

Gender Gaps and Stereotypes in the Long Run

A computational approach to how *Le Monde* got (slightly) demasculinized (1944-2024)

Benoît de Courson^a, Ange Richard^b Gilles Bastin^c

^a Max Planck Institute for the Study of Crime, Security and Law, Freiburg im Breisgau, Germany
(b.decourson@csl.mpg.de)

^b Univ. Grenoble Alpes, PACTE, LIG, 38000 Grenoble, France (ange.richard@univ-grenoble-alpes.fr)

^c Univ. Grenoble Alpes, CNRS, Sciences Po Grenoble - UGA, Pacte, 38000 Grenoble, France.
(gilles.bastin@iepg.fr)

Abstract

A substantial body of research has demonstrated that news media often present a skewed portrayal of society, particularly with respect to gender. This distorted picture manifests in two main ways: a persistent gap in the exposure of men and women in media content, and the perpetuation of entrenched gender stereotypes in the same content. However, the long-term evolution of these gender disparities in the media remains largely unexplored, primarily due to the lack of accessible, diachronic corpora and the methodological divide between the study of exposure gaps and that of stereotypes. This article addresses these challenges by (a) introducing new, generalizable measures to assess gender disparities in media representation, covering both exposure and stereotypes, and (b) applying these measures to a comprehensive dataset of 3.2 million articles, 13.9 million mentions, and 3.7 million citations from *Le Monde*, France's leading highbrow newspaper, spanning the years 1944 to 2024. According to our results, while men still dominate the content produced by *Le Monde*, their share of mentions has decreased from 94% to 74%, and their share of quotes from 98% to 76%. This decline in the masculinity of news has accelerated since 2016 across most news sections except Sports, and in particular in Society and Culture articles. The rise of female journalists in *Le Monde*'s newsroom (reaching 47% at the end of the studied period) contributes to this trend, as women journalists tend to mention and quote more women than male journalists, increasingly so since 2010. However, despite these shifts, our results show that the stereotyping of both men and women mentioned or quoted in *Le Monde*, which had steadily decreased

from 1945 to 2010, has remained stable since then. This suggests that the recent decrease in the content masculinity has not resulted in more equal treatment of men and women in *Le Monde*.

Introduction

The gender gap in the media is a well-documented and enduring bias favoring men across various aspects of media representation. Extensive research demonstrates that this bias, though gradually diminishing, remains widespread across countries, media types, and topics. Gender disparity manifests itself in many ways in the media: not only are women mentioned significantly less often than men in media content, they are also quoted less often than mentioned. Furthermore, when women and men do appear in the media, they are often portrayed through the distorting lens of gender stereotypes, which skews public perception of gender issues and ultimately contributes to the perpetuation of gender inequalities.

How these biases have evolved in the long run remains largely unknown due to limited data and the challenges of effective measurement. A first major obstacle in this research field is that copyright restrictions and the reluctance of media organizations to provide access to their content make it difficult to compile a media corpus suitable for long-term computational analysis. While national libraries, such as Gallica in France, have recently digitized older content, usable corpora remain scarce for the period following World War II due to copyright protections. Fortunately, a few legacy newspapers have digitized their archives and offer access behind a paywall. In France, *Le Monde* has done so, providing access to its full text since its founding in December 1944.

Measuring the diverse dimensions of gender bias in media content using a unified approach is a second central methodological challenge in long-term gender gap research. Gender biases occur at multiple levels such as mentioning and quoting practices – which have traditionally been examined using quantitative research methods – as well as more subtle writing routines resulting in gendered stereotypes, typically examined with qualitative research methods in the literature. Consequently, the relationships between these dimensions of gender gaps in media content are seldom studied together.

In this paper, we address this methodological challenge using a comprehensive corpus of 3.1 million articles from *Le Monde*. We propose a tridimensional approach to gender disparity, measuring: (i) **mentions masculinity**—the overrepresentation of men among those mentioned in the newspaper; (ii) **quotes masculinity**—the overrepresentation of men among those quoted; and (iii) **stereotype level**—the level of gendered patterns in verbs used to introduce quotations, such as cases where a journalist writes that a woman “whispers” (“*murmure*”) while a man “asserts” (“*déclare*”). We believe that this integrated approach allows us to capture a fuller picture of how gender biases in a major newspaper such as *Le Monde* have evolved over time.

Overall, our study confirms a general decline in mention, quotation, and stereotyping gaps over time. However, our longitudinal analysis reveals a surprisingly non-linear evolution. While mentions and quotes gaps remained largely stable—at very high levels of male representation—until the 1970s, they have decreased at an accelerating pace since then. In contrast, the stereotyping gap has shown a more steady decline since the 1940s. In this paper, we argue that these trends can be understood as the combined result of editorial choices made by Le Monde, a significant shift in the gender composition of its newsroom, as well as changes in how women journalists have positioned themselves in relation to the dominant masculinity norm in journalism.

Related work

Gender inequality in media content has been extensively documented in social science research. Since the 1970s, studies have consistently revealed both quantitative disparities—such as the underrepresentation of women—and qualitative differences, particularly gender stereotypes. While these findings vary across media types, countries, and time periods, they broadly support Gaye Tuchman's pioneering concept of the "symbolic annihilation" of women in modern media (Tuchman, 1978).

(a) Measuring gender representation

Early studies, primarily focused on American newspapers, revealed significant gender imbalances in the late 1970s and early 1980s. For instance, women accounted for less than 10% of individuals mentioned in major news outlets during that period (Davis, 1982; Potter, 1985). Jolliffe's (1989) analysis of New York Times coverage in 1885 and 1985 found that, although gender stereotypes declined over time, representation gaps remained unchanged (Jolliffe, 1989). Similarly, Armstrong (2004) found that in 1999, men were still three to four times more likely than women to be featured in 18 leading U.S. newspapers (Armstrong, 2004).

Many studies on media gender gaps continue to rely on content analysis of samples from national media sources (Kinnick, 1998; Ross et al., 2012; Mitchelstein et al., 2019). Using a similar methodology on a broader scale, the Global Media Monitoring Project (GMMP) has been conducted every five years since 1995 across more than 100 countries. Its findings indicate that, following a decrease in gender imbalances and associated stereotypes during the 1990s and early 2000s, a surprising stabilization occurred in the subsequent years. The proportion of women represented in the surveyed media increased from 17% in 1995 to 24% in 2010 ; however, this figure has not significantly changed since then (Macharia, 2021).

Much of the research on measuring gender imbalances is still largely conducted through manual content analysis methods applied to aggregated media corpora at the national level (Kinnick, 1998; Mitchelstein et al., 2019; Ross et al., 2013; D'Heer et al., 2020). This approach has several limitations. First, it is highly dependent on the period selected for data collection, which can introduce temporal bias due to unforeseen events occurring during that time. Additionally, for practical reasons, the sources analyzed are often restricted to prominent media outlets and specific sections deemed most important, which may not provide a comprehensive view of gender representation across all media. Furthermore, the quality

of manual coding, especially when conducted by large groups of coders, can fluctuate, leading to inconsistencies in data interpretation and categorization. Lastly, many of these manual methods rely on aggregated indicators that often fail to clearly distinguish between individuals mentioned (the subjects of news) and those quoted (the sources of news). This lack of precision can obscure the true nature of representation in media.

Increased access to large media content datasets and advancements in Natural Language Processing (NLP) — particularly Named Entity Recognition systems — have recently facilitated the development of computational methods to measure gender inequalities in media. Jia et al. (2016) observed, for example, in a corpus of over 2.3 million online English press articles, that a person's name mentioned in an article has between a 69.5% chance (in Entertainment) and a 91.5% chance (in Sports) of belonging to a man. In their analysis of a large corpus of U.S. newspaper articles published between 1983 and 2008, Shor et al. (2015) identified a « paper ceiling » for women in media content (Shor et al., 2015). Their findings revealed that while the proportion of women mentioned in these articles increased from less than 20% to 27% during this period, the rate of change was notably slow.

Furthermore, in another study, they examined a sample of 20,000 individuals across various professions in the United States and confirmed a significant representation bias — especially in politics — after controlling for factors such as age, social position, and the public visibility of the mentioned individuals on Wikipedia (Shor et al., 2019). To explain this bias, the authors proposed a macro-sociological factor : the stagnation of various indicators reflecting the status of women in society, which had not shown significant progress throughout the 2010s.

(b) Measuring gender stereotypes

The study of gendered media stereotyping has evolved along a different methodological trajectory than research on gendered exposure imbalances. While some broad content analysis approaches spanning various types of media and news beats have incorporated elements of stereotype analysis—such as the Global Media Monitoring Project, which examines gendered roles like authority and agency (Ross et al., 2018; Meadel and Coulomb-Gully, 2011)—research on stereotypes has typically focused on particular media, beats or social groups.

Research on the media coverage of women and men in politics has been particularly insightful in that respect. Overall, studies focused on national-level elections have consistently shown significant gender imbalances (Coulomb-Gully, 2009; Falk, 2010; Heldman et al., 2018). This gender imbalance is frequently coupled with a stereotyping that disadvantages women, as evidenced in various election contexts, including the 2010 British elections (Bystrom et al., 2001; Falk, 2010; Ross et al., 2012). Furthermore, a strong association between gender imbalance and stereotyping has been observed even outside election cycles, particularly in media coverage of U.S. Congress members (Niven, 2005). However, the findings of existing research are not always consistent, as highlighted in a recent meta-analysis of 90 studies covering 750,000 media stories (Van der Pas and Aaldering, 2020).

Gender stereotypes have also been explored in specific media domains, such as women's magazines (Kitch, 2001; Blandin, 2011), advertisement (Goffman, 1979; Grau and Zotos, 2016; Shinoda et al., 2021) or

online news platforms (D'Heer et al., 2019). These studies have largely relied on content analysis, ranging from manual thematic content analysis (Collins, 2011) to critical discourse analysis (Sriwimon and Zilli, 2017) and more recently to topic modeling (Kozlowski et al., 2020). However, this more issue-specific approach to gender inequality in the media presents two major limitations: the corpora examined are often restricted in both size and scope, and there is no clear method for comparing shifts in exposure imbalances with changes in stereotyping.

Recent studies on gender stereotyping have increasingly adopted NLP-based methods, often analyzing the outputs of models trained on news data to observe if these models "learn" gender biases from the underlying text. For instance, Leavy employs machine learning to reveal gender bias embedded in media coverage of Irish Prime Ministers. After training a classifier to predict whether an article was written about a female or male politician, Leavy extracts the key features identified by the model to uncover traces of gendered language in the discourse (Leavy, 2020). In another study, Garg et al. use word embeddings trained on *Google News* and *The New York Times* corpora, measuring gender bias by examining the proximity of vectors related to gender-specific terms and job titles in a projected space. Their findings suggest that stereotypes have diminished over time, corresponding with women's increased labor market participation (Garg et al., 2018; Jones et al., 2020b). This approach has also certain limitations: results, for instance, can be highly sensitive to the choice of words used in vector comparisons, and word embeddings generally offer limited explainability due to their inner complexity.

Research on gender stereotyping reveals that women, in addition to being significantly underrepresented compared to men, are more likely to appear in contexts related to specific themes such as care, entertainment, and motherhood. Women are also assigned distinct roles compared to men, often portrayed in passive positions, while men are more frequently depicted in roles of authority and agency. Furthermore, discourse about women often emphasizes their femininity, emotional states, or sexualized physical attributes (Collins, 2011)

(c) Explaining the gaps

Several types of variables have been used in the literature to explain gender inequalities in media. At the individual level, the most commonly used variable is the gender of the article author or of the editorial decision-makers in the media. This variable is often used to test the hypothesis that male journalists, who have historically outnumbered women in journalism and remain more likely to hold positions of authority within the news media, produce content that favors men. Research in this area has produced mixed results. Armstrong's study of American national and local newspapers finds that female bylines are a strong predictor of gender diversity among those mentioned in the articles (Armstrong, 2004). Similarly, studies on Chilean media (Leiva and Kimber, 2022) and on Austrian media (Riedl et al., 2024). Other studies have shown that female journalists tend to include a greater diversity of sources in terms of gender and ethnicity and to write in less stereotyped ways (Rodgers and Thorson, 2003). However, other research has found that the presence of women in leadership positions does not necessarily reduce gender stereotyping in content (Jolliffe and Catlett, 1994; Meeks, 2024). Similarly, Shor et al. observe a positive effect of female representation in newsrooms on coverage polarity, but do not find a significant impact when women hold senior positions (Shor et al., 2019).

These mixed results underscore the importance of considering organizational factors. Studies based on interviews with female journalists and observations of newsroom practices highlight the role of power structures in newsrooms. First, newsroom feminization is often limited: women in newsrooms still tend to have fewer responsibilities and are frequently assigned to specific types of media or news topics. They also face greater job insecurity than men, as documented in studies on French political journalism (Damian-Gaillard and Saitta, 2011, 2020). This situation is comparable to other contexts where gender stereotyping of female journalists and asymmetrical power over content are pronounced (Djerf-Pierre and Löfgren-Nilsson, 2004). These implicit norms and power dynamics counterbalance efforts by generations of female journalists, often more attuned to feminist issues, to adopt less gendered working standards. Yet, these journalists often find themselves isolated within existing power structures (Byerly, 2004; Djerf-Pierre, 2005). In sum, the gendered substructures of newsrooms temper the effects of feminization on producing less biased content (De Bruin, 2000; De Bruin and Ross, 2004; Melin-Higgins, 2004; Silveirinha et al., 2024).

Finally, some research has focused on how topic selection and editorial treatment contribute to inequalities in content. The GMMP study shows that women are more likely to be mentioned as experts in articles dealing with traditionally “feminine” topics, such as education and social issues, and are notably absent, as experts, from topics like politics, economics, or security (Macharia, 2015). In another study, Jia et al. observe that in a corpus of 2.3 million articles from the English press, names mentioned in the “Entertainment” section are 69.5% male, compared to 91.5% male in the “Sports” section (Jia et al., 2016). More broadly, this literature has shown that the professional norms of journalism have historically constructed “newsworthiness” as a masculine domain—the most “newsworthy” topics (Gans, 1979) are those that primarily involve men, such as politics and international affairs—accompanied by a parallel disqualification and essentialization of women’s voices in media content (Damian-Gaillard et al., 2021).

Limited research has examined the role of structural and external factors in shaping gender inequalities in media content. Examples of these underexplored factors include media ownership structures and reliance on public subsidies (Sjøvaag and Pedersen, 2019). Additionally, political party policies on gender parity and candidate nominations are strong predictors of the restricted media visibility of female candidates (Lühiste and Banducci, 2016).

In light of previous work in the field, we adopt in this paper an approach that emphasizes the need for consistent, comparable methods to measure gender gaps and stereotypes, facilitating a clear analysis of their evolution over time. Prior studies often examine isolated topics across diverse sources, but our research, to the contrary, focuses on great diversity of topics in a single source. This choice allows us to maintain analytical consistency and deepen our data interpretation by leveraging an in-depth understanding of this source’s history. By covering all sections of the newspaper, we aim to provide a more exhaustive perspective on gender representation, complementing earlier findings with a more holistic approach. Furthermore, responding to methodological challenges highlighted in prior literature, we utilize straightforward NLP tools. This approach avoids the complexities and potential biases of advanced machine learning models, promoting transparency and interpretability as we track long-term trends.

Data and Methods

(a) *Le Monde* corpus

This study leverages the complete archives of the French highbrow daily newspaper *Le Monde*, covering all publications since its founding in December 1944 until July 2024. The dataset includes 3.2 million articles, and approximately 1.6 billion words, spanning the entire period, albeit with uneven distribution across years, particularly in the recent years (Fig. S1). The corpus was compiled via web scraping as part of the *Gallicagram* project,¹ a lexicometric tool similar to Google's Ngram Viewer applied to French press archives (Azoulay and de Courson, 2021; de Courson et al., 2023).

In addition to content produced by journalists for the print edition of *Le Monde*, the dataset includes *Le Monde*'s weekly supplements (such as *Le Monde des livres* and *M le Mag*), as well as web-only content like short articles and live coverage, which have increased in volume since the late 1990s. However, articles published before 2006 contain incomplete metadata, particularly lacking information on the section (e.g., "sports," "politics") to which each article belongs. To address this, we used the large language model Hermes-2-Pro-Llama-3-8B to classify all articles into one of eight predefined sections using zero-shot classification (see below). In comparison to a human annotation of 1,000 articles, the LLM achieved 80% accuracy, with the remaining 20% primarily involving debatable cases, as news sections tend to overlap. Importantly, this allows us to tell content about French politics or the French society apart from international content.

The journalist's name appears in 40% of the articles retrieved using our methodology. According to an in-house *Le Monde* documentalist, the remaining 60% consists primarily of brief dispatches from news agencies and short stories that were indeed unsigned in the original newspaper. Additionally, this portion includes articles published between 1987–1994 and 2003–2004, periods affected by known digitization issues at *Le Monde*. To infer the likely gender of articles' authors from their byline, we identified first names where possible, successfully categorizing 31.9% of all articles. On average, signed articles tend to be about twice as long as unsigned ones; thus, this approach allows us to identify the gender of authors for 50% of the total count of words published by *Le Monde* since its inception.

Beyond providing insights into gender gaps in non-American media, choosing *Le Monde* offers several advantages. First, its digitized archives, available since 1944, enable the creation of a large and continuous corpus spanning 80 years of comparable media content. Second, *Le Monde*'s status as France's leading newspaper since its founding—and its significant influence on French society—makes the analysis of gender representation in its content particularly relevant. Third, *Le Monde* has consistently emphasized comprehensive news coverage since its founding, delivering in-depth reporting not only on politics and international affairs—its traditional strengths and the most prestigious sections of the newsroom—but also on economics, culture, and societal issues (Eveno, 2004).

Fourth, despite *Le Monde*'s longstanding editorial policy emphasizing journalistic independence, an elite-oriented news hierarchy, and a commitment to austere reporting (Padioleau, 1985; Eveno, 2001), the newspaper has undergone notable changes since the end of the 1960's, often described as an increased

¹ <https://shiny.ens-paris-saclay.fr/app/gallicagram>

openness to broader societal concerns.² Along with evolving economic pressures, these adjustments reflected shifts in readership and societal expectations (Saitta, 2005). New sections and topics, such as sports and entertainment, were introduced, and supplements were launched to appeal to a broader and more diverse audience, particularly in terms of gender. As a result, the proportion of women among the newspaper's readership increased from 36% to 42.5% between 1961 and 1969, while the share of readers aged 15–24 grew from 21% to 30%.

These efforts to broaden the audience continued into later decades, particularly with the revamped format introduced in the 1990s. The primacy of political news diminished, with the political service being renamed the "France" section in 1995 reflecting a gradual shift in political news from extensive reporting on policy debates to more horse race-style coverage focused on the political game (Kaciaz, 2019). Additionally, the content strategy was revised to include shorter articles, an increased use of narrative journalism (e.g., reportage, profiles), and a greater emphasis on photography.

(b) Measuring gender gaps and stereotypes

By applying NLP-based methods to this corpus, we aim to provide a detailed and updated understanding of how gender gaps and stereotypes are expressed in legacy media content and how they have evolved from the second half of the 20th century to the first two decades of the 21st century. To achieve this, we introduce four distinct metrics to assess the extent of gender gaps and stereotypes in *Le Monde* content. These metrics, deriving from our previous work (Bastin, 2022; Richard et al., 2022), are designed to capture multiple dimensions of gender representation and stereotyping, using linguistic patterns that may reflect broader societal gender dynamics.

Our first metric, the **Mentions Masculinity**³, measures the proportion of men among all people mentioned in an article. This metric is generated by identifying people mentioned in each article using a Named Entity Recognition (NER) algorithm. Each identified entity is then assigned a probability of being male, based on its first word, which is typically a first name — which we gendered using a very large database of first names given to boys and girls in France throughout the XXth century⁴ — or a gendered title.⁵ The Mentions Masculinity, calculated for each article, represents the average of these probabilities for all the named persons recognized by the algorithm. Thus, a Mentions Masculinity of 1 indicates that 100% of the mentions are of men, while a theoretical Mentions Masculinity of 0.5 would indicate equal mentions of men and women.⁶

Our second metric is the proportion of men among all the people quoted in the article — i.e. those who are given a voice in the narrative. We refer to this as the **Quotes Masculinity**. This metric is calculated

2

https://www.lemonde.fr/archives/article/2002/06/03/jacques-fauvet-l-homme-qui-a-ouvert-le-monde-sur-la-societe_278535_1819218.html

³ Here, "masculinity" does not refer to the qualities typically associated with being a man in a given society, nor to the performance of masculine gender roles in everyday interactions (Connell and Messerschmidt, 2005). Instead, it solely refers to journalists' practices of mentioning and quoting individuals in their stories. Thus, it does not involve any assessment of the masculinity of those mentioned as a set of attributes or behaviors.

⁴ <https://www.insee.fr/fr/statistiques/8205621?sommaire=8205628>

⁵ We included the most common titles related to civility, military rank, nobility, and religious status that appeared as the first words of recognized entities with a clearly identifiable gender: M./Mme, général/générale, prince(ss), Mgr (the French title for a bishop), cardinal, Sir, Lord/Lady, comte/comtesse, Père/Soeur, and Abbé.

⁶ We apply the same approach to assign a gender to the author of each article based on their first name.

by first identifying all quotes in the articles using an algorithm that recognizes both direct and indirect quotes by applying a set of syntactic rules (Soumah et al., 2023), along with their cue verbs (the verbs that introduce the quotes) and sources (the quoted speakers). To determine the proportion of men among quoted people, we inferred the gender of the identified sources using various indicators : titles, gender-specific first names (from multiple gendered name dictionaries), and/or gendered job titles (e.g., "directrice," "directeur"). The Quotes Masculinity of an article refers to the percentage of quotes attributed to men out of all gender-specific quotes extracted from the article. We also compare the length of quotes between male and female sources to assess potential cumulative discrimination based on the amount of space allocated to men and women in the newspaper.

Our third metric – the **Gender Segregation Level** – measures the concentration of women in articles. To calculate this metric, we first identify articles that feature at least one woman. Next, we determine the gender ratio of the other individuals mentioned or quoted in these articles. We then compare this ratio to the overall gender ratio in the corpus (for the same section and year). This calculation provides a measure of women's overexposure to other women in *Le Monde's* articles. A value of 1 indicates an independent distribution, while higher values suggest a concentration of women in a limited number of articles.

Finally, we use our quote extraction algorithm to develop a fourth metric — the **Quotes Stereotype Level** — that measures gender stereotyping of male and female sources in media content. This approach leverages a distinctive characteristic of French journalism, particularly at *Le Monde*: the frequent use of diverse verbs to introduce quotes.⁷ Specifically, we calculate a gender odds ratio for each verb by comparing the likelihood of it being used to introduce a male versus a female speaker. This enables us to classify verbs based on their odds ratios and thereby identify typical gendered patterns and stereotypes in journalistic language. In addition to identifying which verbs are typically used to attribute quotes to men versus women, this method allows us to measure the standard deviation of the odds ratios, which we interpret as an indicator of the level of gender stereotyping in the articles. The higher the standard deviation, the more strongly specific verbs are associated with one gender. Conversely, a lower standard deviation indicates that these verbs are used more interchangeably when introducing statements from sources of both genders (see Appendix for details).

Thus, our hypothesis challenges the prevailing view among journalists — expressed for instance in widely used French journalism textbooks (Martin-Lagardette, 2009; Agnès, 2002) — that the choice of verbs is determined solely by stylistic factors, such as avoiding repetition, conveying the tone of the source or creating variety. Instead, we propose instead that verb choice is influenced by gender stereotypes, with certain verbs systematically more likely to introduce quotes from men or women. The metric derived from these odds ratios, which we refer to as Quotes Stereotype Gap, measures the variance in odds ratios between verbs typically associated with male and female sources. A higher Quotes Stereotype Gap indicates a higher degree of gender stereotyping in the way journalists present men and women to newspaper readers.

We do not interpret these metrics as entirely separate dimensions of the gender gap in media content but rather as an attempt to disentangle journalistic implicit gender norms from the structural constraints of

⁷ The list is available at https://github.com/gillesbastin/old_fashion_nlp.

reporting in a gender-unequal world. Specifically, we hypothesize that *Mentions Masculinity* is highly sensitive to the inequalities of the social worlds journalists are tasked with reporting. For instance, journalists cannot fabricate female sources — or are having troubles finding such sources — in domains where women are entirely absent, such as historically male-dominated professional sectors. *Quotes Masculinity*, by contrast, introduces an element of journalistic agency, as journalists and editorial managers have the capacity to prioritize and highlight specific voices by granting them space within their articles. Consequently, *Quotes Masculinity* is less constrained by the external realities of gender disparities in society and reflects more deliberate editorial decisions. Finally, the *Quotes Stereotype Level* metric further isolates journalistic and editorial agency by focusing on verb choice, which carries connotations and meanings that are unlikely to mirror "real-world" gender inequalities. This metric thus provides deeper insight into the role journalists play in perpetuating or challenging gender norms through linguistic framing. These metrics, taken together, allow us to explore how much of the gender gap in media content can be attributed to external social structures versus the implicit biases embedded in journalistic practices.

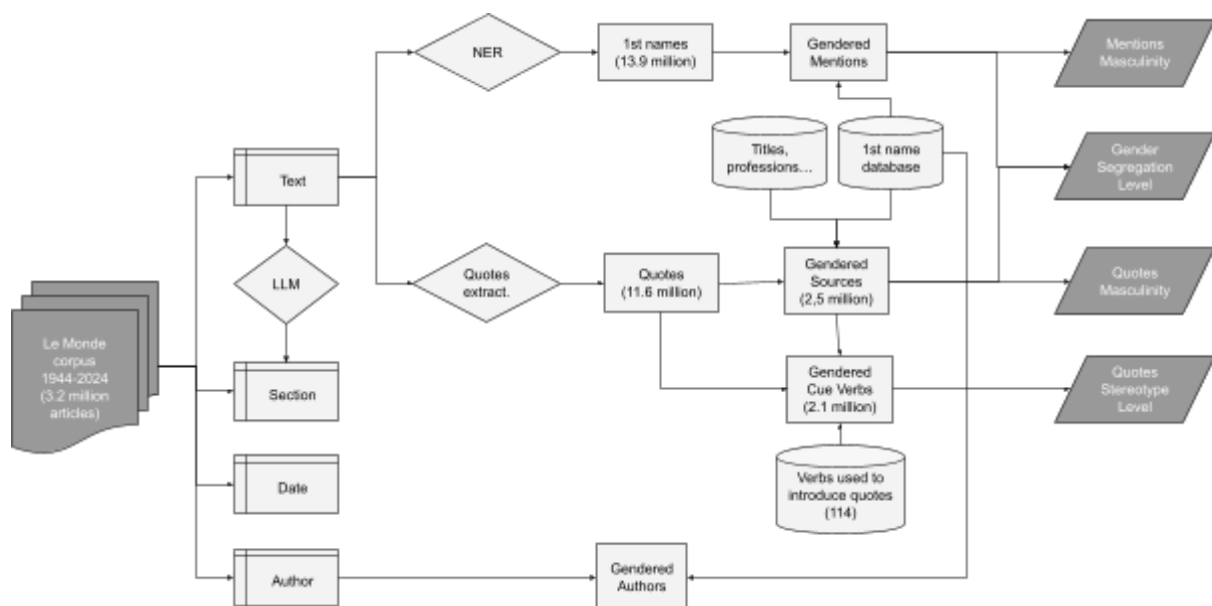


Figure 1. The corpus processing pipeline

Results

a) The male source privilege and its reduction in the long run

Our first key finding reveals that gender representation gaps in *Le Monde* have evolved in a broadly similar pattern over the period studied (Fig. 2). Men's share of both mentions and quotes in *Le Monde* articles has declined, following distinct phases: an almost imperceptible decline from the 1940s through the 1970s; a more noticeable, steady decrease from the 1970s to the 1990s; and a marked acceleration in this downward trend from the 1990s onward (even accelerated after 2010). Overall, the proportion of mentions featuring men dropped from 94% to 74%, while the masculinity of quotes decreased from 98%

to 76%. While the newspaper is still far from gender balance, it represents a significant change, especially considering the extreme initial imbalance. We also observe a trough in the Mentions Masculinity rate between 1950 and 1954, that is specific to the Culture and Society sections (Figure 3). While we could not fully understand its causes, it seems to relate to the momentary presence of a few female journalists in the Culture and Society newsroom (Figure 4), in particular Christine de Rivoyre, who authored 265 articles between 1950 and 1955, featuring 40% of women in her mentions. The trough in masculinity corresponds to the peak of the words ‘robe’ and ‘dance’ (the speciality of Christine de Rivoyre) in the corpus⁸. Both word frequencies collapsed after 1955, suggesting that the theme stopped to be covered. In the Society section, the trough is apparently caused by (i) the high profile serial poisoner Marie Besnard, mentioned over 600 times between 1952 and 1954 and (ii) the publication in 1952 of the serial novel *Que votre volonté soit faite* by Yvonne Chauvin⁹, containing 631 mentions of women and 495 mentions of men.

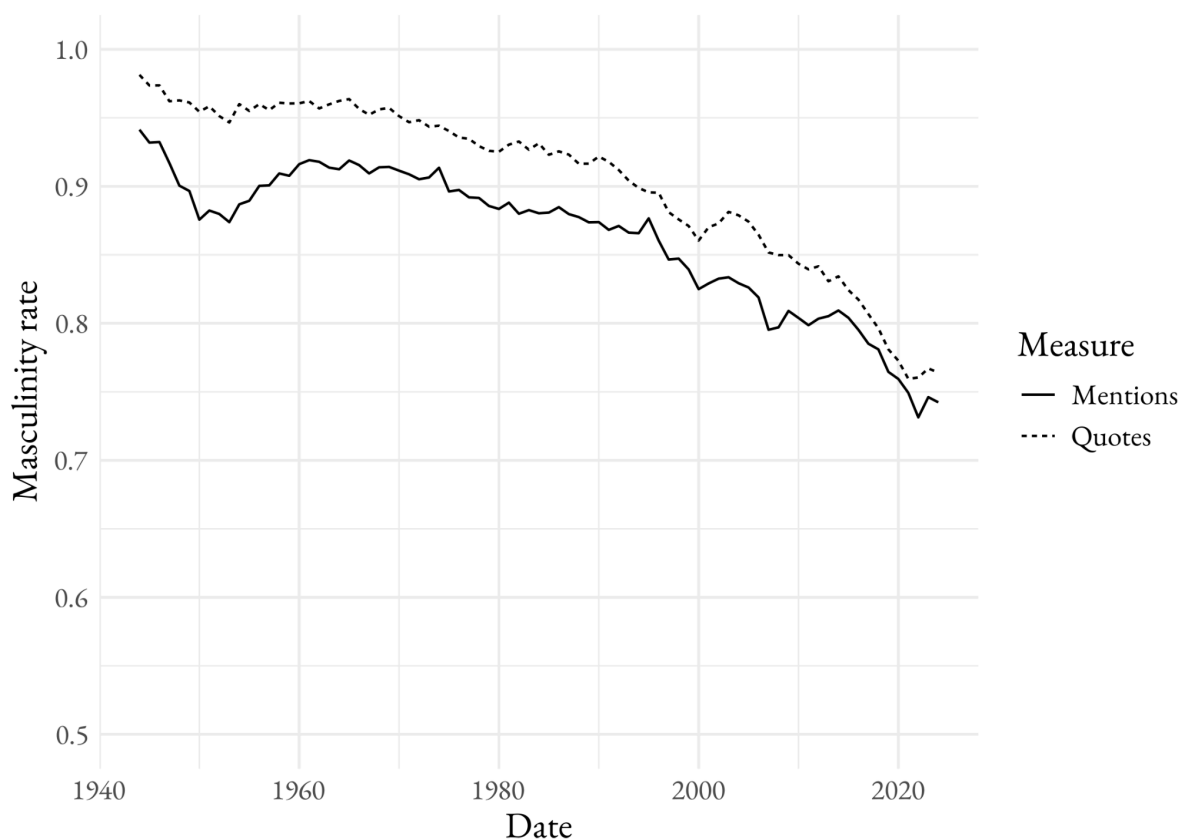


Figure 2: Evolution of Mentions and Quotes Masculinity over time in *Le Monde*

The reduction in the masculinity of mentions and citations in *Le Monde* is not consistent across the different sections of the newspaper (Fig. 3). The only two sections reaching below 70% masculinity after 2020 are Culture and Society, which also started the period at lower levels than the others. Politics, International and Economics are notable for having maintained very high levels of masculinity in both

⁸ The ngrams frequency can be visualized in the [Gallicagram app](#).

⁹ It is the only serial novel ever published in the journal. It is wrongly recognised as part of the Society section as the fictional nature of the document is absent from the text, and the novel relates the life of a family during the German occupation of France.

citations and mentions for a longer time, distinguishing them from the other sections. These sections of the newspaper also have a slower decrease in male sources after the 1990s— especially Economics and International, as Politics experienced two strong demasculinization shocks, likely due to gender parity laws that profoundly changed the gender gap in French politics (Bastin, 2022). The case of the Sport section is very peculiar: its level of content masculinity is far above that of the other sections and it is the only section where the masculinity of mentions and citations is stable until at least 2010.

Another notable finding is the significant narrowing of the gap between the masculinity of mentions and quotes over time. In the 1950s to 1970s, quotes at *Le Monde* were about 5-10 percentage points more masculine than mentions. This gap has steadily decreased over the years, with a brief resurgence in the early 2000s (when, for about 5-10 years, quotes re-masculinized while mentions continued to demasculinize). By the end of the study period, the masculinity levels of quotes had nearly aligned with those of mentions, especially due to a rapid decrease from 2012 on. This suggests that *Le Monde* journalists, who historically favored men even more in quotes than in mentions, now appear to apply nearly the same selection criteria for both.

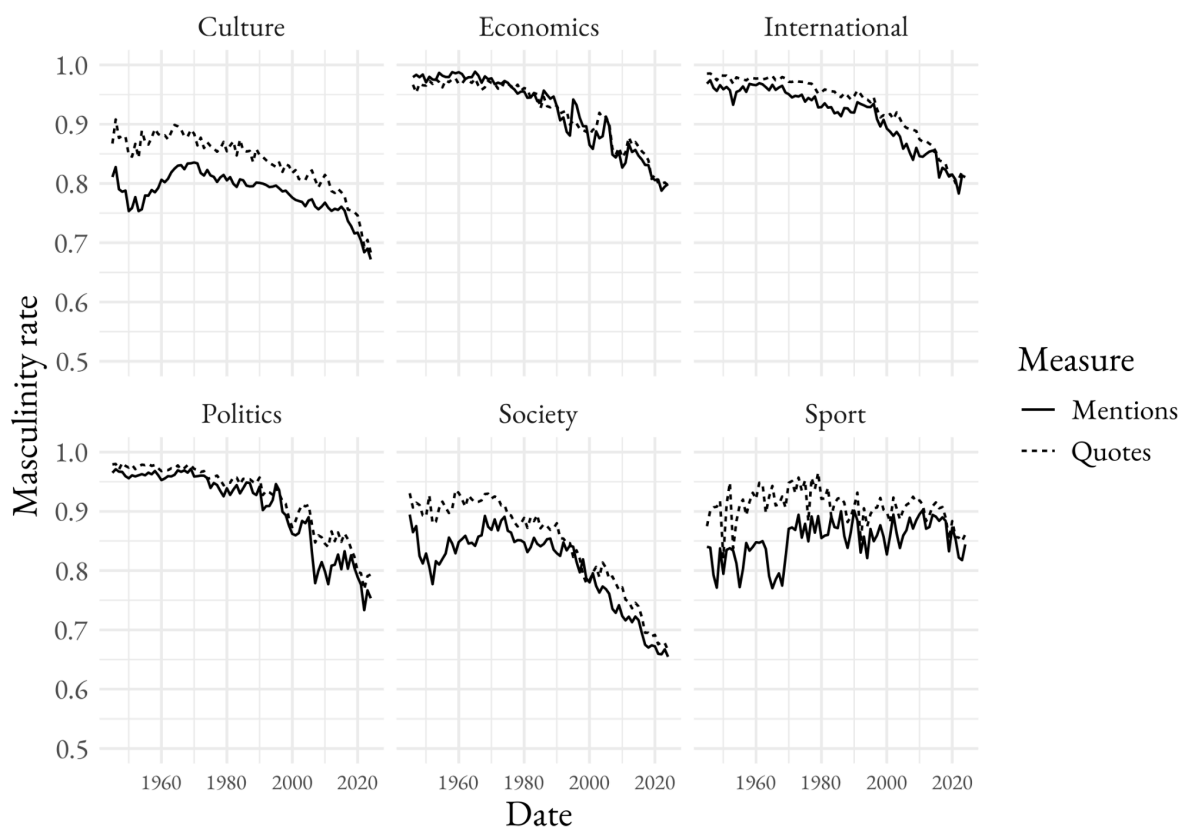


Figure 3: Evolution of Mentions and Quotes Masculinity over time and by sections in Le Monde

Between 1945 and 1970, men were also had longer quotes. On average, they contained more than 20% more characters (Figure 4). By the end of the 20th century, however, quote length became more uniform. Since 2000, women's quotes have tended to be slightly longer than men's.

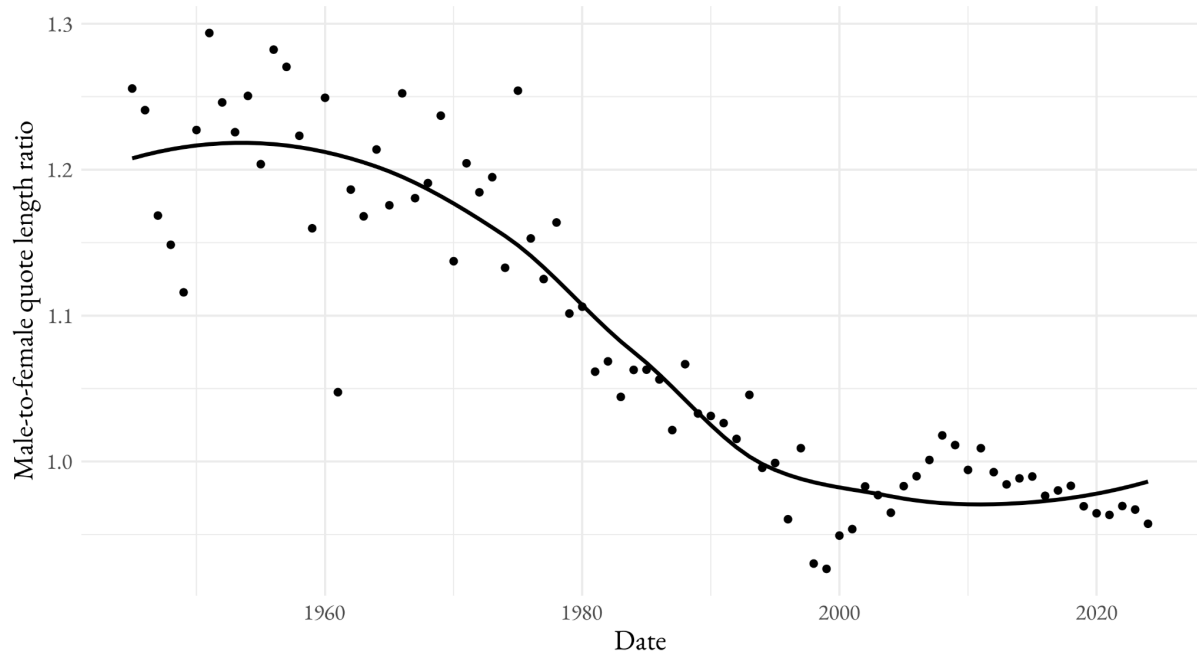


Figure 4: Evolution of the male-to-female quote character count ratio, with a loess regression line

Finally, we investigate the evolution of the Gender Segregation Level, examining whether women's mentions and quotes are dispersed or concentrated in a few articles. Figure 5 shows the overexposure of women to mentions of other women. The dynamics vary across sections. Sports articles exhibit strong and stable gender segregation. In contrast, Culture articles have consistently mixed genders. In other sections, segregation levels have decreased from high levels. Before 1970, women in Economics and International articles were exposed to four times more women than average, and they are now almost uniformly distributed. In Politics articles, the ratio has dropped below 1 since 2020, indicating more gender mixing than random. This shift may result from gender parity laws. Quotes follow the same dynamic, with a consistently higher level of segregation (Figure S4).

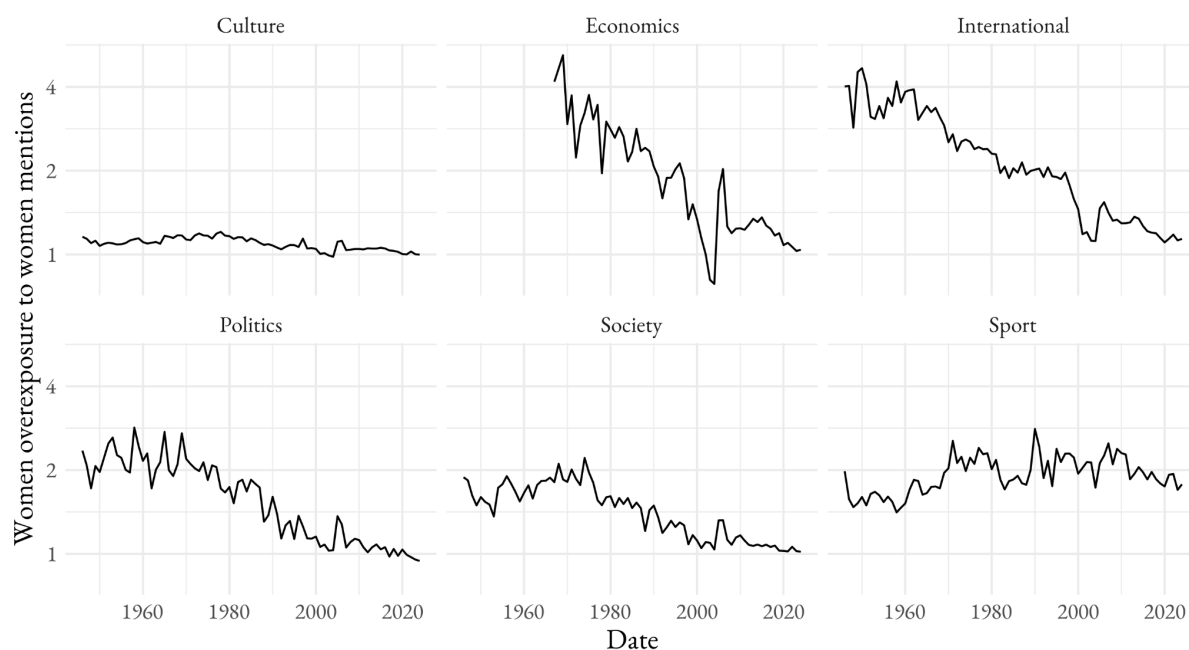


Figure 5: Evolution of the Gender Segregation Level of mentions over time and by sections. The Gender Segregation Level is measured as the gender ratio experienced by women in the articles they are mentioned, divided by the gender ratio in the full corpus in the same year. The y-axis is on a logarithmic scale. Years with fewer than 100 citations by women are omitted.

b) bylines

Journalism at *Le Monde*, like most legacy newspapers, has historically been a predominantly male-dominated occupation. Female journalists were initially rare; the first full-time female journalist was hired in 1948, four years after the newspaper was founded. By 1969, women made up only 16 of 141 newsroom staff members (11.3%) (Eveno, 2004). Their roles were often limited to lower-profile sections, such as lifestyle sections, where the newspaper typically relied on freelance women journalists. Over time, however, the gender gap narrowed significantly. From a predominantly male newsroom in the 1940s and 1960s, the proportion of women signing articles gradually increased, reaching 47% of bylines by the end of the study period (see Appendix S3).

This overall trend, however, hides significant differences between sections of the newspaper (Fig. 4). Sections focused on culture and society have shown a more consistent move toward gender parity. Women now write more than 50% of the content in these areas, reversing the traditional byline gender gap. This shift reflects broader changes in the gender composition of the French journalistic workforce. In contrast, sections such as politics, international news, and business have remained male-dominated for much longer, with only gradual improvements. Since the 2000s, progress in these areas appears to have stalled, with little further movement toward gender parity. The sports section has followed a different pattern. Staffed entirely by men until the middle of the 1980's, it achieved a more balanced gender representation in the 1990s. However, this progress has since reversed, and the section has become increasingly male-dominated once again, bucking the overall trend at *Le Monde*.

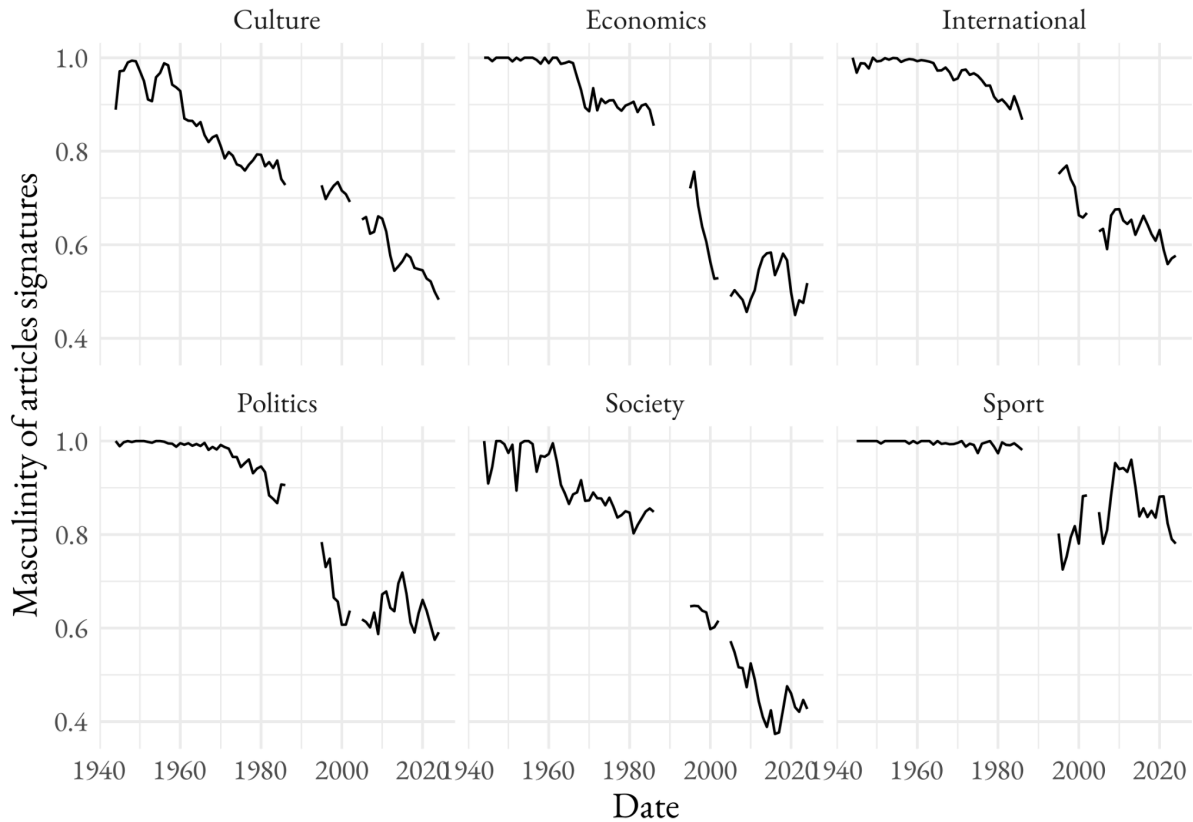


Figure 4: Evolution of byline (signature) masculinity over time and by sections in *Le Monde*. The years 1987-1994 and 2003-2004 have been excluded due to a lack of author metadata.

Did the reduction in the byline gap influence the content gap, defined as the gender balance in mentions and quotes? Based on existing literature, we hypothesized that as women contributed more articles, their work would increasingly amplify women's voices, thereby reducing the masculinity of the content. Our findings reveal a more complex dynamic. The data show that content produced by men consistently exhibits higher masculinity levels than that produced by women (see Fig. 4). Notably, this gap has widened in the recent years, as women have feminized their articles at a faster rate than men. This trend underscores the critical role women journalists have recently played in lowering the masculinity of mentions and quotes at *Le Monde*.

Over the long term, a different trend becomes apparent. Starting at a very low base due to the concentration of women journalists in lifestyle and culture sections at the beginning of the studied period, the masculinity levels in content produced by women increased between the 1950s and 1990s. This rise in the masculinity of women's mentions and quotes is probably due to the combination of their access to new topics, where more male sources were to be encountered, and of a convergence pattern to dominant gender norms in journalism. It was only after women exceeded 20% of the journalistic workforce, post-1990, that women journalists began to actively challenge these norms, reshaping the gendered dynamics of the content. This finding highlights the complex relationship between newsroom composition and content production over time. It also aligns with research on newsroom gender dynamics, which indicates that women are more likely to accept dominant masculine norms in their reporting routines when they are numerously outnumbered.

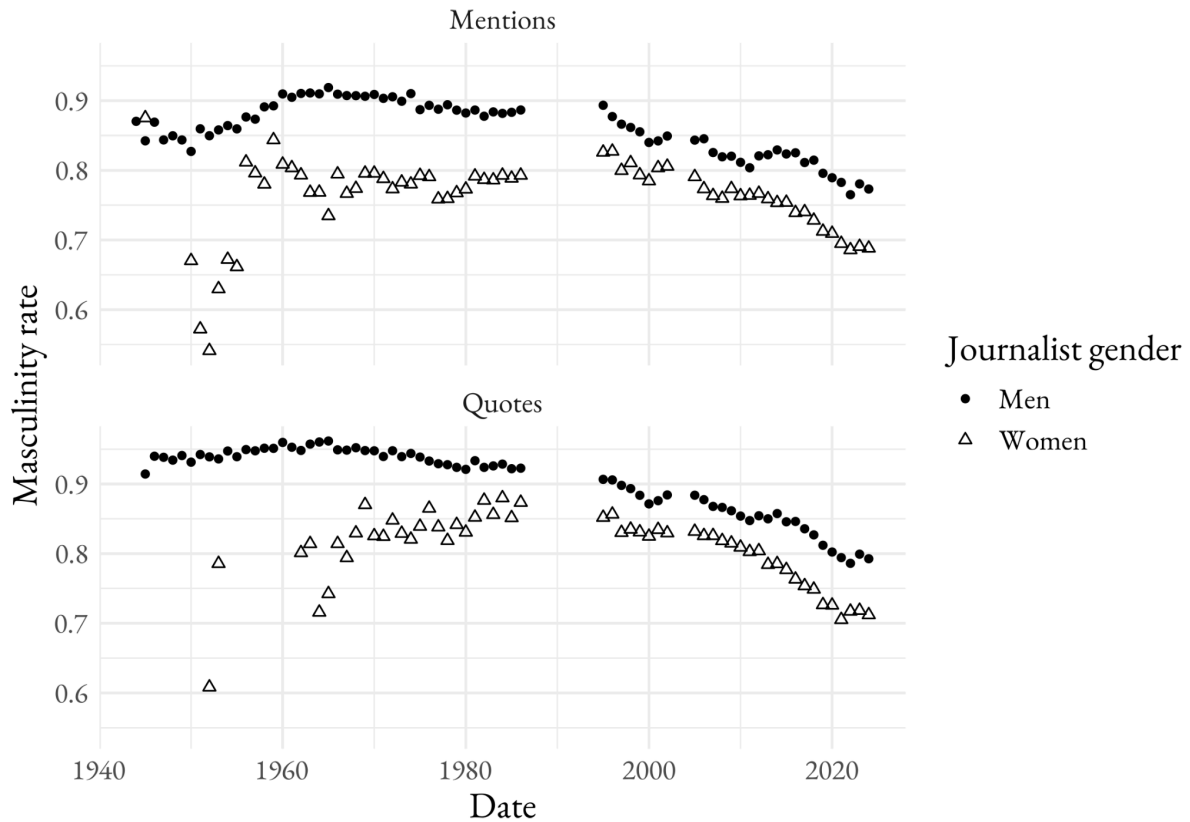


Figure 5: Mentions and citations masculinity over time depending on the journalist gender.¹⁰

c) *Whispering Women, Predicting Men: The Evolution of Gender Stereotypes in Le Monde's Quoting routines*

Gender inequalities in articles published by *Le Monde* can be examined not only through the frequency of mentions and quotes, as demonstrated earlier, but also through the stereotyping associated with the men and women cited. This dimension is captured by our third and final indicator: the Quotes Stereotype Level, which is measured by calculating the odds ratio of specific verbs used to introduce quotes by men versus women.

Analyzing these odds ratios across the entire corpus reveals a clear and pronounced gendered pattern, as illustrated in Figure 6. Some verbs, such as "predict" (*prédire*), "declare" (*déclarer*), "add" (*ajouter*), and "indicate" (*indiquer*), are much more likely to introduce quotes attributed to men. Conversely, verbs such as "whisper" (*murmurer*), "smile" (*sourire*), "remember" (*se souvenir*), "sigh" (*soupirer*), "recount" (*raconter*), "avow" (*avouer*), and "testify" (*témoigner*) are disproportionately used to introduce quotes from women. It is also striking that 'scream' (*hurler*) is disproportionately used to cite women, while 'thunder' (*tonner*), which indicates the same sound volume without a connotation of hysteria, is disproportionately used to cite men. This striking linguistic disparity underscores the persistence of

¹⁰ We only display the years where at least 100 gendered mentions or citations are available. We exclude the 1987-1994 and 2003-2004 years, where signatures are unreliable. See Appendix for the same by section, gap over time, % of women among signatures over time.

gender stereotyping in journalistic practices : male sources in *Le Monde* content are more likely to be introduced as highly authoritative, while women's words are more likely to be introduced with less agency and a more emotional light. We also note that the distribution of log-odds ratio is skewed, with values twice as large for stereotypically feminine verbs. This suggests that women are more assigned to a stereotypical gender role than men.

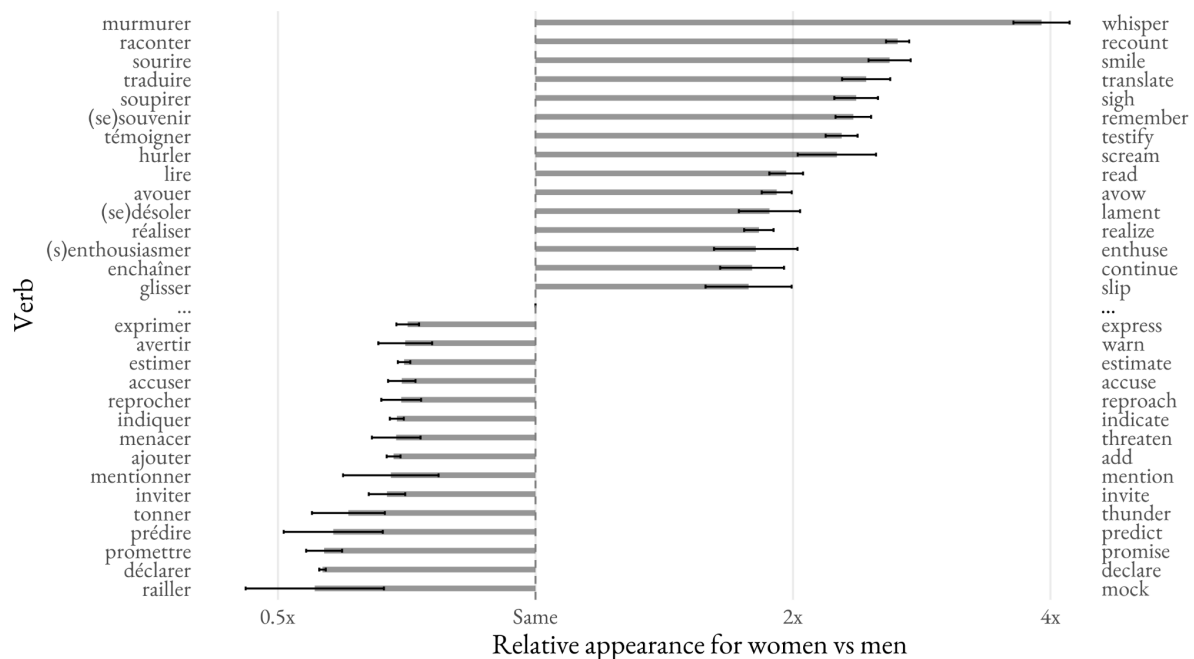


Figure 6: Most stereotyped verbs, measured by log-odd-ratios, with English translation on the right-hand side. Log-odds ratios are computed by decade and aggregated with a method described in Appendix. We only include the 15 verbs most and least associated with women. Error bars represent one standard error.

The level of gender stereotyping in *Le Monde*'s quoting practices has decreased over time. As shown in Figure 7, the gap between the log odds ratios of the most frequently used verbs to introduce male versus female sources has narrowed significantly. This evolution in quoting practices can be summarized by measuring the standard deviation of the log-odds ratio distribution by year (Figure 7B). Notably, the decrease began in the 1950s, predating the decline in the masculinity of the content. This shift in quoting practices can be attributed to two possible factors: either the increased presence of women in positions of power within society, enabling them to be cited in more authoritative ways, or evolving journalistic practices that reflect a reduction in gender stereotyping by journalists. Given the steady decline in stereotyping levels since 1945, coupled with the observation that *Le Monde* journalists began reducing the stereotyping of male and female sources well before they started quoting more women, the latter explanation appears more plausible. However, this decline halted in the early 21st century. In fact, stereotypes appear to have slightly increased over the past decade.

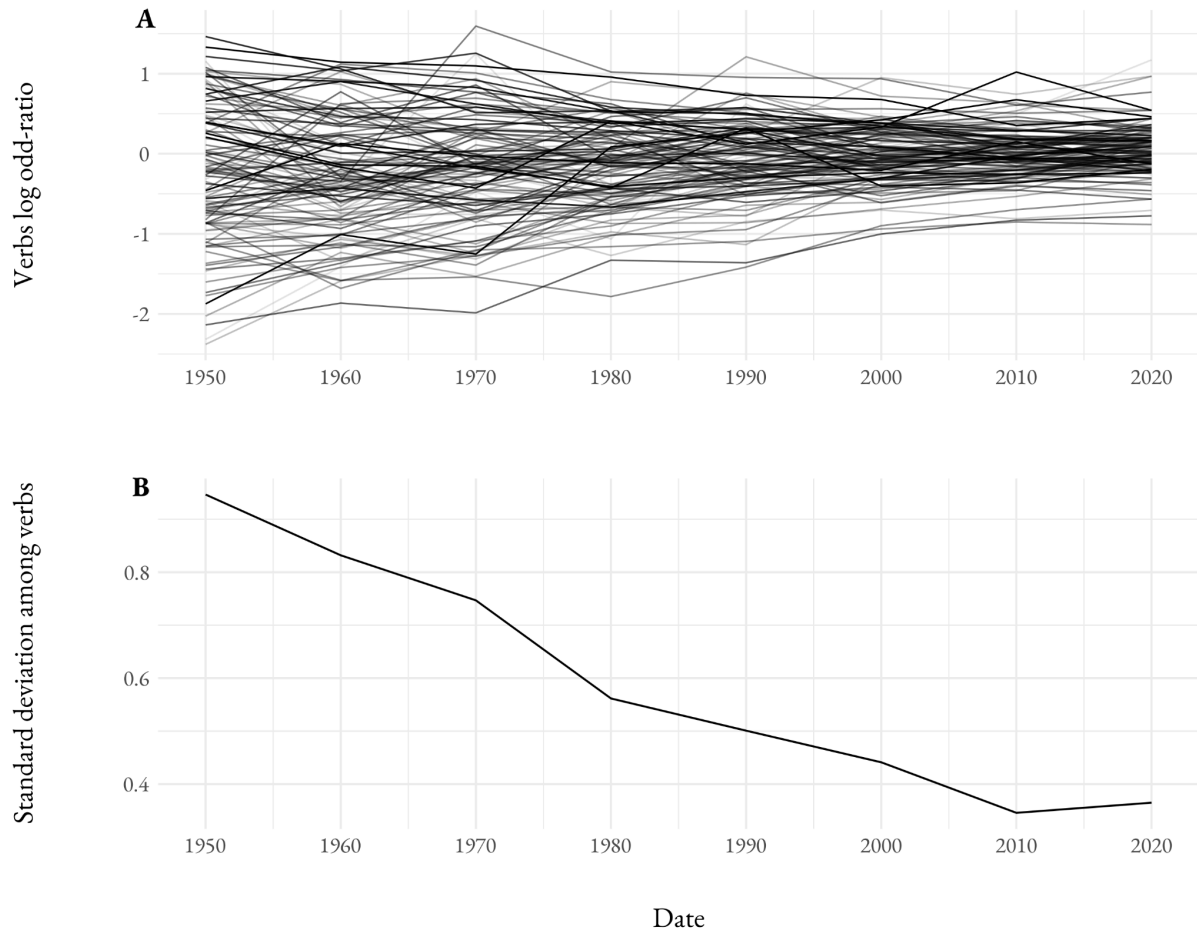


Figure 7: The convergence of log odd-ratios over time, aggregated by decade. We only display verbs that have at least 100 gendered quotes in all decades. More frequent verbs are displayed with a more opaque line (A). The standard deviation among verbs (B) captures the degree of stereotypes over time.

Discussion

This study has highlighted significant trends in the evolution of gender representation and stereotyping in *Le Monde*. First, we observe that masculinity remained approximately stable between 1945 and 1970. After 1970, we document a notable decrease in representation bias over time, although the overall masculinity of the newspaper's content remains high. Importantly, the gap between Mentions Masculinity and Quotes Masculinity has narrowed considerably since the 1990s, particularly in sections like Culture and Society, which have historically included more references to women. This suggests that the privilege historically granted to male voices—men being disproportionately quoted relative to their mentions—has virtually disappeared in *Le Monde's* recent content. Furthermore, the level of quoting stereotypes has shown a steady decline in *Le Monde*, indicating a gradual and limited—yet meaningful—shift in how male and female sources are portrayed. However, the decrease in stereotype intensity came to a halt in the last decade, precisely when the fall in the masculinity of mentions and quotes accelerated.

Our findings align with recent research on gender representation in other cultural mediums, such as books and films, which have similarly documented declines in gender bias (Underwood et al., 2018; Jones

et al., 2020a; Mazières et al., 2021). This suggests that broader social and cultural forces, including shifts in societal norms and expectations around gender, are influencing journalistic practices. The parallels between *Le Monde* and these other mediums point to a larger cultural transformation rather than isolated changes within the newsroom. This shift can be partly attributed to the increasing presence of women journalists in the newsroom, and to the recent reinforcement of their inclination to mention and quote more women (Figure 5).

This study has certain methodological limitations. First, it adopts a binary definition of gender, which precludes an intersectional analysis. As a result, it does not account for the social status of sources or the interaction of gender with other identity markers, such as class. Second, the study's monographic focus limits its generalizability. By concentrating exclusively on *Le Monde*, the findings cannot be confidently extended to other newspapers or media outlets, especially given the absence of comparable data for similar sources.

Third, the performance of our quotes detection algorithm is limited, and seems to be lower for earlier periods in the dataset. Finally, the analysis is constrained by the number of variables available for each article. Incorporating additional metadata—such as article genres or formats, for instance comparing op-eds to news reporting—could yield more nuanced insights. For instance, examining how changes in journalistic formats or the rise of new genres may have contributed to the gradual demasculinization of *Le Monde's* content could offer valuable perspectives.

Despite these limitations, our study provides meaningful insights into the evolution of gender representation in journalism and its broader cultural implications. It underscores the importance of continued efforts to understand and address gender biases, both within the newsroom and in the content produced. In particular, the striking change observed since 2010, when the decline in gender stereotyping ceased — and even slightly reversed — while the decrease in the masculinity of mentions and quotes accelerated, warrants further investigation.

Data availability

A dataset containing the number of mentions and quotes of men and women per article, the list of verbs used to quote and the article metadata (section, date, author and gender of the author) is available on Huggingface: https://huggingface.co/datasets/regicid/lemonde_gender/. The R code used to generate the figures from this dataset is available in the same repository, in the scripts/ folder.

References

- Agnès Y (2002) *Manuel de journalisme. Ecrire pour le journal*. La Découverte.
- Armstrong CL (2004) The Influence of Reporter Gender on Source Selection in Newspaper Stories, *Journalism & Mass Communication Quarterly*, 81: 139–154.
- Azoulay B and de Courson B (2021) Gallicagram: un outil de lexicométrie pour la recherche, <https://osf.io/preprints/socarxiv/84bf3/> (Accessed October 22, 2024).
- Bastin G (2022) Gender Imbalance in the Media: Time Lag or Hysteresis?—French Newspapers, Gender Parity Shocks, and the Long and Winding Road to the Demasculinization of Political Reporting

- (1990–2020), *The International Journal of Press/Politics*, 19401612221143074.
- Blandin C (2011) Elle et le travail des femmes, *Sciences de la société*, 118–135.
- Byerly CM (2004) Feminist Interventions in Newsrooms, *Women and Media : International Perspectives*, John Wiley & Sons, Ltd, pp. 109–131
<http://onlinelibrary.wiley.com/doi/abs/10.1002/9780470776421.ch6> (Accessed May 11, 2021).
- Bystrom DG, Robertson TA and Banwart MC (2001) Framing the Fight: An Analysis of Media Coverage of Female and Male Candidates in Primary Races for Governor and U.S. Senate in 2000, *American Behavioral Scientist*, 44: 1999–2013.
- Collins RL (2011) Content Analysis of Gender Roles in Media: Where Are We Now and Where Should We Go?, *Sex Roles*, 64: 290–298.
- Connell RW and Messerschmidt JW (2005) Hegemonic masculinity: Rethinking the concept, *Gender & society*, 19: 829–859.
- Coulomb-Gully M (2009) Beauty and the beast: bodies politic and political representation in the 2007 French presidential election campaign, *European Journal of Communication*, 24: 203–218.
- de Courson B, Azoulay B, de Courson C, Vanni L and Brunet É (2023) Gallicagram: les archives de presse sous les rotatives de la statistique textuelle, *Corpus*, <https://journals.openedition.org/corpus/7944> (Accessed October 22, 2024).
- Damian-Gaillard B, Montañola S and Saitta E (2021) *Genre et journalisme: des salles de rédaction aux discours médiatiques*. Louvain-la-Neuve (Belgique): De Boeck supérieur.
- Damian-Gaillard B and Saitta E (2011) Le processus de féminisation du journalisme politique et les réorganisations professionnelles dans les quotidiens nationaux français, *Communication. Information médias théories pratiques*, <http://journals.openedition.org/iepnomade-1.grenet.fr/communication/1725> (Accessed May 11, 2021).
- Damian-Gaillard B and Saitta E (2020) La féminisation du journalisme politique : reproduction et subversion des rapports de genre, *Recherches féministes*, 33: 233–250.
- Davis J (1982) Sexist Bias in Eight Newspapers, *Journalism Quarterly*, 59: 456–460.
- De Bruin M (2000) Gender, organizational and professional identities in journalism, *Journalism*, 1: 217–238.
- De Bruin M and Ross K (2004) *Gender and newsroom cultures: Identities at work*. Hampton Press (NJ).
- D’Heer J, Vergotte J, De Vuyst S and Leuven S (2019) The bits and bytes of gender bias in online news: a quantitative content analysis of the representation of women in Vice.com, *Feminist Media Studies*, 20: 1–17.
- Djerf-Pierre M (2005) Lonely at the top: Gendered media elites in Sweden, *Journalism*, 6: 265–290.
- Djerf-Pierre M and Löfgren-Nilsson M (2004) Gender-typing in the newsroom: The feminization of Swedish television news production 1958–2000, *Gender and Newsroom Cultures*. Cresskill: Hampton Press, Inc, 81–104.
- Eveno P (2001) *Le journal Le monde: une histoire d’indépendance*. Odile Jacob.
- Eveno P (2004) *Histoire du journal Le Monde. 1944-2004*. Paris: Albin Michel.
- Falk E (2010) *Women for president: Media bias in nine campaigns*. University of Illinois Press.
- Gans H (1979) *Deciding what’s news. A study of CBS evening news, NBC nightly news, Newsweek, and Time*. New York: Random House.
- Garg N, Schiebinger L, Jurafsky D and Zou J (2018) Word embeddings quantify 100 years of gender and ethnic stereotypes, *Proceedings of the National Academy of Sciences*, 115
<https://pnas.org/doi/full/10.1073/pnas.1720347115> (Accessed November 11, 2024).
- Goffman E (1979) *Gender advertisements*. Macmillan International Higher Education.
- Grau SL and Zotos YC (2016) Gender stereotypes in advertising: a review of current research, *International Journal of Advertising*, 35: 761–770.
- Heldman C, Conroy M and Ackerman AR (2018) *Sex and gender in the 2016 Presidential Election*. ABC-CLIO.
- Jia S, Lansdall-Welfare T, Sudhahar S, Carter C and Cristianini N (2016) Women Are Seen More than Heard in Online Newspapers, *PLOS ONE*, 11: e0148434.
- Jolliffe L and Catlett T (1994) Women Editors at the “Seven Sisters” Magazines, 1965–1985: Did they make a Difference?, *Journalism Quarterly*, 71: 800–808.
- Jolliffe LB (1989) Comparing Gender Differentiation in the New York Times, 1885 and 1985, *Journalism Quarterly*, 66: 683–691.
- Jones JJ, Amin MR, Kim J and Skiena S (2020b) Stereotypical Gender Associations in Language Have Decreased Over Time, *Sociological Science*, 7: 1–35.
- Kaciaf N (2019) *Les pages « Politique »: Histoire du journalisme politique dans la presse française (1945-2006)*. Presses universitaires de Rennes.
- Kinnick KN (1998) Gender bias in newspaper profiles of 1996 Olympic athletes: A content analysis of five major dailies, *Women’s Studies in Communication*, 21: 212–237.
- Kitch CL (2001) *The Girl on the Magazine Cover: The Origins of Visual Stereotypes in American Mass Media*. Univ of North Carolina Press.
- Kozlowski D, Lozano G, Felcher CM, Gonzalez F and Altszyler E (2020) Gender bias in magazines oriented

- to men and women: a computational approach, <http://arxiv.org/abs/2011.12096> (Accessed November 11, 2024).
- Leavy S (2020) Uncovering Gender Bias in Media Coverage of Politicians with Machine Learning, *arXiv:2005.07734 [cs]*, <http://arxiv.org/abs/2005.07734> (Accessed April 27, 2021).
- Leiva R and Kimber D (2022) A Persistent Gender Bias in Chilean Press: The Influence of Journalist's Gender and Editor's Gender, *Journalism & Mass Communication Quarterly*, 99: 156–182.
- Lühiste M and Banducci S (2016) Invisible Women? Comparing Candidates' News Coverage in Europe, *Politics & Gender*, 12: 223–253.
- Macharia S (2015) *Who Makes the News ? Global Media Monitoring Project 2015*. Toronto (Canada): World Association for Christian Communication
https://whomakesthenews.org/wp-content/uploads/who-makes-the-news/Imported/reports_2015/global/gmmp_global_report_en.pdf.
- Macharia S (2021) *Who Makes the News | GMMP 2020 Final Reports*. GMMP - WACC
<https://whomakesthenews.org/gmmp-2020-final-reports/> (Accessed September 20, 2021).
- Martin-Lagardette J-L (2009) *Le guide de l'écriture journalistique / Jean-Luc Martin-Lagardette*. La Découverte. Paris
<https://bibliotheques.ctguyane.fr/ARCHIVES/doc/SYRACUSE/494092/le-guide-de-l-ecriture-journalistique-jean-luc-martin-lagardette> (Accessed November 5, 2024).
- Mazières A, Menezes T and Roth C (2021) Computational appraisal of gender representativeness in popular movies, *Humanities and Social Sciences Communications*, 8: 1–9.
- Meadel C and Coulomb-Gully M (2011) Plombières et jardinières. Résultats d'enquêtes et considérations méthodologiques sur la représentation du Genre dans les médias, *Sciences de la société*, 14–35.
- Meeks L (2024) From Bylines to Headlines: Exploring Gender Bias in the News, *Journalism & Mass Communication Quarterly*, 10776990241284809.
- Melin-Higgins M (2004) Coping with journalism: Gendered newsroom culture in Britain. In de Bruin, M. and Ross, K.(Eds.). *Gender and newsroom cultures: Identities at work* (pp. 197–222), .
- Mitchelstein E, Andelsman V and Boczkowski PJ (2019) Joanne Public vs. Joe Public: News Sourcing and Gender Imbalance on Argentine Digital Media, *Digital Journalism*, 7: 1311–1327.
- Niven D (2005) Gender Bias ? Media Coverage of Women and Men in Congress, *Gender and American Politics: Women, Men, and the Political Process*, M.E. Sharpe.
- Nussbaum MC (1995) Objectification, *Philosophy & Public Affairs*, 24: 249–291.
- Padioleau J-G (1985) *Le Monde et le Washington Post*. Paris: PUF.
- Potter WJ (1985) Gender Representation in Elite Newspapers, *Journalism Quarterly*, 62: 636–640.
- Richard A, Bastin G and Portet F (2022) GenderedNews: Une approche computationnelle des écarts de représentation des genres dans la presse française, <http://arxiv.org/abs/2202.05682> (Accessed May 30, 2023).
- Riedl AA, Rohrbach T and Krakovsky C (2024) “I Can’t Just Pull a Woman Out of a Hat”: A Mixed-Methods Study on Journalistic Drivers of Women’s Representation in Political News, *Journalism & Mass Communication Quarterly*, 101: 679–702.
- Rodgers S and Thorson E (2003) A Socialization Perspective on Male and Female Reporting, *Journal of Communication*, 53: 658–675.
- Ross K, Boyle K, Carter C and Ging D (2018) Women, Men and News, *Journalism Studies*, 19: 824–845.
- Ross K, Evans E, Harrison L, Shears M and Wadia K (2012) The Gender of News and News of Gender: A Study of Sex, Politics, and Press Coverage of the 2010 British General Election, *The International Journal of Press/Politics*, <https://journals.sagepub.com/doi/10.1177/1940161212457814> (Accessed November 12, 2020).
- Saitta E (2005) Le monde, vingt ans après, *Réseaux*, 131: 189–225.
- Santonnicolo F, Trombetta T, Paradiso MN and Rollé L (2023) Gender and Media Representations: A Review of the Literature on Gender Stereotypes, Objectification and Sexualization, *International Journal of Environmental Research and Public Health*, 20: 5770.
- Shinoda LM, Veludo-de-Oliveira T and Pereira I (2021) Beyond gender stereotypes: the missing women in print advertising, *International Journal of Advertising*, 40: 629–656.
- Shor E, van de Rijt A and Fotouhi B (2019) A Large-Scale Test of Gender Bias in the Media, *sociological science*, 6: 25.
- Shor E, van de Rijt A, Miltsov A, Kulkarni V and Skiena S (2015) A Paper Ceiling: Explaining the Persistent Underrepresentation of Women in Printed News, *American Sociological Review*, 80: 960–984.
- Silveirinha MJ, Lobo P and Simões RB (2024) Observing gender in the newsroom: insights from an ethnographic study, *Feminist Media Studies*, 24: 480–496.
- Sjøvaag H and Pedersen TA (2019) Female Voices in the News: Structural Conditions of Gender Representations in Norwegian Newspapers, *Journalism & Mass Communication Quarterly*, 96: 215–238.
- Soumah V-G, Rao P, Eibl P and Taboada M (2023) Radar de Parité: An NLP system to measure gender representation in French news stories, <http://arxiv.org/abs/2304.09982> (Accessed October 22, 2024).
- Sriwimon L and Zilli PJ (2017) Applying Critical Discourse Analysis as a conceptual framework for

- investigating gender stereotypes in political media discourse, *Kasetsart Journal of Social Sciences*, 38: 136–142.
- Tuchman G (1978) The Symbolic Annihilation of Women by the Mass Media, *Hearth and home: Images of women in the mass media*, New York: Oxford University Press.
- Underwood T, Bamman D and Lee S (2018) The Transformation of Gender in English-Language Fiction, *Journal of Cultural Analytics*, 3
<https://culturalanalytics.org/article/11035-the-transformation-of-gender-in-english-language-fiction> (Accessed December 14, 2024).
- Van der Pas DJ and Aldering L (2020) Gender Differences in Political Media Coverage: A Meta-Analysis, *Journal of Communication*, 70: 114–143.

Supplementary Materials

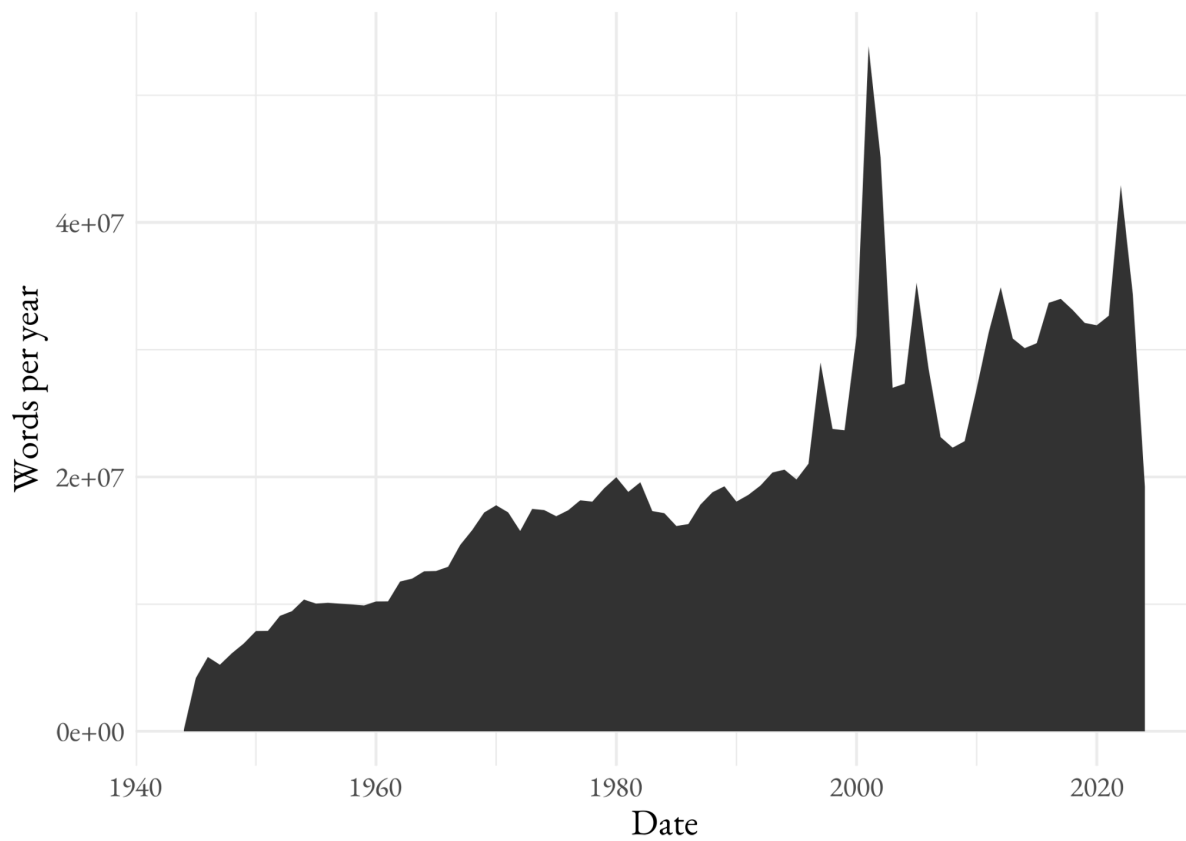


Figure S1: Number of words per year in Le Monde corpus

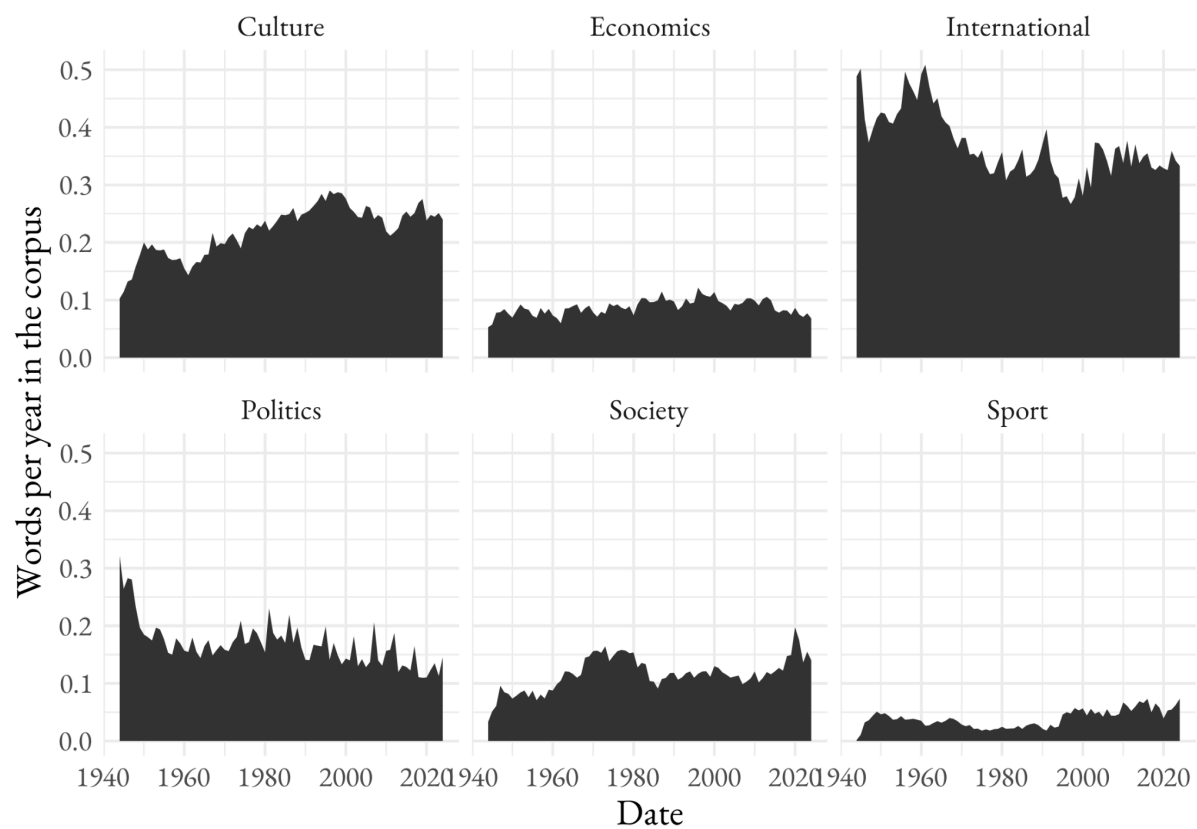


Figure S2: Proportion of words per section over time, omitting unclassifiable articles



Figure S3: Bylines gender gap in Le Monde

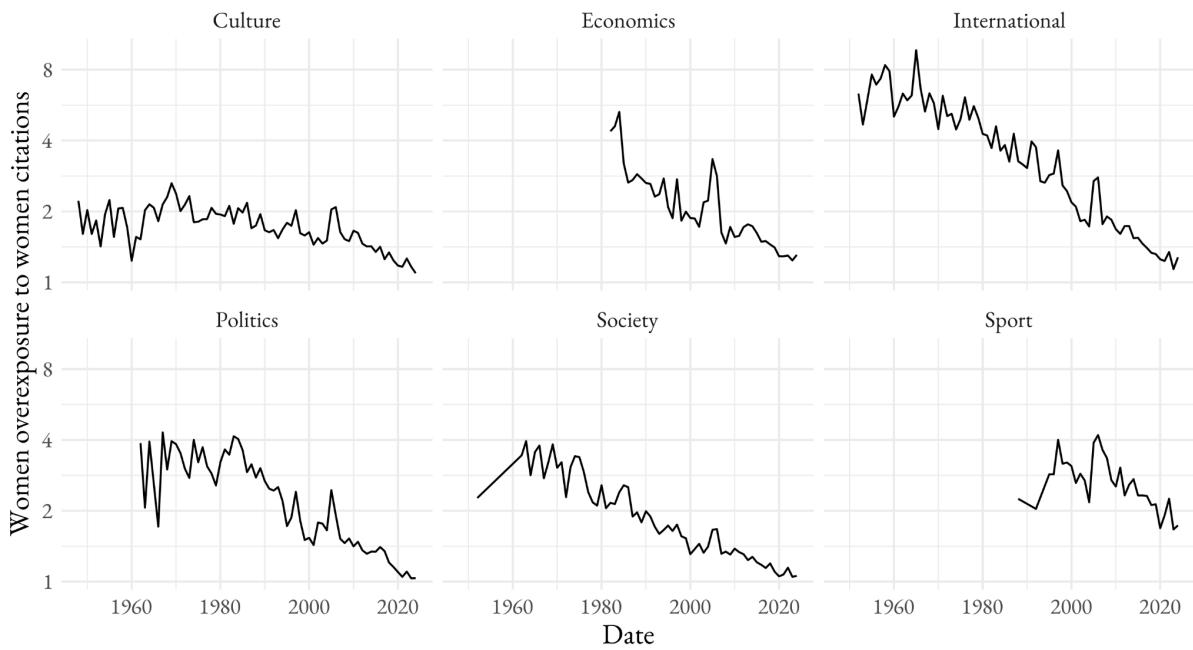


Figure S4: Evolution of the Gender Segregation Level of mentions over time and by sections. The y-axis is on a logarithmic scale. Years with fewer than 50 citations by women are omitted.

Stereotypes methods

To develop the Quotes Stereotype Level measure, we aimed to associate verbs with their propensity to introduce quotes by female speakers relative to male speakers. The standard approach to this problem is to calculate the odds ratio, defined as $(n_male_verb/n_male) / (n_female_verb/n_female)$. However, our 80-year study period presents a significant challenge. Temporal factors influence verb usage, the frequency of female quotes, and the intensity of stereotypes. For example, the verb ‘to tweet’ is predominantly used after 2010, a period when female quotes are more frequent. The odds ratio computed over the entire period is higher than 1, suggesting that ‘to tweet’ is stereotypically feminine. However, when the odds ratio is calculated only for the 2010-2024 period, the conclusion reverses. This discrepancy indicates that the initial result was driven by the higher frequency of female quotes during the timeframe in which ‘to tweet’ is used.

To address this issue, we group quotes by decade (1945-2024, resulting in eight decades). This approach assumes that verb usage, female quote frequency, and stereotypes remain relatively stable within each decade. We identify a consistent list of 123 verbs that appear in at least 100 gendered quotes per decade, ensuring reliable odds ratio estimation. For each decade, we calculate the log-odds ratio (Figure 7A) and the standard deviation among verbs (Figure 7B) to track the decrease in stereotyping over time.

We then aggregate decade-based log-odds ratios to compute period-wide log-odds ratios for each verb. To account for decreasing stereotypes over time, we divide log-odds ratios by the standard deviation of their respective decade, scaled to average 1 so as to remain on the same scale. This implies that within each decade, there is the same amount of dispersion among log-odds ratio. We then average the log-odds ratios, using an arithmetic mean due to their linear scale. We exclude estimates based on fewer than 30 gendered quotes to avoid excessive noise. Although a variance-based exclusion could be used, it would omit extremely stereotypical verbs, which are precisely the ones we are the most interested in, even though they produce noisier odds *ratios*. Instead, we therefore use a total quote count criterion for inclusion. We estimate the variance of average log-odds ratios as the average estimated variance of decade-log-odds ratios, approximated by $1/n_male + 1/n_female$, divided by the number of decades with estimates for each verb. The resulting standard error, representing estimate reliability, is displayed in Figure 6.