Reported speech and gender in the news: Who is quoted, how are they quoted, and why it matters

Maite Taboada
Simon Fraser University, Canada

Abstract
News stories have a well-defined generic structure, consisting of components such as headline, lede, and body, with reported speech a prominent feature, especially in hard news stories. Reported speech serves multiple purposes, from providing evidentiality and intertextuality to contributing to the construction of newsworthiness and to the context creation of news. It is also a site of potential bias in who is cited and how, including with respect to the gender of sources. Using a large corpus of English-language news stories for all of 2023 from the main five mainstream news outlets in Canada (over 370,000 articles from news websites), I examine the gender distribution of those quoted, the syntactic variation in the structure of quotes, and the types of reporting verbs. The study provides a comprehensive overview of the extend of gender bias in contemporary Canadian news, at the same time offering insights into the nature of reported speech in modern news and how it endures and evolves, including in news meant for digital-only publication.

Corresponding author:
Maite Taboada, Department of Linguistics, Simon Fraser University, 8888 University Dr., Burnaby, BC V5A 1S6, Canada.
Email: mtaboada@sfu.ca

Introduction: Reported speech as a feature of news stories

Reported speech is a constant feature of news stories, especially in hard news and current events articles. By *reported speech* I refer here to phrases, most commonly full sentences, that are reported as either direct speech, that is, written between quotation marks, or as indirect speech, both types as complements of a verbal process such as *say, claim,* or *affirm* and traditionally seen as part of the syntactic structure of the clause that contains the reporting verb. Direct speech can also be reported in its own sentence, without a reporting verb, which I refer to as 'floating quotes'. Example (1) provides instances of each from the data analyzed in this study, with quotes in bold. The first quote is direct speech. The second one, in a sentence immediately following, is a frameless or floating quote, with the quote in a sentence by itself. After a couple of other paragraphs, the article also includes an indirect quote, introduced by *said*.[1]

(1)   '**I have guarded optimism, let's just say that**', said Dr. Claudine Storness-Bliss, an obstetrician and gynecologist at SMH who went public this week about a 'scary' resourcing crisis in her unit.
       '**I am glad that we had an open line of communication, that we can continue to raise the issue . . . The assurance that we're being listened to is important, but it's not enough**'.
       [. . .] Storness-Bliss said **the authority's CEO, Dr. Victoria Lee, was not at the meeting Wednesday**.

Reporters quote somebody else's words, whether verbatim or rephrased, for various purposes, among them introducing them as characters in the story, distancing the reporter from the content of the quote, or simply placing the reported speech on the record (Gibson and Zillmann, 1993; Nylund, 2003; Sundar, 1998; van Dijk, 1988). Reported speech can also provide evidentiality and intertextuality (Bednarek and Caple, 2012), contribute to the construction of newsworthiness (Bednarek, 2016) and to the context creation of news (Lukin, 2013).

Many news stories are told in the voice of their protagonists, as politicians present or defend a policy decision, witnesses to an incident or crime retell what they saw, athletes recount a game or competition, and experts reflect on and assess the day's news. This embeddedness of voices is a characteristic of news discourse (Bednarek and Caple, 2012: 90–94; Bell, 1991). Who is quoted and how they are quoted, that is, whose voices we hear in the news, has important consequences for who we think is important in our society. Studies have shown that we see the news' protagonists as those who are quoted in the news (Manning, 2001), whether in print or as sound bytes in audiovisual format, in broadcast news, online news sites, and in social media sites that post video clips. Mainstream news media have extraordinary power in shaping what we think of as newsworthy and who is worthy of our attention, mainly by who they chose to feature and to quote (Bell, 1991; Caldas-Coulthard, 1994; Goodyear-Grant, 2013; Kassova, 2020b).

Against this backdrop, this paper studies how people are quoted in English-language Canadian mainstream news media in all the stories published in 2023, focusing on the structure of the reporting frame and the form of the reported material. Using data from

the Gender Gap Tracker (see Section 3) and analyzing news in digital format (from news websites), I analyze the relative number of quotes by men, women, and other in news stories, focusing on the type of quote (direct, indirect, or floating), and the type of reporting verb. The Gender Gap Tracker was created in a spirit of accountability, providing a visual summary of gender balance across news outlets by extracting quotes and their sources in news stories in Canadian media.

The focus on gender derives from the fact that numerous quantitative and qualitative studies have shown that women are portrayed differently in the media, with women in power (e.g., women in politics or in business) having their gender and their physical characteristics and attire discussed more often (Carlin and Winfrey, 2009; García-Blanco and Wahl-Jorgensen, 2012; Goodall, 2012; Goodyear-Grant, 2013; Power et al., 2019; Trimble, 2018; Trimble et al., 2021; Van der Pas and Aaldering, 2020; Ward and Grower, 2020). Women are not only portrayed differently, but they are also, quantitatively, seen and heard much less often. The Global News Monitoring Project has been studying the gender gap in media across the globe since 1995. It has consistently found that there are fewer women protagonists of news stories and fewer women sources and that (assuming a target of 50% women in news stories), at the rate of progress they observed since 1995, it would take another 67 years to reach parity (Gallagher, 2005; Macharia, 2015, 2020; Ross and Carter, 2011). Numerous other studies, including longitudinal studies (Shor et al., 2015), have found that women are seen and heard less often than men. Non-binary people are very rarely included in studies but, when they are, results show that they are hardly ever present in the news. A series of studies sponsored by the Gates Foundation (Kassova, 2020a, 2020b, 2023) found that the situation was similar in different countries and for specific events, like the Covid-19 pandemic. Explanations for the phenomenon tend to conclude that the persistent underrepresentation of women in the top spheres of power is the ultimate culprit. Studies have also found that bringing awareness to the problem by keeping track of protagonists and sources may increase the representation of women in news stories (British Broadcasting Corporation, 2024; Hawkins-Gaar, 2019; Yong, 2018), thus providing role models and normalizing the presence of women (and men and non-binary people) across all areas of society. These studies typically examine counts, that is, proportion of women and men mentioned in headlines or in news stories; proportion of women and men quoted; or proportions across different sections of the paper (politics, finance, sports, arts, etc.). There are few studies, however, that also study the linguistic structure of the reported speech and whether there are differences not only in how many times women are quoted, but also in *how* they are quoted (but see Caldas-Coulthard, 1994).

The present study contributes to our understanding of (digital) news today and the role of news to shape and be shaped by the social context in which news media operate. The increasing demand for diversity in news stories is driven by sociological changes in societies such as Canada that see diversity and inclusion as important goals for all aspects of society. At the same time, this push for diversity is shaping the kinds of stories news outlets tell, and the range of voices they include, contributing to the agenda-setting role of news organizations. Although this is a study at a point in time and place (Canada in 2023), longitudinal studies may reveal the extent to which demands for diversity are successful, using this study as a baseline for comparison. Thanks to the wide availability of

news in digital format, the analyzed articles were scraped from the websites of news organizations. The study thus also contributes to our understanding of digital news in that it provides an up-to-date analysis of one characteristic of news, reported speech, making use of new tools for its analysis that can be applied to large datasets (see Section 3).

The next section presents an overview of the linguistic literature on reported speech first, followed by a summary of studies in media studies and political science on the influence of whose voices we hear in the news. Section 3 introduces the project that led to a large-scale data collection, the Gender Gap Tracker, and the subset of data collected for this study. The main analysis of the results is presented in Section 4, with a discussion in Section 5, and why who is quoted in the news matters.

Quotations, verbs, and sources in news stories

Reported speech shows extraordinary variation across the world's languages and is often the site of markers of evidentiality and stance, where the speaker interprets and evaluates the reported speech, indicating affiliation or distance and the source of the speech, for example, as a witness or through hearsay (Chafe and Nichols, 1986). Goddard and Wierzbicka (2019) explore some of that typological diversity across languages, proposing that reported speech is a pivotal human phenomenon: We live 'among other people's utterances: they are the stuff of our daily life, our dreams, memories, thoughts, and stories, the fabric of our mental, emotional and social lives' (Goddard and Wierzbicka, 2019: 197). Our stories have traditionally included reported speech. It was natural then, that news articles, which we often refer to as news *stories*, also feature reported speech prominently.

Thus, quotes in news media are an important part of the story, a characteristic feature of journalistic discourse. Journalism textbooks teach that quotes, especially direct speech, contribute to make the story more believable and understandable (Gibson and Zillmann, 1993); they 'lend credit to speakers who use them in their messages' (Zelizer, 1989: 369). Quotes are so central to news stories that news items have been characterized as 'talked into being': 'news content revolves around the practice of quoting: the (co-) construction, selection, editing and representation of comments, explanations, interpretations, speculations, praise and blame, among others' (Nylund, 2003: 844). For journalists, obtaining quotes and attributing them to credible sources are essential aspects of journalism discourse, 'the bread and butter of a news story' (Sundar, 1998: 56). Quotation practices, and the function of quotes, is different in journalistic discourse than in conversation or in fiction (Waugh, 1995). According to Nylund (2003), quotes serve narrative purposes in the story, among them: confirmation of claims of newsworthiness (novelty, validity, public relevance); criticism and blame (providing conflict and drama, sometimes with a 'balanced' perspective, i.e., with conflicting viewpoints); evaluation (e.g., establishing that something is a problem) and emotion (which the reporter does not want to convey themselves); subjective experience and sense of presence and validity (in contrast, with the reporter providing a first-person point of view); or solutions to problems (media finding solutions and sources who will speak to them). These functions can be summarized as engagement and credibility (Zelizer, 1989). Quotes are also deployed as argumentative devices, to either advance a thesis or to provide evidence for it (Smirnova,

2009). Quotes in journalism have to persuade the reader that the reporter's perspective is the correct one, thus contributing to truthfulness, factuality, believability (van Dijk, 1988), and reliability (Caldas-Coulthard, 1994). Quotes can also be challenged by those being quoted, therefore making the reporter accountable.

The people quoted in the news are referred to as 'sources' and are typically politicians or bureaucrats who are either responsible for problems or trying to solve them, experts who identify or elaborate on the problems, and victims who suffer because of the problems (Nylund, 2003). In more positive news stories of the uplifting kind, sources may be people who achieved something (award winners, athletes, lottery winners). Manning (2001), in an influential treatment of news sources, showed the importance of which people and organizations journalists get their information from. He argued that not all sources enjoy the same degree of access and the same ability to communicate their perspectives. Arguably, the issue of who is quoted is central to representation in media: 'the ability to speak in the news is important for influencing the terms of broader social and political contestation' (Benson and Wood, 2015: 803). Lazaridou et al. (2017) argue that examining quotes may be the most straightforward and quantitatively feasible way to identify media bias. Compared to selection bias (which news are covered), coverage bias (which aspects of the event are covered) or framing bias (how the event is described), quotes can be studied at large scale and can reveal bias in who is chosen as a source and how they are presented. Lazaridou et al. (2017) found that, in two UK newspapers, reported speech is more frequently by politicians of the governing party and that the two newspapers differ in how faithfully they quote the original speaker. A study by Niculae et al. (2015) uncovered political bias across news outlets by studying how often and how extensively they cite former US President Barack Obama. This study deals with gender bias and who is quoted, by gender, trying to address an important issue of representation.

In addition to the choice of sources, how journalists convey the voice of those sources is also important. A great deal of research in journalism has studied the effect of quotes in perception of newsworthiness and credibility, in addition to readers' engagement, with particular focus on whether direct or indirect speech increases any of those. Direct quotes are thought to render news stories more lively and trustworthy (Clark and Gerrig, 1990; Gibson and Zillmann, 1993; Short et al., 2002; van Krieken, 2020; Vis et al., 2015), in part because direct speech is supposed to reproduce the exact wording of the source, despite studies that point that the assumption of verbatim reproduction does not always hold (Lehrer, 1989; López Pan, 2010; Short, 1988; Waugh, 1995), and that the surrounding text does much of the work for quote interpretation (Nylund, 2003). Some studies have found that direct quotes do not have a noticeable impact on credibility or engagement and that the conflicting evidence about the importance and credibility of quotes may depend on whether the study is conducted on online news stories or offline (i.e., paper) articles (Sundar, 1998; van Krieken, 2020), because the default reading mode in online news stories 'might be one of distrust and low credibility' (van Krieken, 2020: 156). Other factors that influence credibility and engagement, such as trust in the media (Matthews, 2012) or whether the story has a narrative format and its topic (Kelly et al., 2003), may be more important than the type of quote present in the story.

I have, thus far, mentioned quote types and sources as fundamental parts of reported speech. We know something is reported speech because it points to a source, another

subjectivity or voice in the text other than the author's. We also know because there is an introductory, reporting, or quoting verb, a verbal process like *say* or *claim* that points to that source. Together, the named source and the reporting verb constitute the quoting frame. And, naturally, we know we are reading reporting speech because there is a quote, content either in direct or indirect speech that is considered important enough to be repeated or summarized. Thompson (1996) argues that reported speech (language reports, in his terminology), include a fourth component, the attitude, that is, the evaluation by the present reporter of the message of the original speaker (see also Bednarek, 2006; Scollon, 1998). Such evaluation can already be contained in the reporting verb, as there is a world of difference between saying something and claiming it, a linguistic manifestation of the heteroglossia inherent in language (Bakhtin, 1986; Martin and White, 2005). Thompson (1996) points out that quoting verbs may also indicate attitude towards the speaker (rather than towards the message), with examples such as *brayed, wittering on*, or *fulminated.* Jullian (2011) adds that the evaluation or appraisal inherent in quotes is also an appraisal on the part of the journalist towards the events described and their role in their world, that is, reporting verbs encompass the ideology of the reporter (White, 2000).

Taking into consideration this complexity of reported speech in the news, this paper deals with who the sources are in contemporary Canadian news articles, as well as how they are introduced, together with the structure of the quotes the journalist attributes to them. To do so, it makes use of the Gender Gap Tracker (GGT), which allows large-scale analysis of potential gender bias in news in digital format, including rich language analyses regarding the characteristics of reported speech. I discuss the background of the GGT in the next section.

The Gender Gap Tracker

The Gender Gap Tracker is a collaboration between an Ottawa-based non-profit organization, Informed Opinions,[2] and the Discourse Processing Lab[3] at Simon Fraser University. Data collection for the project started in October 2018 and it was officially launched in 2019, with a public web page that tracks the proportion of men and women quoted in Canadian news media in English.[4] A French-language version, the *Radar de parité*,[5] was launched in spring 2023. The code to scrape the news articles and the system to extract quotes is publicly available,[6] and the quotation tool is also available as a service through the Australian Text Analytics Platform;[7] albeit without gender analysis; see Bednarek et al. (2024).

The goal of the GGT was to bring awareness to the underrepresentation of women in Canadian media. Informed Opinions had carried out some manual studies (Morris, 2016), but continuous tracking was difficult and thus technology was proposed as a solution. A team of researchers at the Discourse Processing Lab built a Natural Language Processing (NLP) system to identify the people mentioned in an article and the reported speech found therein. After that, a coreference and matching step lines up the names of people with the quotes found in the news article. The system uses a combination of rule-based NLP (to find the beginning and end of segments between quotation marks),

syntactic parsing (to determine which clausal complements are complements of verbs of saying), and neural methods (to build coreference chains). We then use external services and a large list of people's names to assign gender to the speakers of those quotes. The GGT extracts a large amount of information. Importantly for this study, information includes the reporting verb (*says, claimed, has stated*, etc.) and the type of quotation (direct, indirect, floating, and subtypes of those). With that information and the gender of the speaker, we can provide rich analyses of the relationship between gender and type of quote. I present additional details of this extracted information at the beginning of Section 4, before analyzing the results.

Throughout this article, and on the GGT website, our gender categories are 'women', 'men', and 'other' or 'unknown'. I would like to acknowledge here that this is an unsatisfactory solution to a complex reality. The world is not simplistically divided into women and men. Our solution is based on names, that is, we only assign a quote to a person if we can assign it to a full name. We do not assign 'woman' to a quote that starts with *She said that*, but instead find the antecedent for *she* and match it to a full name. This is feasible in news articles, which always quote a person's full name when they are first introduced. Then, we have two approaches, self-presentation and common association. In the first case, we assign gender based on full name and the self-presentation of that person in public. When a source is not a public figure, then the GGT assigns gender based on the most common association of that first name with a gender, using large databases of genders such as GenderAPI. Because of this reliance on external information, inclusion of non-binary and gender-diverse categories is very poor. In cases when public figures self-present as non-binary (e.g., they use *they/them* pronouns), we assign them to the 'other' category. This is also the case for names that are commonly associated with men, women, or non-binary people (Alex, Amir, Ash).

Evaluation of the GGT shows quite high accuracy in detection of people's names, quotes, and gender prediction. Most importantly, we found that there is only a small bias in predicting gender. Our main concern is that errors would disproportionately affect one gender or the other, that is, that we would more often assign 'woman' to a source that is a man, or vice versa. An error analysis of the top sources showed that we had a slight bias against women, that is, that we assigned the label 'men' to women sources slightly more often. The error rates, however, are very small: 0.1% of the cases for men and 0.2% of the cases for women (for full details of the evaluation process, see Asr et al., 2021). Thus, we are confident that the results shown in Section 4 below reflect reality rather accurately.

Technical details and results for the first few years of data are available in two published papers (Asr et al., 2021; Rao and Taboada, 2021). Several blog posts and opinion pieces by our group have drawn attention to the problem of underrepresentation (Rao et al., 2021; Taboada, 2020; Taboada and Asr, 2019; Taboada and Chambers, 2020).

Analysis: Who is quoted, how are they quoted

The GGT database contains about 2 million articles since its inception on October 1, 2018. For this paper, I used all articles published in 2023, a total of 371,724 articles,

Table I. Numbers of articles and quotes included in the study.

News organization	No. of articles	No. of quotes	No. of women quoted	No. of men quoted	% Women quoted
CBC News	52,771	656,971	37,726	64,802	36.70
CTV News	76,118	658,675	36,609	73,274	33.26
Global News	40,902	355,798	18,090	36,030	33.37
National Post	50,351	367,258	15,499	54,514	22.09
The Globe and Mail	43,277	427,806	20,326	59,785	25.35
The Star	108,305	549,246	24,671	77,698	24.06
Total	371,724	3,015,754	152,921	366,103	29.41

divided as shown in Table 1 by news organization. The table also includes the numbers of quotes by women and men, and the percentage of those by women.[8]

The table, thus, addresses the first objective of this section, to answer the question who is quoted. It turns out that it is most often men, even in data as recent as 2023. Women, roughly 50% of the population, are quoted on average 29% of the time. There is some variation across news outlets. At the top of the list is CBC, the Canadian Broadcasting Corporation, a public broadcaster with a commitment to equity, diversity, and inclusion.[9] At the bottom we find The National Post, a right-of-center broadsheet.[10] This shows that, despite the realities of gender representation in the real world, news organizations do have some control over whose voices they feature.

The second objective of the analysis is to study how sources are quoted in the news, and whether this differs according to their gender. To do that, I first briefly summarize the type of information that the GGT extracts and, in the rest of this section, discuss how sources are quoted through the type of reporting verb and the syntactic structure of the quotes.

For each article and each quote within the article, the NLP system in the GGT extracts a great deal of information, including all the people mentioned in the article, all the people quoted, the quotes themselves, and the structure of the quotes (reporting verb, type of quote, and length of the quoted material). For example, for the three quotes in Example (1) above, the system would extract the information presented in Table 2. Acronyms for quote types are explained below; for instance, 'QCQ' means 'quotation mark – content – quotation mark'. Note that token counts include both words and punctuation.

In Section 4.1 I provide quantitative results of the types of reporting verbs found in the data, while in 4.2 I expand on the structure of reported speech. General trends are reported, and potential gender differences are also investigated.

Reporting verbs

A large variety of verbs introduce quotes, in either direct or indirect speech. In the roughly 3 million quotes in our dataset (see Table 1), about 2.5 million had a reporting

Table 2. Quote information for quotes in Example (1).

Quote	Speaker	Verb	Quote type	Is floating?	Token count
'I have guarded optimism, let's just say that',	Dr. Claudine Storness-Bliss	Said	QCQVS	No	10
'I am glad that we had an open line of communication, that we can continue to raise the issue . . . The assurance that we're being listened to is important, but it's not enough'.	Dr. Claudine Storness-Bliss	–	QCQ	Yes	36
the authority's CEO, Dr. Victoria Lee, was not at the meeting Wednesday.	Storness-Bliss	Said	SVC	No	15

verb, with the rest being floating quotes. Of those verbs, *say* is by far the most frequent, certainly when only the lemma is considered (accounting for over 70% of all reporting verbs), but also in its many conjugated forms, with *said* and *says* at the top of unlemmatized forms of verbs, as shown in Table 3. The next most used form of quotative form is *according to*, not strictly speaking a verb. These results seem to align with common trends in news coverage, with similar results found in non-Canadian data (see, e.g., Bednarek et al., 2024: 20), including a general preference for neutral reporting verbs in British (Bednarek, 2006: 141) and US news (Garretson and Ädel, 2008). The unlemmatized side of the table tells us that most of the reporting verbs are in the past. On the lemmatized side, we see common verbs of saying (*say, tell, add, note, announce, report, suggest, argue, explain, confirm, warn*). Two verbs to note are *write*, which indicates that many of the verbal reports we see in news come from written sources, perhaps from press releases or digital communication (email, social media posts). The other somewhat surprising verb is *think*, which seems to be used to introduce a summary of a quote to come. For instance, in (2) the indirect quote introduced by *thinks* foregrounds the direct quote in the next paragraph. All quotes in the example are in bold.

(2) O'Donnell has high hopes that decriminalization will remove some of the sting of stigma for people who are struggling with addiction. And he thinks **other provinces could follow B.C.'s example.**
'If we do succeed in helping people, I'm sure the rest of Canada could do the same', he said, quickly adding **that greater access to a regulated safer supply of prescription drugs, like hydromorphone tablets or fentanyl patches, is part of the answer to stemming the tide of the overdose crisis.**

After the top 15 verbs in Table 3, the rest of the distribution has a very long tail of 3100 different verb lemmas, with many verbs appearing in a handful of quotes each, such as *attach, endorse, equate, excoriate, mourn*, or *project*, indicative of a more formal register and perhaps an effort towards stylistic variation in the reporting verbs. A few of the reporting verbs refer to manner of speaking, such as *croak, spit*, or *mumble*, reflecting attitudes towards the speaker (Thompson, 1996). There are very few examples of informal reporting verbs, such as *be like* or *go* and, when they appear, they tend

Table 3. Top conjugated and lemmatized forms of reporting verbs.

Verb, conjugated	Instances	%	Verb, lemmatized	Instances	%
Said	1,389,627	55.33	Say	1,764,019	70.23
Says	250,713	9.98	According to	110,195	4.39
According to	110,195	4.39	Tell	106,900	4.26
Told	97,280	3.87	Add	66,983	2.67
Say	95,331	3.80	Write	32,875	1.31
Added	43,725	1.74	Note	31,958	1.27
Wrote	29,131	1.16	Announce	27,717	1.10
Saying	28,348	1.13	Report	19,550	0.78
Announced	25,324	1.01	Suggest	19,507	0.78
Noted	17,984	0.72	Think	19,111	0.76
Adding	17,121	0.68	Argue	18,302	0.73
Reported	15,809	0.63	Explain	17,786	0.71
Explained	11,167	0.44	Confirm	13,293	0.53
Confirmed	10,846	0.43	Warn	12,551	0.50
Think	9987	0.40	State	11,316	0.45

to be in the speech of a source, as in Example (3), where the quote introduced by *go* (in bold) is inside a floating quote which reports what Wainwright said.

(3) 'I think what I enjoy most about it is just my love for the game of baseball and pitching in general', Wainwright said. 'I love to sit on the bench next to our guys, next to our pitchers, and go **"Why do you think he threw that pitch?"** or "What do you think the hitter's thinking right now?" . . . I love watching for that and I love talking about it'.

As mentioned in Section 2, reporting verbs can carry attitude/evaluation and other connotations, which makes them relevant to an investigation of gender bias. In terms of gender distribution, there are actually few differences in the relative proportion of reporting verbs by men and women. To carry out this analysis, I extracted a smaller sample of quotes (a total of 937,131) which I could be sure were clearly attributed to either a man or a woman (recall that the system also produces 'unknown' quotes). Table 4 shows that the top verbs are the same, and in very similar proportions. These results may suggest that the general preference of news reporting for neutral verbs reduces any potential gender bias in the use of non-neutral reporting verbs, perhaps even any kind of bias, with reporters inclined to use very general verbs to avoid the impression that they are interpreting or evaluating the source's words.

Perhaps more interesting is an analysis of verbs that are used by men and not by women, and vice versa, rather than relying on the most frequent verbs which tend to be neutral (Table 4). Caldas-Coulthard (1994) found, like this study, that men are quoted much more often than women, but also that women are more likely to scream, yell, and nag than men. To investigate such potential bias in contemporary Canadian data,

Table 4. Top 10 reporting verbs, by gender.

Verb, lemmatized	Men		Women	
	Instances	%	Instances	%
Say	485,884	78.78	258,697	80.75
Add	28,892	4.68	14,325	4.47
Tell	26,859	4.35	12,937	4.04
Write	8854	1.44	4806	1.50
Note	8591	1.39	4255	1.33
Explain	5606	0.91	3485	1.09
Think	3224	0.52	1587	0.50
Argue	2770	0.45	1016	0.32
Suggest	2609	0.42	998	0.31
According to	2256	0.37	1165	0.36

I extracted verbs that were only used by either men or women, with samples shown below in (4) and (5):

(4) Verbs that introduce only quotes by men: *admonish, articulate, attack, bellow, bemoan, brag, charm, disdain, excoriate, grouse, mutter, object, portray, preach, push, rant, snarl, threaten.*

(5) Verbs that introduce only quotes by women: *bitch, chalk, curse, diagnose, freak, hypothesize, lack, legislate, purr, recollect, retell, screech, shudder, spew, stammer, strive, wallow, yearn.*

There does seem to be some verb usage which corresponds to gender stereotypes (e.g., aggressivity for men, emotionality for women). However, it is difficult to make generalizations, as the numbers are very small. For instance, *threatened* was only used by men, but only 13 times. The verb *curse* is only used by women, twice. These low frequencies support the hypothesis that journalistic norms about neutral reporting expressions have important discursive effects on gender bias in reported speech, namely that gender bias may be reflected in the proportion of those quoted, but not necessarily in the verbs used to introduce the quotes.

Quote types

As discussed in section 2, differences in quote type (especially direct/indirect) have been said to influence both credibility and readers' engagement, although these effects are still debated. In order to accurately parse the different types of quotes, the GGT created a classification system to separate indirect, direct, and floating quotes. Indirect quotes are those introduced as part of the subordinate structure of a sentence with expressions such as *They said that*, and are different from direct quotes in that the latter are graphically identified with quotation marks. Floating quotes are not part of the syntactic structure of a matrix clause, but appear in a sentence on their own, also identified by quotation marks.

In the literature, floating quotes are referred to as reported speech without a framing clause (McGregor, 1994), unintroduced dialogue (Tannen, 1986), 'defenestrated' speech (Spronck and Nikitina, 2019), or 'insubordination' (McGregor, 2019) because the speech that is reported appears without a matrix or main clause. Following our previous work on the Gender Gap Tracker, I use the term 'floating quotes' (Asr et al., 2021).

Indirect quotes were classified according to the order of quoting frame versus content, that is, introductory subject (S) and reporting verb (V) versus content or quote (C). For instance, a quote of type SVC is the prototypical *The Minister said that . . .* Thus, the possibilities for indirect speech are those in Example (6), with the quote type by acronym at the beginning of each example, with the relevant quote content in bold type.

(6) SVC: The last witness of the day was Brian MacRury, who spent 27 years with the Sudbury police, much of that time as a canine track officer. McRury led the canine track the morning of Sweeney's murder. He said **his dog, a German shepherd named 'Oakey' was trained to track human scents**.

CSV: Coverage of recent presidential elections, the coronavirus pandemic, protests against police killings of Black Americans and other events convinced Janis Fort that the media can't be believed. **One station will cover a story that others ignore**, she said, leaving viewers not sure whom to trust.

CVS: **For the pharmacy sector, the biggest issue in transitioning away from the fax machine will be who covers the cost if the digital communications platforms that replace it are expensive**, says Ng.

VCS: Noting **that Telefilm administered $158.7-million in funding support over the course of 2022-23 '16 per cent more than the previous year'** board chairman Robert Spickler emphasized that production is back on track, cinemas have reopened and film festivals have returned to their prepandemic in-person sizes and strengths.

Direct speech is represented with the same three letters (S, V, C), plus Q to capture quotation marks. Some examples are provided in (7). Floating quotes are a special example of direct speech, where the quote appears in a sentence by itself, as shown in the last example in (7).

(7) QCQSV: Three minutes into the game, sophomore point guard Zakai Zeigler, who brings energy on offense and defense, went down with an injury to his left knee. **'We all hurt for Zakai',** Tennessee coach Rick Barnes said

QCQVS: The Timoteo Circus is one of the best known of Chile's 120 circuses. **'All Chileans know Circo Timoteo, it's like an institution',** said Stéfano Rubio, a conductor and administrative manager of the circus.

SVQCQ: Ciotti had announced his party would not vote for either of the two motions of censure – meaning there would not be enough votes to stop the law. Reacting to the vandals, Ciotti tweeted: '**I will never give in to the new disciples of terror'.**

QCQ	Thomas had made the comments during a committee meeting on Thursday morning as she began asking St-Onge a second round of questions.
	'Minister, I noticed that you answer my questions in French, but other English questions you answer in English, if they're from your Liberal colleagues', Thomas said.
	'I realize it's completely your choice, we're a bilingual country, but if at all possible, I would love to have it in English'.

Heuristic quotes are a special type of direct speech that spans across multiple sentences. We developed a heuristic method that, when a direct quote was found, would also search back across sentences to find the beginning of the quote, so as not to limit search to within sentence boundaries. Heuristic quotes are of many syntactic types, which is probably why they are the second most frequent type of quote (see Table 5 below). They are typically floating quotes, as shown in the first example in (8), where a direct quote of type QCQSV is followed by a floating quote that spans four sentences. It is the latter that we capture as heuristic. Heuristic may also include the speaker and the reporting verb within the sentence, as we see in the second example in (8), where *Rich agrees* serves as SV to a quote that spans three sentences.

(8)	Heuristic, floating	'There's a lot of meetings that happen', Singh told Raj.
		'Our critics meet with the ministers on a regular basis, pushing for different elements of the agreement. We've got an oversight committee that works on it. We've got our House leaders and whips that work on it. And then I have my meetings with the prime minister'.
	Heuristic, SVQCQ	'This sounds like a Canadian edition of OpenAI', I suggest. Rich agrees: '**That is actually a pretty good way to think about it. Open AI as it used to be. Open AI when it was non-profit, before it became commercial'.**

Table 5. Types of quotes.

Quote type	Instances	%	Quote type (cont.)	Instances	%
SVC	1,190,862	39.49	VSC	1,211	0.04
Heuristic	404,119	13.40	VSQCQ	496	0.02
QCQSV	384,971	12.77	SCV	426	0.01
CSV	349,790	11.60	QSCQV	89	0.00
QCQ	303,814	10.07	VQCQS	14	0.00
QCQVS	150,592	4.99	SQCQV	4	0.00
According to	110,195	3.65	QCSQV	3	0.00
CVS	93,902	3.11	VQSCQ	2	0.00
SVQCQ	23,745	0.79	QVCQS	1	0.00
VCS	1518	0.05			

The last general type of quote that I will illustrate here are those that are introduced with *according to*. These also appear in a variety of syntactic patterns, with the prepositional phrase *according to* introducing the speaker and the quote either before or after the prepositional phrase (Example 9).

(9)	*according to*, before	According to a National Institute of Aging study published this month, **58 per cent of Canadians older than 50 experience loneliness.**
	according to, after	**Canada's inflation rate dropping to 2.8 per cent in June is a 'milestone moment' that Canadians should find some relief in,** according to Deputy Prime Minister and Finance Minister Chrystia Freeland.

Turning to the relative frequency of all these types, we see, in Table 5, that SVC (*The Prime Minister said that . . .*) is the prototypical and, indeed, the most frequent type of quote in news stories. Heuristic quotes appear second, perhaps not unsurprisingly, as they may be composed of many different patterns. What is relevant about the second place of heuristic quotes is that it means that a large amount of reported speech includes multiple sentences. After SVC and heuristic, frequent types include CSV, with a quote followed by subject and verb (*Quote, the Prime Minister said*) and floating quotes, where there is no reporting verb. Again, similar (but not identical) trends were found in non-Canadian news datasets (see Bednarek et al., 2024: 18), indicating a potential broader reach of such journalistic conventions.

It is interesting to note that, whereas in indirect speech the preferred pattern is SVC (*The Prime Minister said that quote*), in direct speech it is more common to place the quotation at the beginning, that is, QCQSV (*'Quote', the Prime Minister said*). The second most frequently type of direct speech is floating quotes, with no reporting verb. The third most frequent is SVQCQ, that is, *The Prime Minister said 'quote'*. After that, other orders of subject, reporting verb, and quote are vanishingly rare, with only a handful of cases (see Table 5).

The relative frequency of SVC versus QCQSV may indicate that reporters choose direct speech when they want to foreground the quoted content at the beginning of the sentence. In other words, the choice may not be between *The Prime Minister said that* and *The Prime Minister said 'quote'*, but rather between placing the quote after or before the quoting frame. When the quote works better after the quoting frame, then writers choose indirect speech. When the quote is placed before the quoting frame, then reporters choose direct speech.

Turning now to the comparison across genders, and using a smaller subset as above, where the gender of each quote could be clearly identified, we see that there is no difference between the two genders. The summary table (Table 6) shows that the proportions of each type are almost identical for men and women.

To investigate in/direct reported speech, and using a smaller subset of quotes where speakers were clearly identifiable by gender, I grouped the many types into three larger classes. The class of indirect speech includes all the types without a Q: SVC, CSV, CVS, VCS, VSC, SCV, as shown in Example (6) earlier. The second class includes all direct

Table 6. Types of quotes, by gender.

Quote type	Men		Women	
	Instances	%	Instances	%
SVC	252,988	41.02	127,487	39.79
QCQSV	156,210	25.33	80,509	25.13
CSV	110,231	17.87	60,781	18.97
Heuristic	43,453	7.05	21,809	6.81
QCQVS	26,897	4.36	15,485	4.83
CVS	17,043	2.76	10,042	3.13
SVQCQ	6906	1.12	2771	0.86
According to	2256	0.37	1165	0.36
VCS	320	0.05	137	0.04
VSC	292	0.05	101	0.03
VSQCQ	151	0.02	79	0.02
SCV	10	0.00	3	0.00
VQCQS	4	0.00	1	0.00
Total	616,761	100.00	320,370	100.00

Table 7. Relative frequency of direct, indirect, and *according to* quotes, by gender.

Reported speech type	Men		Women	
	Instances	%	Instances	%
Indirect speech	380,884	61.76	198,551	61.98
Direct speech	233,621	37.88	120,654	37.66
According to	2256	0.37	1165	0.36

speech, that is, all the types with Q surrounding the content, plus heuristic quotes, which are always direct speech. The third class includes all instances of *according to*. The relative frequency is shown in Table 7. We see that, overall, reporters seem to prefer indirect speech over other forms of quoting, and indicating that, again, there is no difference across genders in this general classification of quotes.

In general, the more frequent use of indirect speech offers an avenue to 'allow journalists to intertwine their own voices with news sources' expressions' (van Krieken and Sanders, 2019: 402). Van Krieken and Sanders, in a historical analysis of Dutch news stories, show that direct speech is becoming more frequent, perhaps as a way to untangle that mix of the reporter's and the source's voice (see also Vis et al., 2015). While direct speech may be increasing (although I do not have historical data for English), it is still the case that indirect speech is the more frequent type, especially if we include *according to* quotes as a form of indirect speech.

This preference for indirect speech over direct speech seems to agree with observations by Waugh (1995), who suggests that indirect speech is the unmarked form of

quotation in news stories, and with quantitative corpus analysis results by Semino and Short (2004), who found that news texts contain more indirect speech than other genres such as fiction or biography. It also aligns with very similar findings in a specialized corpus of US print news: It is remarkably close to that found by Garretson and Ädel (2008) in US newspapers during the 2004 presidential election, at 38% vs. 62% across several newspapers. This result suggests that linguistic conventions from older (print) news endure in contemporary digital news. It is also noteworthy that Table 7 shows that there appears to be no gender bias in contemporary Canadian digital news regarding credibility and reader engagement as (potentially) influenced by the use of indirect/direct speech.

Discussion: Why it matters

Bringing together analysis of sources (who is quoted) with analysis of reporting expressions and of quote types (how are they quoted) can provide a richer and more nuanced picture of gender representation in the news. Thus, this study offered insights into diverse rather than uniform usage: There was evidence of strong gender bias in the selection of sources, with women being quoted much less frequently than men. At the same time, there seemed to be no gender bias in the reporting verbs used or in the structure of quotation types. Such nuanced findings are important for initiatives that aim to decrease gender bias in the news, as they can inform us about particular issues that need to be addressed versus aspects that are, in fact, not problematic.

The most striking quantitative result of this data analysis is the lack of balance in gender representation in news stories. Sources are predominantly men, as the Gender Gap Tracker dashboard has shown since we started data collection in October 2018, with average percentages fluctuating between 27 and 32%. The causes, naturally, are not exclusively within news organizations themselves, as they have no control over who is elected to Parliament or appointed as CEO of a company. We saw this clearly during the peak of the Covid-19 crisis, when the percentage of women increased, mostly because the public health officers and ministers of health who were giving daily press briefings tended to be women, as women tend to be overrepresented in such roles (Taboada, 2020). News organizations and reporters, however, do exert control over some of their sources, particularly academics and experts, and some experiments have shown that keeping track of sources helps bring gender balance to news stories (British Broadcasting Corporation, 2024; Yong, 2018).

From a linguistic point of view, the analysis shows a relative lack of variety in reporting verbs, with the single verb *say* accounting for over 70% of the verbs chosen to introduce quotes. Reporting verbs tend to be in the content or factual class (*say, tell*), as opposed to evaluative verbs (*claim, argue*). This helps produce the impression of objectivity or neutrality. Although the structure of quoting frames and quotations shows more variation than the use of reporting verbs, the prototypical SVC structure is the most frequent and indirect speech is more frequent than direct speech, with no gender differences in quote types. One possible conclusion is that, although the most frequent reporting verbs are factual in nature, reporters still exert control over how they rephrase the content of quotes. In addition, the study's findings indicate that linguistic conventions from print news can still be found in digital news. The fact that there is no gender bias in quote type

is a positive finding that could not have been predicted in advance, and shows the advantage of using new tools to identify trends that would be difficult to analyze manually in large datasets.

I have, in this paper, analyzed one year's worth of data in English. The Gender Gap Tracker and the *Radar de parité* have such rich data that many other further analyses present themselves. In previous work, we have analyzed the types of sources present in English-language news, by classifying them into politicians, athletes, non-profit leaders, lawyers, experts, or witnesses (Asr et al., 2021), for the years 2018–2020. Further analyses would show whether the same trends continue, and whether they are similar in French-language media. Research by Calsamiglia and Ferrero (2003) examined the differences between academics showing caution in their reporting verbs, while the organizations being quoted were more assertive. Similar analyses could show differences across women and men and across different types of sources. Future research could also analyze the much larger GGT dataset, comprising data going back to October 1, 2018, with continuous updates for the foreseeable future. Examination of the data in French, where the proportion of women quoted in 2023 was similar, at 30%, would highlight potential differences with the English data.

The analyses presented here, and the larger goals of the Gender Gap Tracker, contribute to our understanding of news today, including news available in digital format. The push for more diversity in many of our public institutions leads us, first, to question assumptions about who is important and who needs to be present in news stories. It also pushes advances in corpus and computational projects such as this one, which in turn illuminate aspects of our news discourse.

Notes

1. 'Surrey Memorial Hospital doctors say meeting with minister was a start but 'not enough' without funding'. CBC News. May 31, 2023. https://www.cbc.ca/news/canada/british-columbia/surrey-memorial-hospital-doctors-meet-with-health-minister-1.6861319 (Accessed May 3, 2024).

2. https://informedopinions.org/
3. https://www.sfu.ca/discourse-lab
4. https://gendergaptracker.informedopinions.org/
5. https://radardeparite.femmesexpertes.ca/
6. https://github.com/sfu-discourse-lab/GenderGapTracker
7. https://www.atap.edu.au/tools/
8. Although longitudinal analyses are not the focus of this study, the overall proportion of women and men has seen a slight steady increase since the beginning of data collection in 2018; see Rao et al. (2021).
9. https://strategies.cbcrc.ca/en/equity-diversity-and-inclusion/
10. https://mediabiasfactcheck.com/national-post/

References

Asr FT, Mazraeh M, Lopes A, et al. (2021) The Gender Gap Tracker: Using natural language processing to measure gender bias in media. *PLoS ONE* 16(1): e0245533.

Bakhtin M (1986) *Speech Genres and Other Late Essays*. Austin: University of Texas Press.

Bednarek M (2006) *Evaluation in Media Discourse: Analysis of a Newspaper Corpus*. London: Continuum.

Bednarek M (2016) Voices and values in the news: News media talk, news values and attribution. *Discourse, Context & Media* 11: 27–37.

Bednarek M and Caple H (2012) *News Discourse*. London: Bloomsbury.

Bednarek M, Schweinberger M and Lee KK (2024) Corpus-based discourse analysis: From meta-reflection to accountability. *Corpus Linguistics and Linguistic Theory*. Epub ahead of print 16 April 2024. DOI: 10.1515/cllt-2023-0104.

Bell A (1991) *The Language of News Media*. Oxford: Blackwell.

Benson R and Wood T (2015) Who says what or nothing at all? Speakers, frames, and frameless quotes in unauthorized immigration news in the United States, Norway, and France. *American Behavioral Scientist* 59(7): 802–821.

British Broadcasting Corporation (2024) *50:50 The Equality Project*. Available at: https://www.bbc.co.uk/5050 (accessed 4 May 2024).

Caldas-Coulthard CR (1994) On reporting reporting: The representation of speech in factual and factional narratives. In: Coulthard M (ed.) *Advances in Written Text Analysis*. New York: Routledge, pp.309–322.

Calsamiglia H and Ferrero CL (2003) Role and position of scientific voices: Reported speech in the media. *Discourse Studies* 5(2): 147–173.

Carlin DB and Winfrey KL (2009) Have you come a long way, baby? Hillary Clinton, Sarah Palin, and sexism in 2008 campaign coverage. *Communication Studies* 60(4): 326–343.

Chafe W and Nichols J (1986) *Evidentiality: The Linguistic Coding of Epistemology*. Norwood, NJ: Ablex.

Clark HH and Gerrig RJ (1990) Quotations as demonstrations. *Language* 66: 764–805.

Gallagher M (2005) *Who Makes the News? Global Media Monitoring Project 2005*. London: World Association for Christian Communication.

García-Blanco I and Wahl-Jorgensen K (2012) The discursive construction of women politicians in the European press. *Feminist Media Studies* 12(3): 422–441.

Garretson G and Ädel A (2008) Who's speaking? Evidentiality in US newspapers during the 2004 presidential campaign. In: Ädel A and Reppen R (eds) *Corpora and Discourse: The Challenges of Different Settings*. Amsterdam: John Benjamins, pp.157–188.

Gibson R and Zillmann D (1993) The impact of quotation in news reports on issue perception. *Journalism Quarterly* 70(4): 793–800.

Goddard C and Wierzbicka A (2019) Direct and indirect speech revisited: Semantic universals and semantic diversity. In: Capone A, García-Carpintero M and Falzone A (eds) *Indirect Reports and Pragmatics in the World Languages*. Cham: Springer, pp.173–199.

Goodall H (2012) Media's influence on gender stereotypes. *Media Asia* 39(3): 160–163.

Goodyear-Grant E (2013) *Gendered News: Media Coverage and Electoral Politics in Canada*. Vancouver: University of British Columbia Press.

Hawkins-Gaar K (2019) Journalism has a gender representation problem. Bloomberg is looking for a solution. *Poynter*, 30 January.

Jullian PM (2011) Appraising through someone else's words: The evaluative power of quotations in news reports. *Discourse & Society* 22(6): 766–780.

Kassova L (2020a) *The Missing Perspectives of Women in COVID-19 News: A special report on women's under-representation in news media*. London: AKAS Consulting.

Kassova L (2020b) *The Missing Perspectives of Women in News: A report on women's under-representation in news media; on their continual marginalization in news coverage and on the under-reported issue of gender inequality*. London: AKAS Consulting.

Kassova L (2023) *From Outrage to Opportunity: How to Include the Missing Perspectives of Women of All Colors in News Leadership and Coverage*. London: AKAS Consulting.

Kelly J, Knight J, Peck LA, et al. (2003) Straight/narrative? Writing style changes readers' perceptions of story quality. *Newspaper Research Journal* 24(4): 118–122.

Lazaridou K, Krestel R and Naumann F (2017) Identifying media bias by analyzing reported speech. In: *2017 IEEE International Conference on Data Mining (ICDM)*, New Orleans, LA, USA, 18–21 November 2017. New York: IEEE, pp.943–948.

Lehrer A (1989) Between quotation marks. *Journalism Quarterly* 66(4): 902–941.

López Pan F (2010) Direct quotes in Spanish newspapers: Literality according to stylebooks, journalism textbooks and linguistic research. *Journalism Practice* 4(2): 192–207.

Lukin A (2013) What do texts do? The context-construing work of news. *Text & Talk* 33(4–5): 523–551.

Macharia S (2015) *Who Makes the News?: Global Media Monitoring Project*. World Association for Christian Communication, London and Media Monitoring Project, South Africa. Available at: https://whomakesthenews.org.

Macharia S (2020) *Who Makes the News? 6th Global Media Monitoring Project*. Global Media Monitoring Project. World Association for Christian Communication, London and Media Monitoring Project, South Africa. Available at: https://whomakesthenews.org.

Manning P (2001) *News and News Sources: A Critical Introduction*. London: Sage.

Martin JR and White PRR (2005) *The Language of Evaluation*. New York: Palgrave.

Matthews J (2012) Source attribution and perceptual effects. In: *Proceedings of the ISLC 2012: International symposium on language and communication*, Izmir, Turkey, 10–13 June 2012, pp.85–96.

McGregor WB (1994) The grammar of reported speech and thought in Gooniyandi. *Australian Journal of Linguistics* 14(1): 63–92.

McGregor WB (2019) Reported speech as a dedicated grammatical domain–and why defenestration should not be thrown out the window. *Linguistic Typology* 23(1): 207–219.

Morris M (2016) *Gender of Sources Used in Major Canadian Media*. Ottawa: Informed Opinions.

Niculae V, Suen C, Zhang J, et al. (2015) Quotus: The structure of political media coverage as revealed by quoting patterns. In: *Proceedings of the 24th International Conference on World Wide Web*, Florence, Italy, May 18–22, 2015, pp.798–808.

Nylund M (2003) Quoting in front-page journalism: Illustrating, evaluating and confirming the news. *Media, Culture & Society* 25(6): 844–851.

Power K, Rak L and Kim M (2019) Women in business media: A critical discourse analysis of representations of women in Forbes, fortune and Bloomberg Businessweek, 2015–2017. *Critical Approaches to Discourse Analysis Across Disciplines* 11(2): 1–26.

Rao P and Taboada M (2021) Gender bias in the news: A scalable topic modelling and visualization framework. *Frontiers in Artificial Intelligence* 4(82): 66477.

Rao P, Taboada M and Graydon S (2021) What we can learn from three years of data on the gender gap in news reporting. *Poynter*.

Ross K and Carter C (2011) Women and news: A long and winding road. *Media, Culture & Society* 33(8): 1148–1165.

Scollon R (1998) *Mediated Discourse as Social Interaction: A Study of News Discourse*. London: Longman.

Semino E and Short M (2004) *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing*. London: Taylor & Francis.

Shor E, van de Rijt A, Miltsov A, et al. (2015) A paper ceiling: Explaining the persistent underrepresentation of women in printed news. *American Sociological Review* 80(5): 960–984.

Short M (1988) Speech presentation, the novel and the press. In: van Peer W (ed.) *The Taming of the Text*. London: Routledge, pp.61–81.

Short M, Semino E and Wynne M (2002) Revisiting the notion of faithfulness in discourse presentation using a corpus approach. *Language and Literature* 11(4): 325–355.

Smirnova AV (2009) Reported speech as an element of argumentative newspaper discourse. *Discourse & Communication* 3(1): 79–103.

Spronck S and Nikitina T (2019) Reported speech forms a dedicated syntactic domain. *Linguistic Typology* 23(1): 119–159.

Sundar SS (1998) Effect of source attribution on perception of online news stories. *Journalism & Mass Communication Quarterly* 75(1): 55–68.

Taboada M (2020) The coronavirus pandemic increased the visibility of women in the media, but it's not all good news. *The Conversation*, 25 November.

Taboada M and Asr FT (2019) Tracking the gender gap in Canadian media. *The Conversation*, 3 February.

Taboada M and Chambers L (2020) *Who is Quoted and Who is Elected? Media Coverage of Political Candidates*. Richmond Hill, ON: Canadian Science Policy Centre.

Tannen D (1986) Introducing constructed dialogue in Greek and American conversational and literary narrative. In: Coulmas F (ed.) *Direct and Indirect Speech*. Berlin: Mouton, pp.311–332.

Thompson G. (1996) Voices in the text: Discourse perspectives on language reports. *Applied Linguistics* 17(4): 501–530.

Trimble L (2018) *Ms. Prime Minister: Gender, Media, and Leadership*. Toronto: University of Toronto Press.

Trimble L, Curtin J, Wagner A, et al. (2021) Gender novelty and personalized news coverage in Australia and Canada. *International Political Science Review* 42(2): 164–178.

Van der Pas DJ and Aaldering L (2020) Gender differences in political media coverage: A meta-analysis. *Journal of Communication* 70(1): 114–143.

van Dijk TA (1988) *News as Discourse*. Hillsdale, NJ: Lawrence Erlbaum.

van Krieken K (2020) Do reconstructive and attributive quotes in news narratives influence engagement, credibility and realism? *Journalism Studies* 21(2): 145–161.

van Krieken K and Sanders J (2019) Historical trends in the pragmatics of indirect reports in Dutch crime news stories. In: Capone A (ed.) *The Pragmatics of Indirect Reports: Socio-Philosophical Considerations*. Cham: Springer, pp.401–418.

Vis K, Sanders J and Spooren W (2015) Quoted discourse in Dutch news narratives. In: Lardinois A, Levie S, Hoeken H, et al. (eds) *Texts, Transmissions, Receptions*. Leiden: Brill, pp.152–172.

Ward LM and Grower P (2020) Media and the development of gender role stereotypes. *Annual Review of Developmental Psychology* 2: 177–199.

Waugh LR (1995) Reported speech in journalistic discourse: The relation of function and text. *Text & Talk* 15(1): 129–173.

White P (2000) Media objectivity and the rhetoric of news story structure. In: Ventola E (ed.) *Discourse and Community. Doing Functional Linguistics*. Tübingen: Gunter Narr, pp.379–397.

Yong E (2018) I spent two years trying to fix the gender imbalance in my stories. *The Atlantic*. 6 February.

Zelizer B (1989) 'Saying' as collective practice: Quoting and differential address in the news. *Text* 9(4): 369–388.

Author biography

Maite Taboada is Distinguished Professor of Linguistics at Simon Fraser University. Her research intersects discourse analysis and computational linguistics, with a focus on sentiment analysis, social media language, and misinformation.