

Using search engines to trace metalinguistic discussions of the noun *female* in digital discourse

Marie Flesch

Université de Lorraine / Nancy, France
marie.flesch@univ-lorraine.fr

Abstract

This contribution uses the web as corpus to investigate metalinguistic discourses of the noun “female” between 2001 and 2019. It presents the construction of a diachronic dataset using search engines and compares results provided by Google, DuckDuckGo and Yandex, and, for Google, searches made with four different combinations of parameters. It then uses the resulting corpus to create a repertoire of arguments and ideologies used by online critics of “female”. Results show that, even though Google generated the most relevant results by far, supplementing it with other search engines was beneficial. They also suggest that, by 2015, the metalinguistic discourse of “female” had mostly solidified, with no new arguments emerging in the corpus, and that the debates are a response to new (at the time) and broader social tensions, specifically the rise of masculinist discourses in online spaces.

1 Introduction

Since the 1970s, feminist scholars have been uncovering the asymmetries in references to women and men (for example, the tendency to refer to men as “men” and to women as “girls”, Lakoff, 1975; Sigley and Holmes, 2002), and the process of semantic pejoration undergone by many terms referring to women, seen in pairs like “master and mistress” (Schulz, 1990). Metalinguistic debates about terms for women long predate second wave feminism, however, and have unfolded in dictionaries, style books, and usage guides. One such debate revolves around the noun “female”, which fell out of favour in the 19th century, coming to be seen as “vulgar”, “contemptuous” (Hall, 1917) and “impolite” (Fowler, 1926, p. 175). Yet, when second wave feminists began reclaiming the word “woman”, their efforts primarily targeted the words “lady” and “girl” rather than “female”; for example, Robin Lakoff (1975) does not discuss “female”

and even uses it as a noun to refer to women. Many academic works about language and gender, including some discussing gender-inclusive guidelines, continued to use the noun in reference to women in the first decades of the 21st century (for example, Hellinger, 2010; Mills and Mullany, 2011; Baker, 2014 — as well as my own work, for ex. Flesch and Bailly, 2020). Recently, however, the debate around “female” seems to have picked up, as evidenced by newer editions of some style guides, such as the 2022-2024 edition of *The Associated Press Stylebook* and the 2024 edition of *The Chicago Manual of Style*, which now note the negative connotations of the nouns “female” and “male” and recommend to not use them in reference to humans (Froke et al., 2022; The University of Chicago Press Editorial Staff, 2024).

In attempting to retrace the developments of the renewed discussion about the noun “female”, I found few traces of it in academic works, in feminist theory and beyond (for example in Salleh, 1992; Frye, 2005; Peters et al., 2023; Montell, 2024). My hypothesis is that the debate mostly re-emerged from “below”, in non-academic discussions held on the internet, precisely because the noun “female” has been repoliticized online through its use by incels and other men rights activist movements, and more broadly in reaction to the development of a “networked misogyny” (Banet-Weiser and Miltner, 2016). A notable example of these grassroots efforts is the Reddit forum r/MenAndFemales, which is dedicated to calling out the use of the noun and frequently features incel content (JenJens, 2026).

This is why, in order to retrace the trajectory of the “female as a noun” debate in the 21st century, I turned to the web as corpus. Three search engines (Google Search, Yandex, and DuckDuckGo) were used to search for the expression “female as a noun” and to build a diachronic corpus of metalinguistic discussions in blogs, discussion forums,

social media platforms, media articles, and comments sections. Using search engines to create such a corpus comes with obvious limitations, notably the lack of reproducibility, the instability and inconsistency of search results, and the opacity of the algorithms used (Kehoe, 2020). However, it also has advantages: it is more lightweight than using the massive Common Crawl or Reddit archives, does not require knowledge of programming languages, and provides access to discussions in their context, allowing for more situated linguistic analyses. Moreover, it provides access to content that predates Common Crawl archives, which (as of the time of this writing) are accessible only for the year 2013 and onwards¹. While the corpus makes no claim to exhaustivity or representativity, it constitutes a valuable roadmap of the recent developments of the “female vs. woman” debate.

The contribution this paper makes is twofold:

1. It presents a comparison of the results of searches using the exact expression “female as a noun” provided by three search engines (Google, Duckduckgo, Yandex) and, for Google, using combinations of different search parameters: with and without “omitted” results; global search (all times) and searches in custom ranges of times. The goal is to establish whether, in order to get as many quality results as possible for the study at hand (specifically for older web data), it is beneficial to (a) complement Google with other search engines and (b) combine different types of Google searches. This could be useful to researchers who want to conduct diachronic studies of linguistic or sociocultural phenomena using the web as corpus.
2. It investigates metalinguistic discourse in the resulting corpus (available on OSF²), looking at the ways internet users frame “female” as a noun: it uncovers the repertoire of arguments used by critics and the social categories associated with the usage of the term. Results suggest that, by 2015, the metalinguistic discourse had already solidified. They also reveal that it is not only a response to evolving norms in language use, but also to broader social tensions, particularly the rise of the manosphere.

¹<https://data.commoncrawl.org/crawl-data/CC-MAIN-2012/index.html>

²https://osf.io/hq62x/overview?view_only=ec480386076144d48e21147b4d10b44b

2 Related work

2.1 Search engines

Research has shown that results provided by various search engines can differ significantly. Because of its overwhelming market dominance, Google usually outperforms other search engines in terms of relevance, ranking and coverage of results (Hajian Hoseinabadi and CheshmehSohrabi, 2024). However, search engines are not simply entry points to the internet. Because it is also a content and platform provider, with services like Google Books, YouTube or Google Maps, Google tends to give priority to its own services and collections in its search results (Lewandowski, 2023). It has also been shown that, when looking specifically for Reddit results, Google disproportionately excludes discussions related to politics, pornography, bots, and cryptocurrency, acting as a “digital gatekeeper” that determines what information is accessible and shaping online discourse (Poudel et al., 2025).

The shortcomings of the “web as corpus” approach, which emerged in the late 1990s, have been pointed out by corpus linguists. Because the actual size of the internet is unknown, it cannot be considered a “corpus” and cannot be used to create representative samples (Sinclair, 2005). The other issue, when using search engines, is that of reproducibility: it is made impossible by the frequent changes in algorithms and by the “bubble effect”, i.e. the fact that search engines tailor their results based on derived users’ preferences (Ćurković and Košec, 2018). Difficulties in using search engines to create corpora is compounded by the decrease in options offered by search engines, which makes them less and less suitable for the extraction of linguistic data (Kehoe, 2020). The shift toward AI-driven search could further reduce the reliability of search engines (Memon and West, 2024).

These biases have serious implications for any research that uses solely search engines as a corpus building tool. There are alternatives, such as using Common Crawl web archives. However, using them is problematic: it requires access to big data infrastructure and skills, documents are not preprocessed, and the archives contain many near duplicates of pages (Panchenko et al., 2018). For the present study, which aims to retrace the developments of a metalinguistic discourse as far back in time as possible, the fact that Common Crawl only provides content crawled since 2013 constitutes an additional limitation.

2.2 The “female” as a noun debate

Metalinguistic debates around “female” have been documented by John Lesslie Hall (1917) and Dennis E. Baron (1986). They are largely restricted to the noun, while the adjective “seems to have no enemies” (Hall, 1917, p. 89). In use as a synonym for “woman” since the 14th century, “female” began to meet some opposition by writers in the 18th century, and then came under attack in the 19th century because of its negative connotation. The writer Sarah Josepha Hale, who carried the first campaign against the noun in the American magazine *Godey’s Lady’s Book*, noted that it “reduce[s] [women] to the level of animality” (LC, 2018). Other arguments have been put forward in dictionaries and style guides: it was described as “impolite” Fowler (1926, p. 175), “contemptuous” (Evans and Evans, 1957, p. 560), “clinical” (Miller and Swift, 1980, p. 76), “problematic” (American Psychological Association, 1995, p. 55), and retaining a “sense of non-humanness” (Doyle, 1995, p. 31). However, in reclaiming the word “woman”, second wave feminists mostly focused on “lady” and “girl” (cf Lakoff, 1975). The debate has recently been reignited, with William Safire (2007) writing in *The New York Times* that “feminists everywhere have begun to turn on the word female”. In the same article, Deborah Tannen notes that many feminists avoid the term because it “connotes a biological category”. Amanda Montell also argues that the woman vs. female debate is fuelled by the “gender versus sex question” (2024), while Deborah Cameron turned to the British National Corpus to show that “female” as a noun is often used by men to make disparaging comments on women (Cameron, 2016). Recent corpus linguistics studies suggest that the noun is undergoing a process of pejoration (Robinson, 2018; Ferley, 2022; Joensuu, 2023).

However, it seems that most of the metalinguistic discourse about “female” is not taking place in academic spaces, but in digital discourse, in forums, blogs, comment sections, and social media. For example, it has been the main focus of the subreddit r/MenAndFemales since its creation in 2020. To the best of my knowledge, no study has investigated systematically the recent developments of the “female” debate: this contribution aims to address this gap, by using the web as corpus.

3 Search engines comparison

3.1 The search expression

The goal was to use an expression that would most likely provide results containing metalanguage, and not return too many false positives (general discussions about language or just simple usage of the term). From my ongoing work about the topic, “female as a noun” emerged as a phrase that is frequently used in discussions about the word. For example, it is the title of one chapter in Hall (1917)’s English usage guide; and it features in the first rule of the r/menandfemales subreddit (“‘Female’ as an adjective is okay. ‘Female’ as a noun is not”³). The quotation marks search operator was used in order to force exact-match searches.

3.2 The search engines

Search engines were selected on the basis of their popularity. According to statcounter (2026), the five most used search engines are Google Search with a market share of 91.31%, Microsoft Bing (market share: 4.47%), Yahoo! Search (market share: 1.24%), Yandex Search (market share: 1%) and DuckDuckGo (market share: 0.65%). However, Bing and Yahoo! had to be discarded because initial searches showed that using the quotation marks search operator did not work, returning results that contained both “female” and “noun” but not necessarily “female as a noun”. The fact that both search engines behave in the same way is not surprising, seeing that Yahoo! Search uses Microsoft Bing to generate results (BBC News, 2009). The way searches were performed varied depending on the affordances of each search engine. Because the Russian search engine Yandex Search does not allow for time-specific searches, a simple global search was conducted. For DuckDuckGo, a search engine created by an independent software company that focuses on internet privacy, custom time ranges were defined: 2000-2010, 2011-2015, 2016-2020, 2021-2022, and 2022-2026. Google also offers a date range parameter; and, depending on the results of a search, it provides the option to display “omitted results”, i.e., results that are not considered the most relevant. Given Google’s dominance as a search engine and the fact that, in exploratory searches, it was clear that it returned a lot more results than Yandex Search and DuckDuckGo, I decided to compare results provided by

³<https://www.reddit.com/r/MenAndFemales/>

different combinations of parameters: (1) global results (without setting a date range), only showing the default results; (2) global results (without setting a date range) including the omitted results; (3) results obtained using a custom date range, only with the default results; and (4) results obtained using a custom date range, including omitted results. The date range was set as January 1 to December 31 of each year starting in 2015. Before that date, there were simply too few results (sometimes none). The earliest date range was thus set to January 1, 2000 to December 31, 2014. No omitted results were available for this period. For all search engines, the region was set to “Any Region” and the language to “English”.

3.3 Extraction of URLs

All searches were conducted in Nancy (France) on July 19th 2026, with the Chrome web browser. URLs and their metadata (name of website, title of page, snippet of its content, etc.) were retrieved using the Instant Data Scraper browser extension⁴, an automated data extraction tool operated by Flavr Technology that extracts structured data from web pages and exports it in a tabular form, as .csv or .xlsx files. The tool offers a crawling function that extracts several pages of results. However, there were some issues: the crawling was sometimes interrupted; and it was simply impossible to use with Yandex Search (the search results were simply copied and pasted in a .csv file).

3.4 Annotation of search results

The various .csv files generated by the Instant Data Scraper extension were combined in R (version 4.5.1). Overall, the searches generated 2309 search results; there is a lot of overlap, as was to be expected, with only 1604 unique URLs. Less expected, maybe, is the extent of the difference between the number of unique hits returned by DuckDuckGo (n=90) and Yandex (n=56) on one hand, and Google on the other (n=1512). The results were then manually annotated to assess their quality. The first step of the annotation process consisted in determining the date of each page. For Reddit and Quora pages, which account for a large portion of the results (see Section 3.7), this involved specific measures. Since dates of posts are not visible on Reddit’s default interface, the old interface (“Old Reddit”) was viewed by replacing “www” by “old”

in URLs. For Quora, it was necessary to create an account and to add “/log” at the end of URLs to view a post (or “question”)’s full history, including its publication date and answers. In some rare cases, such as dictionary entries and encyclopedia articles, no date was available.

Then, each page was inspected to determine if it was relevant, i.e., (1) if it contained the phrase “female as a noun” and (2) the phrase was used in a metalinguistic comment or discussion. This process was far from straightforward. For example, for Facebook results, it was often necessary to scroll far down a comment section to view the relevant comment(s), if they existed. In Reddit discussion threads, comments are sometimes hidden and must be “loaded” by clicking on a link, or can be viewed by accessing the pages in JSON format by appending “.json” to the end of URLs. Given the tediousness of the task, I decided to restrict the analysis presented in this paper to pages published until 2019 included (283 unique URLs). While this leaves out the major portion of the dataset, it still allowed me to meet the main objective of the study by focusing on earlier discussions.

All results considered not relevant didn’t contain the search expression “female as a noun” – or, the search expression did not appear in the main discussion or article, but somewhere else on the page (for example, in “related posts” or “related topics” on discussion forums). This was frequent in Reddit and Quora results. The search expression always featured in metalinguistic discussions; thus, all pages that contained “female as a noun” were annotated as relevant. 167 results were annotated as relevant and 117 as not relevant.

3.5 Comparison between search engines

Combined, the four types of Google searches provided the bulk of relevant results (150/167 pages). Yandex only returned 20 relevant pages, and DuckDuckGo 16. Of the 16 pages not detected by Google, 8 were returned by Yandex and 8 by DuckDuckGo, with no overlap between the two search engines. 12 of these 16 pages not detected by Google were discussion threads posted on discussion forums, mostly Reddit but also smaller forums. Figure 1 represents the distribution of results by year; it only shows the results detected by each search engine and not by the others (i.e., the “Yandex” category shows the results returned by Yandex but not by Google and DuckDuckGo). It shows that very few results, or no result at all, were returned

⁴<https://chromewebstore.google.com/detail/instant-data-scraper/ofaokhiedipichpaobibbnahnkdoiiah?hl=en>

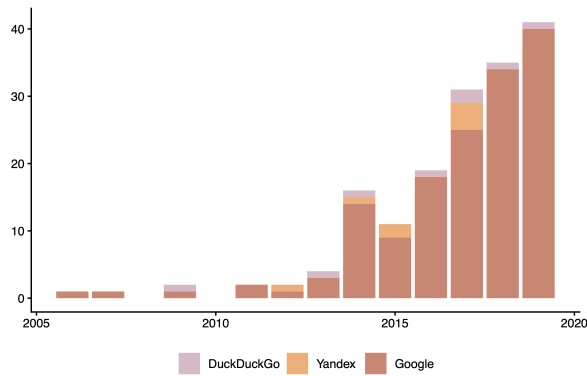


Figure 1: Distribution of relevant results detected exclusively by each search engine (pre-adjustments made in section 4.1)

for the years 2005-2013, and that most of these were returned by Google searches.

3.6 Comparison between Google searches

Google searches are compared for the years 2015 to 2019 (because a single combination of parameters was used for years prior to 2015). By far, the searches performed with a yearly time range and including omitted results returned the most results ($n=93$), followed by the global search with omitted results ($n=43$). Searches that didn't include omitted results yielded fewer results ($n=35$ with time range, $n=39$ without).

3.7 Composition of the dataset

The dataset is heterogeneous, with pages from 53 websites (social media, blogs, the Q&A website Quora, forum, media and academic websites). Forums, which are the earliest sources, constitute about half of the dataset (84 pages, or 50.3%). Most forum pages come from Reddit (55 pages, 32.93% of dataset). Quora (25 pages, 14.97%) and Facebook (16 pages, 9.58%) are the other two major sources of data. The distribution of data sources per year (Figure 2) reflects sociocultural shifts in internet use, with the decrease in blog content and the rise of social media. Despite Twitter popularity, only two results originate from the microblogging website, which may be due to recent changes affecting Twitter/X pages' indexing by Google (Peters, 2023).

4 Metalinguistic discourse analysis

4.1 Adjustments made to the dataset

The dataset analysed in this section differs slightly from the dataset presented above. Some results

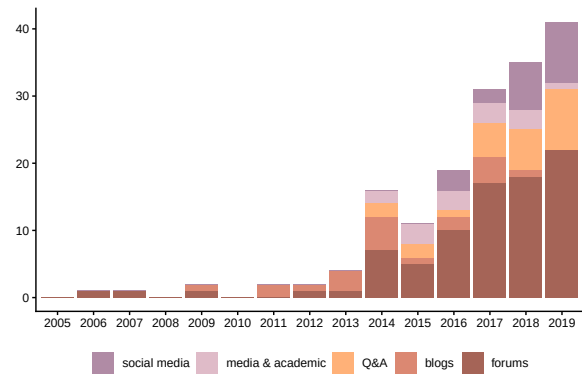


Figure 2: Composition of the dataset (post-adjustments)

considered relevant had to be excluded because they were actually duplicates: for example, in forums, “female as a noun” sometimes appeared on several different pages (and thus URLs) of a single discussion thread. Here, several pages of a same discussion thread are considered a single observation. In one instance, a discussion thread spanned two different time intervals (the years 2006 and 2012); since the study focus on the diachronic developments of the “female” metalinguistic discourse, I considered the sets of comments posted during these two years as two different observations. In one case, an interview was published on two different websites; I only retained one occurrence. Only textual content was included in the dataset, meaning that two videos (posted on Facebook and Instagram) were excluded. Furthermore, the dataset was expanded by performing searches within websites using their search function when available; and by including web pages linked by internet users in forums and articles, some of which no longer existed and had to be retrieved using the Internet Archive Wayback Machine⁵. Overall, 12 new web pages were added to the dataset. The final dataset contains 167 observations, dated 2001 to 2019. It is available on OSF ([link](#)); it contains links to pages, websites names, dates, and annotations (but no verbatim comments).

4.2 The framing repertoire

The framing repertoire of “female” as a noun was constituted by manual annotation of metalinguistic comments by a single annotator (the author). Because the goal was to identify the set of arguments used rather than counting their frequency on each page, each discursive space was treated as a

⁵<https://web.archive.org/>

whole, with arguments found in a discussion thread or comment section documented only once, even when there were repeated by several people. Two types of framing were explored:

- The arguments mobilised by internet users to criticise and reject the usage of “female” as a noun.
- The social categories attached to the use of the noun; or, in other words, according to critics, who typically use “female” as a noun?

4.2.1 Arguments

Arguments were identified through a two-step process, beginning with the extraction of qualifiers, and then the coding of these qualifiers into distinct argumentative categories. 16 different arguments were identified; they are presented in Table 1, with examples (paraphrased, as well as all excerpts from the corpus cited in the paper, so that authors are not identifiable), the frequency of each category, and the year it first appears in the corpus. Some arguments are semantically quite similar (e.g. “de-humanising” and “for animals”; “derogatory” and “misogynistic”); they were retained as separate categories to capture the nuances of metalinguistic discourse. Because of the non-exhaustive and biased nature of the corpus, the exact date or origin of each argument in online discourse can obviously not be determined. However, what Table 1 shows is that, between 2015 and 2019, and despite the much larger number of pages in that period, no new arguments were introduced in the corpus. This suggests that, by 2015 (and at least until 2019), the metalinguistic debate had solidified and possibly reached a point of saturation.

4.2.2 Social categories

Descriptors referring to “typical” users of “female” as a noun, according to the authors in the corpus, were manually extracted in statements like “It’s often used by incels which is disgusting” [Reddit, 2018] or “it makes you sound like Ferengi”. [The Straight Dope Message Board, 2015]. Nuances were preserved, specifically in descriptors referring to misogynistic communities or movements, such as pick-up artists, MGTOW (Men going to their own way), MRAs (Men rights activists), red-pillers, etc. Table 2 shows the descriptors that appear at least twice in the corpus, with their overall frequency and year of first mention. Descriptors referring to misogynistic communities (incels, MRAs,

etc.) or individuals (misogynists, anti-feminists), and figures of hegemonic masculinities (bros, neckbeards, nerds/geeks, “nice guys”; cf [Ensmenger, 2015](#)), highlighted in bold, constitute the majority of mentions (96/162, or 59.26%). “Ferengi” refers to the extraterrestrial species depicted in the American science fiction franchise *Star Trek*, who are characterised by their extremely capitalistic and sexist social norms and who systematically refer to women as “females”. Not only is “Ferengi” the most frequent and earliest descriptor, but metalinguistic comments also frequently feature images of Ferengi or references to their pronunciation of the word: “They objectify us. We’re just ‘Feeeeeeemaaaaales’. (Picture a Ferengi saying it)” [Reddit, 2013]. This suggests that references to Ferengi have become a staple of metalinguistic discourse of the noun “female”, which is supported by the fact that the subreddit *r/MenAndFemales* (created in 2020 and devoted to calling out uses of “female” as a noun) features an image of a Ferengi as its profile image.

The use of “female” by the other social categories presented in Table 2 is not always seen as a derogatory. Metalinguistic commentary is often nuanced, recognising the importance of context: “In a scientific conversation [...] then sure, “female” as noun is fine. But used in a casual conversation, as it often is these days, it carries an underlying implication of otherness.” [Reddit, 2017]. It also acknowledges other types of sociolinguistic variation linked to professional or linguistic background: “If the speaker isn’t a native English speaker, I’m more inclined to give them the benefit of the doubt.” [Quora, 2018]; “If he has a military or law enforcement background, I tend to give him the benefit of the doubt.” [The Straight Dope Message Board, 2012].

Finally, like with arguments, the analysis shows that, despite the limited size of the corpus before 2014, the majority of descriptors are already present by that date: ideologies connected to the use of “female” as a noun are already crystallized in metalinguistic discourse.

5 Conclusions

This study used three search engines as points of entry to the web as corpus, in order to constitute and analyse a corpus of metalinguistic comments about the noun “female”. The comparison between results provided by Google, Duck-

Argument	Examples	1st mention	Freq.
for animals	A kangaroo of the fairer sex is a female, but a human with two X chromosomes is a girl or a woman. [The Straight Dope Message Board, 2001]	2001	60
asymmetry	Men and Women is okay. Males and Females too. Men and Females... stupid [Resetera forum, 2017]	2001	43
weird	Referring to a woman using female as a noun is just weird [Ask Metafilter, 2007]	2001	37
dehumanising	As a rule of thumb, I never use “male” or “female” as a noun when describing people because “man” or “woman” is a lot less dehumanizing. [Reddit, 2019]	2006	79
derogatory	“female” is never followed with a positive comment [Resetera forum, 2017]	2006	52
sex vs. gender	For humans, sex and gender are two distinct concepts. A person can identify as a woman without being biologically female, and vice versa. [Reddit, 2014]	2006	16
misogynistic	I think using “female” as a noun to refer to women should be banned. [...]. Guys who do that are usually just misogynists. [The Snark Theater blog, 2018]	2009	33
a slur	It seems like OP is using the word “female” because she can’t say “slut.” [Ask A Manager comment section, 2014]	2009	9
reduces women to reproductive functions	I think that ‘female’ is outdated. Also, it reduces women to their reproductive capacity in contexts where it plays little to no role. [Media Diversity Institute website, 2019]	2011	25
a red flag	I have no idea why, but people using female as a noun sends up red flags in my mind. [Quora, 2019]	2011	18
distancing	I feel like they do this to distance themselves from us, to define women as the “other” [Reddit, 2017]	2012	40
grammatically incorrect	Female is an adjective, not a noun. Woman is a noun. [Reddit, 2018]	2012	32
transphobic	Not all women are biologically female, and equating “female” with “woman” invalidates gender-nonconforming and trans individuals. [Buzzfeed, 2014]	2013	15
a dog whistle	Now it’s a bit of a dog-whistle. When a man uses it, it’s often to disparage women [Reddit, 2014]	2014	1
a sign of discomfort with “woman”	Using “female” as a noun also suggests discomfort with the word “woman” as if it were indecent or inappropriate.. [Reddit, 2018]	2015	4
racist	In my experience people who use “females” in a derogatory way only reinforce the misogynoir that Black women face. [The Straight Dope Message Board, 2015]	2015	3

Table 1: Arguments used by critics of the noun “female”

DuckGo, and Yandex shows, unsurprisingly, that Google vastly outperforms other search engines in terms of number of relevant results returned.

However, the few results provided by Duck- DuckGo and Yandex (and not by Google) were valuable for the present study; specifically, they

Descriptors	1 st mention	Freq.
Ferengi	2006	27
men terrible with women	2007	3
law enforcement	2011	13
misogynists	2011	15
the military	2011	24
bro	2012	2
medical/science professionals	2012	18
MRAs	2012	9
neckbeards	2012	7
nerds/geeks	2012	4
weird guys/douchebags	2012	3
Black men	2013	6
pick-up artists	2014	5
Hispanic men	2015	2
incels	2017	13
“nice guys”	2018	3
anti-feminists	2018	3
ESL speakers	2018	3
MGTOW	2018	2

Table 2: Social categories associated with the use of “female” as a noun

included results from smaller forums (like city-data.com, english.stackexchange.com, italki.com and neogaf.com), which helped diversify the corpus. Similarly, the comparison between various Google searches showed that it can be useful to use various combinations of parameters in order to get as many relevant results as possible. Choosing the option “repeat the search with the omitted results included” and setting a time range improved the performance. One-year intervals were used as the time range and it is possible that using shorter time spans would have yielded even more results.

While the corpus created provides a partial and biased view because of the limitations of search engines and of the methodology used, its analysis provides new insights on the “female” debate. First, it shows that the new “female” debate builds on arguments put forward by earlier critics of the noun: its dehumanising and derogatory connotations, and the animal comparison. When allowances are made, for example in medical and scientific discourse, it is only under the condition that usage is symmetric (“males and females”), a requirement also specified by style guides (for ex., [Miller and Swift, 1980](#); [Froke et al., 2022](#); [American Psychological Association, 2010](#)).

Other arguments (“female” is racist, transphobic, and does not refer to gender) reveal the influence of feminist and transgender activism. The corpus also contains newer arguments, which go beyond noting pejorative connotations: some internet users see “female” as a slur, not synonymous with “woman” but with “bitch” or “slut”. The “red flag” and the “dog whistle” arguments implicitly tie the use of the noun to misogynistic individuals and communities. The analysis of “typical” users of “female” shows that the pejoration of the term (and possible increase in usage) is connected to the masculinist ideologies of the manosphere. The linguistic argument (that “female” is an adjective and not a noun, and thus using it as a noun is grammatically incorrect) is also new. It is sometimes countered with references to dictionary entries. Some users invert the critique, arguing that widespread misuse of “female” has led dictionaries to include the noun: “The misuse of this word has become so widespread that some dictionaries now list “noun” as a secondary meaning. That doesn’t make it acceptable. Stop. This. Right now. . “[Facebook, 2019]. The prescriptive force of this new norm exemplifies what Deborah Cameron (2012) terms “verbal hygiene”, a form of metalinguistic discourse that, while nominally about grammar, reflects “deeper anxieties which are not linguistic, but social, moral and political” (p. vii). Finally, looking diachronically at the corpus reveals that, despite the small number of pages in the earlier years (2001-2015), by 2015, the framing repertoire of the debate has mostly been constituted, with no new arguments emerging. It is possible, however, that new arguments and ideological stances have emerged since 2019. Exploring this will be the focus of my future research.

Limitations

The nature of search engines makes this research unreproducible and the corpus fundamentally biased: it is impossible to know how many pages were filtered by search engines, and why. Other limitations are the facts that using search engines do not give access to websites that have disappeared and pages or comments that have been deleted (unlike archives), and that the dataset was annotated by a single annotator. Moreover, the search expression “female as a noun” was deliberately chosen for its limited scope, with the intent to minimise false positives. This means that metalinguistic discus-

sions which do not include the expression “female as a noun” were not included in the corpus; it also means that the relatively limited number of search engine page results made it possible to manually examine all pages published in a 19-year period, which may not be realistic for studies focusing on more frequent words or expressions.

Acknowledgments

This research was funded by the French National Research Agency as part of the project “ANR-24-AERC-0011-01”.

References

- American Psychological Association. 1995. *Publication manual of the American Psychological Association*, 4th ed edition. Washington, DC.
- American Psychological Association. 2010. *Publication manual of the American Psychological Association*, 6th ed edition. Washington, DC.
- Paul Baker. 2014. *Using corpora to analyze gender*. Bloomsbury, London.
- Sarah Banet-Weiser and Kate M Miltner. 2016. [Masculinitysofragile: Culture, structure, and networked misogyny](#). *Feminist media studies*, 16(1):171–174.
- Dennis E Baron. 1986. *Grammar and gender*. Yale University Press.
- BBC News. 2009. [Microsoft and Yahoo seal web deal](#).
- Deborah Cameron. 2012. *Verbal hygiene*. Routledge, London; New York.
- Deborah Cameron. 2016. [To gender or not to gender? \(thoughts prompted by the death of Zaha Hadid\)](#).
- Margaret Doyle. 1995. *The A-Z of non-sexist language*. The Women’s Press.
- Nathan Ensmenger. 2015. [“Beards, sandals, and other signs of rugged individualism”: masculine culture within the computing professions](#). *Osiris*, 30(1):38–65.
- Bergen Evans and Cor Evans. 1957. *A dictionary of contemporary American usage*. Random House, Inc.
- Andrew Ferley. 2022. [The semantic derogation of female](#). Master’s thesis, York University.
- Marie Flesch and Sophie Bailly. 2020. *Mapping the itineraries and interests of cisgender and transgender internet users with the RedditGender corpus*, page 93–112. L’Harmattan.
- Henry Watson Fowler. 1926. *A dictionary of modern English usage*. Clarendon Press, Oxford.
- Paula Froke, Anna Jo Bratton, Andale Gross, Jeff McMillan, Pia Sarkar, Jerry Schwartz, and Raghuram Vadarevu. 2022. *The Associated Press stylebook 2022-2024*, 56 edition. The Associated Press, New York.
- Marilyn Frye. 2005. Categories in distress. In Barbara S. Andrew, Jean Keller, and Lisa H. Schwartzman, editors, *Feminist interventions in ethics and politics: feminist ethics and social theory*, page 41–58. Rowman Littlefield Publishers.
- Azadeh Hajian Hoseinabadi and Mehrdad CheshmehSohrabi. 2024. Proposing a new combined indicator for measuring search engine performance and evaluating Google, Yahoo, DuckDuckGo, and Bing search engines based on combined indicator. *Journal of Librarianship and Information Science*, 56(1):178–197.
- John Lesslie Hall. 1917. *English usage: Studies in the history and uses of English words and phrases*. Scott, Foresman.
- Marlis Hellinger. 2010. Guidelines for non-discriminatory language use. In Ruth Wodak, Paul E Kerswill, and Barbara Johnstone, editors, *The SAGE handbook of sociolinguistics*, page 565–582. SAGE Publications. Publisher: Sage.
- JenJens. 2026. [Is the incel website content too much?](#)
- Santeri Joensuu. 2023. [The plural “females” and its derogatory use. A corpus study on the sexist aspects in the use of the word](#). Master’s thesis, Tampere University.
- Andrew Kehoe. 2020. Web corpora. In Magali Paquot and Stefan Th. Gries, editors, *A practical handbook of corpus linguistics*, page 329–352. Springer.
- Robin Tolmach Lakoff. 1975. *Language and woman’s place*. Harper Row.
- LC. 2018. [“The best human benefactor of WOMAN”, Matthew Vassar and Sarah Joseph Hale](#).
- Dirk Lewandowski. 2023. *Understanding search engines*. Springer.
- Shahan Ali Memon and Jevin D West. 2024. [Search engines post-ChatGPT: How generative artificial intelligence could make search less reliable](#). *arXiv preprint arXiv:2402.11707*.
- Casey Miller and Kate Swift. 1980. *The handbook of nonsexist writing*, 1st edition. Lippincott Crowell, New York.
- Sara Mills and Louise Mullany. 2011. *Language, gender and feminism: Theory, methodology and practice*. Routledge.
- Amanda Montell. 2024. *Wordslut: a feminist guide to taking back the English language*. Taylor Francis.

- Alexander Panchenko, Eugen Ruppert, Stefano Faralli, Simone Paolo Ponzetto, and Chris Biemann. 2018. Building a web-scale dependency-parsed corpus from commoncrawl. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*.
- Jay Peters. 2023. [Tweets aren't showing up in Google results as often because of changes at Twitter](#).
- Jessica R Peters, Allison Stumper, Katja M Schmalenberger, Andy J Taubman, and Tory A Eisenlohr-Moul. 2023. [Improving rigor through gender inclusivity in reproductive psychiatric science](#). *Comprehensive Psychoneuroendocrinology*, 16:100194.
- Amrit Poudel, Yifan Ding, Tim Weninger, and Jürgen Pfeffer. 2025. [Digital gatekeepers: Google's role in curating hashtags and subreddits](#). In *Proceedings of the 63rd annual meeting of the Association for Computational Linguistics (volume 1: Long papers)*, page 2402–2415.
- Melissa Aubrey Robinson. 2018. *A man needs a female like a fish needs a lobotomy: The role of adjectival nominalization in pejorative meaning*. Master of arts thesis, University of North Texas.
- William Safire. 2007. Language: Woman vs. female. *The New York Times*.
- Ariel Salleh. 1992. The ecofeminism/deep ecology debate. *Environmental Ethics*, 14(3):195–216.
- Muriel R. Schulz. 1990. The semantic derogation of women. In Deborah Cameron, editor, *The feminist critique of language: A reader*, 1st edition, page 135–147. Routledge.
- Robert Sigley and Janet Holmes. 2002. [Looking at girls in corpora of english](#). *Journal of English Linguistics*, 30(2):138–157.
- John Sinclair. 2005. *Corpus and text - Basic principles*. Oxbow Books, Oxford.
- statcounter. 2026. [Search engine market share world-wide](#).
- The University of Chicago Press Editorial Staff. 2024. *The Chicago Manual of Style*, 18 edition.
- Marko Ćurković and Andro Košec. 2018. [Bubble effect: including internet search engines in systematic reviews introduces selection bias and impedes scientific reproducibility](#). *BMC medical research methodology*, 18(1):130.