *Invisible women: Exposing data bias in a world designed for men*. Random House.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- D. Stanley Eitzen and Maxine Baca Zinn. 1993. [The Sexist Naming of Collegiate Athletic Teams and Resistance to Change](#). *Journal of Sport and Social Issues*, 17(1):34–41.
- Batya Friedman and Helen Nissenbaum. 1996. [Bias in Computer Systems](#). *ACM Trans. Inf. Syst.*, 14(3):330–347.
- Andrew Gaut, Tony Sun, Shirlyn Tang, Yuxin Huang, Jing Qian, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang, and William Yang Wang. 2020. [Towards Understanding Gender Bias in Relation Extraction](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2943–2953, Online. Association for Computational Linguistics.
- M. Hellinger and H. Bussmann. 2001. *Gender Across Languages: The linguistic representation of women and men. Volume 1*. John Benjamins Publishing Co.
- M. Hellinger and H. Bussmann. 2003. *Gender Across Languages: The linguistic representation of women and men. Volume 3*. John Benjamins Publishing Co.
- Sungchul Kim, Kristina Toutanova, and Hwanjo Yu. 2012. [Multilingual Named Entity Recognition using Parallel Data and Metadata from Wikipedia](#). In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 694–702, Jeju Island, Korea. Association for Computational Linguistics.
- Daniël de Kok. 2014. [TüBa-D/W: a large dependency treebank for German](#). In *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13)*, page 271.
- Piotr Konieczny and Maximilian Klein. 2018. [Gender gap through time and space: A journey through Wikipedia biographies via the Wikidata Human Gender Indicator](#). *New Media Soc.*, 20(12).
- Gorka Labaka, Iñaki Alegria, and Kepa Sarasola. 2016. [Domain adaptation in MT using titles in Wikipedia as a parallel corpus: Resources and evaluation](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 2209–2213, Portorož, Slovenia. European Language Resources Association (ELRA).
- Robin Lakoff. 1973. [Language and Woman’s Place](#). *Language in Society*, 2(1):45–79.
- Shyong K. Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R. Musicant, Loren G. Terveen, and John Riedl. 2011. [WP:Clubhouse? An Exploration of Wikipedia’s Gender Imbalance](#). In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration, 2011, Mountain View, CA, USA, October 3-5, 2011*, pages 1–10. ACM.
- Michela Menegatti and Monica Rubini. 2017. [Gender Bias and Sexism in Language](#). *Oxford Research Encyclopedia of Communication*.
- Cynthia Fabrizio Pelak. 2008. [The Relationship Between Sexist Naming Practices and Athletic Opportunities at Colleges and Universities in the Southern United States](#). *Sociology of Education*, 81(2):189–210.
- Joseph Reagle and Lauren Rhue. 2011. [Gender bias in Wikipedia and Britannica](#). *International Journal of Communication*, 5:21.

- Christina Richards, Walter Pierre Bouman, Leighton Seal, Meg John Barker, Timo O Nieder, and Guy T'Sjoen. 2016. [Non-binary or genderqueer genders](#). *International Review of Psychiatry*, 28(1):95–102.
- Rachel Rudinger, Jason Naradowsky, Brian Leonard, and Benjamin Van Durme. 2018. [Gender Bias in Coreference Resolution](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 8–14, New Orleans, Louisiana. Association for Computational Linguistics.
- Katja Geertruida Schmahl, Tom Julian Viering, Stavros Makrodimitis, Arman Naseri Jahfari, and Marco Tax, David andj Loog. 2020. [Is Wikipedia succeeding in reducing gender bias? assessing changes in gender bias in Wikipedia using word embeddings](#). In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, pages 94–103, Online. Association for Computational Linguistics.
- Deven Santosh Shah, H. Andrew Schwartz, and Dirk Hovy. 2020. [Predictive Biases in Natural Language Processing Models: A Conceptual Framework and Overview](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5248–5264, Online. Association for Computational Linguistics.
- Jeanette Silveira. 1980. Generic masculine words and thinking. *Women's Studies International Quarterly*, 3(2-3):165–178.
- Wojciech Paweł Sosnowski and Joanna Satoła-Staškowiak. 2019. A contrastive analysis of feminitives in Bulgarian, Polish and Russian. *Cognitive Studies| Études cognitives*, (19).
- Dagmar Stahlberg, Friederike Braun, Lisa Irmen, and Sabine Sczesny. 2007. Representation of the sexes in language. *Social communication*, pages 163–187.
- Tony Sun, Andrew Gaut, Shirlyn Tang, Yuxin Huang, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang, and William Yang Wang. 2019. [Mitigating Gender Bias in Natural Language Processing: Literature Review](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1630–1640, Florence, Italy. Association for Computational Linguistics.
- Rachael Tatman. 2017. [Gender and Dialect Bias in YouTube's Automatic Captions](#). In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 53–59, Valencia, Spain. Association for Computational Linguistics.
- Francis M Tyers and Jacques A Pienaar. 2008. Extracting bilingual word pairs from Wikipedia. *Collaboration: interoperability between people in the creation of language resources for less-resourced languages*, 19:19–22.
- Eduard Fosch Villaronga, Adam Poulsen, Roger Andre Søråa, and B. H. M. Custers. 2021. [A little bird told me your gender: Gender inferences in social media](#). *Inf. Process. Manag.*, 58(3):102541.
- Claudia Wagner, David García, Mohsen Jadidi, and Markus Strohmaier. 2015. [It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia](#). In *Proceedings of the Ninth International Conference on Web and Social Media, ICWSM 2015, University of Oxford, Oxford, UK, May 26-29, 2015*, pages 454–463. AAAI Press.
- Claudia Wagner, Eduardo Graells-Garrido, David García, and Filippo Menczer. 2016. [Women through the glass ceiling: gender asymmetries in Wikipedia](#). *EPJ Data Sci.*, 5(1):5.
- Kellie Webster, Marta R. Costa-jussà, Christian Hardmeier, and Will Radford. 2019. [Gendered Ambiguous Pronoun \(GAP\) Shared Task at the Gender Bias in NLP Workshop 2019](#). In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, pages 1–7, Florence, Italy. Association for Computational Linguistics.
- Kellie Webster, Marta Recasens, Vera Axelrod, and Jason Baldrige. 2018. [Mind the GAP: A Balanced Corpus of Gendered Ambiguous Pronouns](#). *Transactions of the Association for Computational Linguistics*, 6:605–617.
- Amir Zeldes. 2017. [The GUM Corpus: Creating Multilayer Resources in the Classroom](#). *Language Resources and Evaluation*, 51(3):581–612.
- Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. 2018. [Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 15–20, New Orleans, Louisiana. Association for Computational Linguistics.