



Marie Flesch*, Julie Abbou and Heather Burnett

Gender-inclusive speaking: a quantitative study of noun phrases referring to humans in a corpus of spoken French

<https://doi.org/10.1515/ling-2024-0105>

Received June 4, 2024; accepted August 23, 2025; published online January 7, 2026

Abstract: This contribution presents the first quantitative study of gender-inclusive language in spoken French, based on a corpus of sociolinguistic interviews conducted in Montreal, Paris, and Marseille with feminist and queer activists. Focusing on noun phrases referring to human beings, we quantify the proportion of generic masculines in the corpus and analyze the various strategies used by the speakers to avoid them. Results show that the overall rate of use of masculine-marked expressions whose intended referents are not necessarily men is extremely low (around 5 % of all the noun phrases in the corpus). We show that this result arises from three unrelated sources: differences between the spoken and standard written language, use of person-centered language, and replacement of so-called generic masculines with syntactic doublets.

Keywords: corpus linguistics; sociolinguistics; gender-inclusive language; person-centered language; feminism

1 Introduction

In the past decade, how the gender of human beings is represented in language, and what effects the linguistic encoding of social gender might have on society, has been a major topic of inquiry and discussion for both lay people and specialists (linguists, psychologists, and sociologists). Of course, this is not new: the role of language in producing and reproducing gender inequality has been a major focus for scholars and activists in the anglophone world, in francophone Canada, and in Sweden since the 1970s (see Pauwels 1998 for English; Arbour et al. 2014; Vachon-L'Heureux 1992

*Corresponding author: **Marie Flesch**, ATILF (UMR 7118), Université de Lorraine, Nancy, France,
E-mail: marie.flesch@univ-lorraine.fr. <https://orcid.org/0000-0003-2139-1517>

Julie Abbou, Department of Cultures, Politics, Society, University of Turin, Turin, Italy,
E-mail: julie.abbou@unito.it. <https://orcid.org/0000-0002-1958-6254>

Heather Burnett, Laboratoire de Linguistique Formelle (UMR 7110), CNRS - Université Paris Cité, Paris, France, E-mail: heather.burnett@cnrs.fr. <https://orcid.org/0000-0002-7389-0679>

for French Canada; and Hornscheidt 2003 for Swedish), and since the 1980s and 1990s in countries such as Germany (Bußmann and Hellinger 2003), Belgium (Arbour et al. 2014), and France (Burr 2003; Houdebine-Gravaud 1998; Yaguello 1979; among others). Nevertheless, the past ten years has seen renewed interest in how gender is encoded in language, and how languages might be changed to be more egalitarian. In some languages, like English, the primary interest is in how to include people whose gender does not fall into the binary categories male/female (for example, non-binary, gender fluid, or genderqueer individuals, see Baron 2020; Conrod 2018 for overviews). In other languages such as French, the recent attention to the inclusion of non-binary people is just a small part of a much larger revival of discussions around male bias in these languages and what should (not) be done about it. As Abbou et al. (2018) and Burnett and Pozniak (2021) describe, since 2017, there has been an explosion of discourse about French gender-inclusive language in traditional and social media, in academic publications, and in political, activist, and even legal circles. These debates concern not only how/whether to represent non-binary people in the language, but also whether/how so-called “generic” masculines should be replaced with non-masculine-marked forms. *So-called generic masculines* are masculine-marked expressions whose intended referents are not limited to men, e.g., using *les étudiants_M* to mean simply ‘the students’.¹

An important feature of the current intensity of “verbal hygiene” (Cameron 1995) on this topic in France and elsewhere in the francophone world is that it is almost exclusively focused on the written language. Indeed, the phrase *écriture inclusive* has become the dominant way of referring to a wide variety of feminist linguistic practices, replacing other older terms less tied to the written language such as *langage/communication inclusive/epicene/non-genré* ‘inclusive,epicene/non-gendered language/communication’, or *féménisation* ‘feminization’ (see Abbou 2023; Elmiger 2021b; Toma 2021, for reviews), and guidelines about how to talk in an inclusive or gender-neutral way take a backseat to how to write in one (see Crémier 2023). Likewise, the recent scientific literature on the relationship between grammatical and social gender in French has focused on how francophones use (Burnett and Pozniak 2021; Diaz Colmenares 2021; Diaz Colmenares and Heap 2020; Flesch and de Beaumont 2023; Simon and Vanhal 2022; among others) and interpret (Liénardy et al. 2023; Pozniak et al. 2023; Spinelli et al. 2023; Xiao et al. 2023; among others) gender-inclusive forms in writing only. The main goal of this paper is therefore to center the spoken language in a way that is not generally done when describing gender-inclusive strategies, either in

¹ In this paper, we refer to such “generic” masculines as so-called generic masculines, because (1) their reference is not always semantically generic, and (2) all the psycholinguistic literature on their interpretation (see Brauer and Landry 2008; Gygax et al. 2008; Richy and Burnett 2021; among many others) has shown that they have a male-biased interpretation even when they are used generically.

activist or in scientific discussions, and document the ways in which grammatical gender marking is expressed in speech.

Our investigation of spoken language referring to humans has the potential to yield surprising results that have consequences not only for language policy, but also for descriptive and theoretical linguistics. In the past 50 years (since at least Blanche-Benveniste 1986), the availability of large corpora of naturalistic speech has brought to light the ways in which varieties of French used on both sides of the Atlantic differ when they are written and spoken. Morpho-syntactic differences identified include the role of pronominal subjects (Auger 1994; Culbertson 2010; Liang 2023; Roberge 1990), the structure of negation (Barra-Jover 2004; Palasis 2013), the distribution of the subjunctive mood (Kastronic 2016; Poplack 1993), and the pronominal inventory (Blondeau 2001; Laberge and Sankoff 1979), among many others (Berrendonner 2004; Blanche-Benveniste 2007; Blanche-Benveniste et al. 1987; Deulofeu 2001). Indeed, some researchers even propose that francophones live in a situation of diglossia (Massot 2010; Palasis 2013; Rowlett 2011; among others), so great are the grammatical differences between the oral and written forms of the language. Therefore, this paper also investigates the descriptive question (which has theoretical implications) of whether there are important differences between spoken and written grammatical gender marking in French. In order to do so, we present a quantitative study of gender marking in noun phrases referring to humans in the *Cartographie linguistique des féminismes* (CaFé) corpus (Abbou and Burnett 2025), a corpus of 102 sociolinguistic interviews with feminist and queer activists in three large multicultural francophone cities: Paris, Marseille, and Montreal. We have chosen to study the speech of activists because these are people who are likely to be leading changes in the use of the grammatical gender system, which may then diffuse in francophone communities more generally.

The results of our study paint a picture of the use of the French grammatical gender system in oral language which contrasts with common assumptions made about written French, both in grammatical descriptions and in gender-inclusive language guides. We find that the overall rate of use of masculine-marked expressions whose intended referents are not necessarily men is extremely low: only at most around 5 % of all the noun phrases in the corpus (and 12.5 % of noun phrases with gender-neutral reference). This also contrasts with the proportion of generic masculines found in the rare studies which have quantified the proportion of various types of nouns in corpora of written data: 31 % of all nouns referring to gender-mixed groups in a corpus of Belgian news and political texts (Simon and Vanhal 2022), and 24.97 % of all nouns referring to humans in a corpus of German press texts (Müller-Spitzer et al. 2024). We show furthermore that the use of these so-called generic masculines is conditioned by social factors including age, geographical location, and education in such a way that, for younger, less educated speakers from Montreal, the

rate of using a so-called generic masculine is even lower: 3.89 %. We argue that these surprising results can be understood to arise from three different, unrelated, sources which, together, conspire to almost eliminate so-called generic masculines from the spoken language of feminist and queer activists:

1. Differences between the spoken and standard written language such that a lot of gender marking present in writing is neutralized in speaking in Paris, Marseille, and Montreal (e.g., *ami_M* and *amie_F* ‘friend’ are both pronounced [ami] in these three locations).
2. The introduction of *person-centered language*, following proposals by disability activists in anglophone North America, which replaces noun phrases referring to people headed by property nouns, which are often so-called generic masculines like *les sourds* ‘the deaf’ or *les noirs* ‘black people’, with a noun phrase headed by *personne* (i.e., *les personnes sourdes*; *les personnes noires*), which is grammatically feminine.
3. The replacement of isolated so-called generic masculines (e.g., *les étudiants*) with syntactic doublets (*les étudiants et étudiantes*), following proposals by influential gender-inclusive language guides.

Two of the three (socio)linguistic phenomena described above have no direct relation to feminist linguistic activism; however, as we will show in the paper, they combine with the doublets to render the spontaneous spoken language of feminist and queer activists highly gender inclusive. In fact, we will show that the contribution of the doublets, which currently occupy the most central place in the discourse around gender-inclusive language in French, to reducing the use of so-called generic masculines is far lower than the contribution of both person-centered language and, especially, the neutralizations created by written and spoken French grammar. We argue that our results have implications for gender-inclusive language policy not only for French but also for other languages: guidelines cross-linguistically are almost always formulated with the written language in mind; however, the spoken language can differ greatly from the written form, and is just as important when it comes to the production and reproduction of gender inequality. In order to be useful, language policies and guides should not neglect it. Our paper also makes a contribution to descriptive and theoretical linguistics: our results suggest that, in our corpus (and possibly also in other corpora of spoken French), masculines play a very minor role when it comes to accomplishing generic, gender-neutral, or gender-mixed reference, something that is at odds with many linguists’ descriptions of French grammatical gender, and on which they base morphological, syntactic, and semantic theories.

The structure of this article is as follows: in Section 2, we present a brief overview of the types of noun phrases in written and spoken French, as well as some of the

major proposals that have been made to render the written language more gender inclusive. In Section 3, we describe the CaFé corpus, and in Section 4, we present the results of our quantitative study of noun phrases referring to humans in this corpus. Section 5 concludes with a discussion of the implications of our findings for language policy and linguistic descriptions of French.

2 French grammatical gender and *écriture inclusive*

In languages that typologists describe as having grammatical gender, such as French, there is often some relation between the meaning of at least some of the nouns and whether they give rise to feminine or masculine agreement marking patterns. Broadly speaking, for nouns referring to humans, there are four cases in French. The first case comprises nouns whose basic semantic meaning incorporates gender, and their grammatical gender lines up with the social gender of the referent. Following Hellinger and Bußmann (2003), we call these nouns *gender-specific*, and this class includes words like *femme* ‘woman/wife’, *homme* ‘man’, *frère* ‘brother’, *sœur* ‘sister’ and *mari* ‘husband’. The second case consists of nouns whose basic semantic meaning includes people of all social genders; however, the noun itself has a single fixed grammatical gender. Terminology describing this second class varies greatly according to author and grammatical tradition; however, in this paper, we will call these nouns, such as *personne*, *parent*, *victime*, and *gens*, *epicene* nouns. The vast majority of French nouns referring to humans come in masculine-feminine pairs of two types. The first type consists of noun pairs where the masculine and the feminine have a different form, for example, *boulanger_M* – *boulangère_F* ‘baker’. The masculine form appears with masculine-marked dependents (*le beau boulanger*) and the feminine form appears with feminine-marked dependents (*la belle boulangère*). The second type consists of cases where there is a single form of the noun (e.g., *journaliste* ‘journalist’) but this form can appear with either masculine or feminine-marked dependents: *le beau journaliste* – *la belle journaliste*. In this paper, we will follow Corbett (1991) in calling nouns like *journaliste* “common gender” nouns. In almost all cases, the interpretations of nouns with feminine endings and common gender nouns in feminine-marked noun phrases are clear: feminine-marked noun phrases (almost) always refer to women. The interpretation of masculine-marked noun phrases, on the other hand, is more controversial.

In French grammars (see, for example, Grevisse and Goosse 2016), masculine-marked noun phrases referring to humans are described as allowing two types of reference, one referring specifically to men (specific masculine), the other to mixed-

gender groups or persons whose gender is irrelevant or unknown (generic masculines). Many descriptive and theoretical linguists also describe French masculine noun phrases in this way (see Atkinson 2012; Dubois and Lagane 1973; Ihsane and Sleeman 2016; Riegel et al. 1994; Schafroth 2003; among many others). In other words, the claim is that, while *la boulangère* or *la journaliste* can only be interpreted as referring to a woman, *le boulanger*, *les boulangers*, or *le journaliste* can be used to refer to men, women, or people of other genders.

The controversy revolves around whether the masculine marking on noun phrases such as *tous les citoyens* ‘all citizens’ or *tout citoyen canadien* ‘All Canadian citizens’ is biased towards a male interpretation; that is, are readers of these expressions more likely to think that they apply to men than women (or people of other genders)? In fact, many researchers working in a feminist and/or a psycholinguistic perspective have argued that these so-called “generic” uses of masculine noun phrases are in fact male-biased in this way. Some, for example Michard (1996, 1999), Michel (2016) among others, have highlighted naturally occurring written examples in which noun phrases that intuitively ought to have a gender-neutral interpretation are instead interpreted as male oriented. Others (Brauer and Landry 2008; Gygax et al. 2008; Richy and Burnett 2021) have conducted controlled experiments in which francophone participants read sentences containing gender-marked noun phrases, and show, using various response measures, that the masculine ones reliably trigger faster and easier male mental representations. Both of these lines of research have fueled feminist linguistic activism aiming to eliminate the so-called “generic” (yet male-biased) uses of masculine-marked noun phrases in favor of expressions that, depending on the political orientation of the activist, would either make women more visible or remove any inferences related to gender entirely.

In what follows, we present the major strategies present in francophone guidelines. As Elmiger, who has compiled the most comprehensive cross-linguistic database of gender-inclusive guidelines to date, notes, the proposals are as varied as they are numerous (Elmiger 2021a). For this reason, we will present only proposals that appear in the most influential guidelines in France and Quebec. For France, we follow Elmiger, who identified the 2015 (revised 2022) *Guide pratique pour une communication publique sans stéréotype de sexe*, published by the French government’s *Haut conseil d’égalité entre les femmes et les hommes* (Bousquet and Abily 2015/2022), and the 2016 (revised 2019) book *Manuel de l’écriture inclusive*, published by the communications agency *Mots-Clés*, which popularized the now ubiquitous term *écriture inclusive* (Haddad 2016/2019). For Quebec, we include the guidelines elaborated by the Université du Québec à Montréal (Lamothe et al. 1992), the guidelines written by the highly influential *Office québécois de la langue française* (OQLF) (2012, revised 2020), and the book *Manuel de grammaire non sexiste et inclusive* by Lessard and Zaccour (2018). Note that some of the Quebec references are

older: this reflects the fact that feminist linguistic activism is older in Canada than in France, being a product of Second Wave feminist interest in language in the 1970s and 1980s (Vachon-L'Heureux 1992).

– Composed forms

The idea is to replace a masculine that would be used to refer not only to men with an expression that combines aspects of the masculine and feminine forms of the noun, separated by some punctuation. *Étudiant* can thus be replaced by *étudiant-e*, *étudiant·e*, *étudiant.e*, or *étudiant(e)*. These forms can be inclusive of all gender identities, including those beyond the binary.

– Masculine – feminine doublets

The second strategy that is proposed by all the guides is to write both masculine and feminine forms of the nouns, either in a conjunction (*les étudiants et étudiantes*), disjunction (*les étudiants ou étudiantes*), or simply repeating them (*les étudiants, étudiantes*). In French, these are often described as *doublets*.

– Neutralization

A third strategy found in all the guides is the use of expressions where grammatical gender is neutralized. Under certain morphophonological conditions, it is impossible to tell whether a noun phrase is masculine or feminine. For example, while the common gender noun *journaliste*'s membership in the *masculine* class is visible in *le_M gentil_M journaliste* 'the nice journalist', if we wanted to say 'the incredible journalist', we would use the vowel-initial adjective *incroyable*, and the vowel in the determiner would be elided, making *l'incroyable journaliste*. Likewise, if we wanted to pluralize: since the French definite plural determiner *les* is not gender marked, the noun phrase *les journalistes* is not either.

– Epicene and collective nouns

Closely related to the neutralization strategy, the final strategy found in all the guides is to use an epicene noun, like *personne* or *gens*, or a noun referring to a collective. For example, instead of writing *le directeur pense que...*, one can write *la personne en charge du projet pense que...* or *la direction pense que...* 'Management thinks that...'. Collective nouns (e.g., *la foule* 'the crowd' or *le groupe* 'the group') are like epicene nouns in that they do show gender marking, but their semantic meaning is gender neutral.

The guidelines only occasionally make reference to the spoken language; however, when they do it is to predominantly discuss what place the composed forms, which make special use of punctuation, might have in speech. Bousquet and Abily (2015/2022) simply considers the composed forms irrelevant for speaking and proposes to use doublets. Other guides give instructions about how to pronounce the composed forms: Lamothe et al. (1992: 16) propose that if one is reading a composed form, one should read it as a doublet in which the masculine is read before the feminine, and Haddad (2016/2019) propose that composed forms should also be read as doublets starting with the feminine form.

3 The *Cartographie linguistique des féminismes* (CaFé) corpus

The CaFé corpus comprises 102 sociolinguistic interviews (of approximately 90 min each) with people who are engaged in what we described as “feminism, women’s issues and/or activism for queer and sexual rights” in Paris (42 interviews), Montreal (40 interviews), and Marseille (20 interviews), and in which we collected their positions on issues related to gender and sexuality, and their link with language. Two interviews were done with multiple people: one in Montreal (two people) and one in Paris (three people). The corpus was collected between 2021 and 2022, and consists of 168 hours of speech, transcribed with an enriched orthographic transcription (approximately 2 million words). For more details about the content and technical aspects of the corpus, see Abbou and Burnett (2025); however, we highlight here that many of the topics discussed were political and theoretical; therefore, they were conducive to talking about people and things *in general*, and so conducive to the production of noun phrases with generic or gender-mixed reference.

As is to be expected in a corpus collected from feminist activists, the majority of participants are women (cis or trans); however, it also includes a small minority of cis and trans men and non-binary people. Paris, Montreal, and Marseille are all large multicultural cities, and the composition of feminist spaces reflects this. Many of our participants are bilingual (French-Arabic, French-Kabyle, French-Portuguese, French-English, French-Spanish, among others), and we have chosen not to exclude the participants whose first language is not French, provided French is the language of their feminist engagement. As a result, we have three speakers who learned French after childhood: one whose native language is Spanish, one whose native language is Nigerian English, and one whose native language is Swedish.

The 105 recorded speakers were recruited on the basis of their engagement in feminist movements according to five categories:

- **Academics:** gender studies scholars, but also people who promote a feminist reading of science (scientists, doctors) or promote women in STEM (science, technology, engineering, and mathematics).
- **Professionals:** diversity practitioners, salaried community activists, lawyers, and instructors.
- **Associative:** activists, spokespeople, and others from associations, collectives, political parties, or unions.
- **Media:** authors, editors, librarians, translators, and journalists.
- **Collective:** volunteers in online activism, grassroots collectives, informal networks, etc.

The speakers in the corpus range from 19 to 83 years old, with a median age of 35 (SD = 22). In the analyses below, we divide the speakers into three age groups: participants under 35; participants over 60; and participants between 35 and 59. The speakers in CaFé are also highly educated, with all but nine having at least an undergraduate degree, and around a quarter of the corpus (27) having a PhD. The breakdown of participants by corpus according to age, education, and feminist engagement is shown in Table 1.

Table 1: Breakdown of age, education, and feminist engagement in the CaFé corpus.

	Marseille	Montreal	Paris	Total
Age group				
1 (under 35)	9	23	18	50
2 (35–59)	7	12	14	33
3 (60 and over)	4	6	12	22
Education (highest diploma)				
High school	3	5	1	9
Undergraduate	6	16	11	33
Master's	5	13	18	36
PhD	6	7	14	27
Engagement				
Associative	2	16	10	28
Collective	3	7	5	15
Media	5	6	11	22
Professional	4	3	9	16
Academic	6	9	9	24

4 Quantitative studies of gender-inclusive speaking

This section presents a quantitative study of grammatical gender marking on noun phrases referring to humans in the CaFé oral corpus. Of course, grammatical gender marking is present on other linguistic expressions in the language, particularly pronouns, adjectives, and participles; however, we have chosen to study noun phrases for practical reasons: first of all, studying gender marking on pronouns requires that, for each occurrence of the masculine and feminine pronouns *il* and *elle*, we identify whether its antecedent is human or non-human and, if human, what the social gender is. Given that there are 17,408 occurrences of *il(s)* and 7,017 occurrences of *elle(s)* in CaFé, it would be extremely time-consuming to process these data manually, and automatic identification of coreference chains in French is still a work in progress, despite recent advances (see e.g., Wilkens et al. 2020). The same observation holds for predicative adjectives and participles in morpho-syntactic agreement configurations.

Neologisms, like *iel* ‘they’ and *toustes* ‘all’, are certainly relevant to our central research question concerning gender marking and inclusive speaking. Non-metalinguistic uses of these expressions are found in our corpus; however, they are very infrequent: only 17 non-metalinguistic occurrences of *iel*, 13 of which are uttered by speaker *Paris 24*, the youngest speaker in the corpus (19 years old), and nine non-metalinguistic occurrences of *toustes*. We therefore limit our study to noun phrases referring to humans. The data and code (R notebook) used for the analyses are available on the OSF platform (https://osf.io/dc3v7/?view_only=2a294f3347764a89a04fc937b07330a).

4.1 Data extraction and coding

Part-of-speech annotation was performed in the TXM software program (Heiden et al. 2010) using the spoken French parameter of TreeTagger (Schmid n.d.). It allowed us to retrieve all nouns from the corpus. As some nouns were mislabeled, we also extracted all adjectives, past participles, and present participles. The dataset was then manually filtered in order to only retain nouns referring to humans. In a second step, we manually coded each human noun phrase for different categories based on the different strategies for gender-neutral communication described in Section 2. Crucially, our coding was always based on the *spoken* form of the noun phrase. The first category is masculine-marked expressions. We decomposed this under two subcategories: specific masculine noun phrases which, from context, we determined referred to men (such as those in (1)), and so-called generic masculine noun phrases,

for which it was possible, given the context, that the speaker was not referring only to men (such as those in (2)).

(1) a. *s'était inspiré du du meurtre euh d'un d'un jeune a- d'un ado américain qui s'était fait battre à mort* (Montreal 29)
'Was inspired by the murder of a young – an American teenager who got beaten to death'

b. *les sales affaires nous révèlent des comportements effectivement répréhensibles de la part d'enseignants et d'étudiantes très jeunes moins de vingt ans qui finalement consentent mais un drôle de consentement naturellement* (Paris 03)
'The sordid stories reveal to us behavior that is indeed reprehensible on behalf of teachers and very young students, less than 20 years old, who consented in the end, but naturally it was a weird consent'

(2) a. *mais quand même ce que j'ai en tête derrière c'est ces vieux académiciens croulants [...] qui sont d'un autre monde quoi* (Marseille 11)
'But actually what I have in mind are those old decrepit members of the Académie Française who are in another world'

b. *si plus de de blancs avaient pris la défense des esclaves noirs ça se serait peut-être passé plus vite* (Paris 14)
'if more white people had stood up for black slaves it might have happened faster'

In (1a), the noun phrase *un ado américain* 'an American teenager' is a specific indefinite referring to Matthew Shepard, a male victim of a homophobic hate crime, and *enseignants* 'teachers' is a masculine whose male interpretation emerges through the contrast with *étudiantes très jeunes* 'very young female students'. In contrast, *ces vieux académiciens croulants* in (2a) refers to the members of the Académie Française, six of whom were women at the time of the interview. Likewise, *blancs* in (2b) clearly refers to male, female (and maybe non-binary) white people. Partitioning the masculine-marked noun phrases into categories based on whether or not they exclusively refer to men requires knowing the speakers' intended referents, which we obviously do not always know with certainty. In the cases where it was not clear whether a speaker meant a masculine to refer only to men or not, we decided to err on the side of caution and classify it with the "so-called generic masculines". A consequence of this coding schema is that our results will possibly overestimate the proportion of so-called generic masculines in the corpus, but we prefer this to the alternative which is to miss some masculines that can include women and non-binary people.

The second main category that we coded for is feminines. Feminine-marked noun phrases almost always refer exclusively to women (such as those in (3));

however, there is one exception: *travailleuse(s) du sexe* ‘sex worker’. Some of the occurrences of *travailleuse(s) du sexe*, such as (4), seem to include male sex workers. *Travailleuse(s) du sexe* being the only example of a so-called generic feminine in our corpus, we group it with the other feminines in the overall results presented below.

(3) a. *ça a été génial d'ailleurs euh ma cheffe m'a laissé beaucoup de liberté genre euh quand euh il fallait faire des projets professionnels* (Marseille 06)
 ‘It was great, by the way, my boss gave me a lot of freedom, like when it came to doing professional projects’

b. *c'est devenu une très très bonne amie on s'est vue dimanche dernier d'ailleurs* (Paris 01)
 ‘She became a very, very good friend. In fact, we saw each other last Sunday’

(4) *juste pour les droits des travailleuses du sexe c'est malade ce qu'elle a fait* (Montreal 22)
 ‘Just for sex workers’ rights, it’s amazing what she did’

The third category of noun phrases that we distinguish are neutralizations: expressions where gender marking is neutralized (as in (5)). Consequently, these noun phrases can refer to people of all genders. These noun phrases are built around common gender nouns in particular morphophonological environments. Note that in other environments, common gender nouns show gender marking and are coded as masculine or feminine appropriately. For example, while *quelques camarades* ‘a few friends’ is coded as neutralization, *certains_M camarades* ‘some friends’ and *ta_F camarade* ‘your friend’ are coded as masculine and feminine respectively.

(5) *Je parle avec mes amies² anglophones pis ils ont des enjeux pis là mais c'est pas du tout du même ordre* (Montreal 30)
 ‘I talk with my anglophone friends and they have issues, but it’s not at all at the same level’

The fourth category is collective noun phrases. An example of a noun phrase coded as *collective* is shown in (6), where *rédac* ‘editorial board’ refers to the set of editors at a magazine.

(6) *plus le fait qu'au bout d'un moment la rédac aussi a un peu euh vieilli dans le sens où tu es toujours euh toute jeune mais voilà* (Paris 27)
 ‘Plus the fact that after a time, the editorial board had also aged a bit in the sense that you’re still very young but, you know’

² Here and in other examples from the CaFé corpus, we use the following inclusive spelling convention: the silent “e” in words in which gender marking is not audible in speech is greyed.

The fifth category is *epicene* noun phrases. The members of this category are noun phrases built around *personne* ‘person’ and *gens* (7), but also *parent* ‘parent’, *bébé* ‘baby’, *victime* ‘victim’, *témoin* ‘witness’, and others.

(7) a. *alors pour moi les ennemis c'est vraiment les vraiment les gens au pouvoir euh les les politiques au pouvoir* (Paris 39)
 ‘So for me, enemies are really people in power, uh politicians in power’

b. *bah je sais pas il y a des personnes qui sont complètement aveugles aux questions de race ou de classe* (Paris 33)
 ‘Bah I don't know, there are people who are completely blind to questions of race or class’

Finally, we also distinguished a category of *doublets*, where a whole determiner phrase can be doubled (8), or some of its subparts (9).

(8) a. *j'ai commencé à rencontrer des wikipédiens des wikipédiennes dans la vraie vie* (Paris 34)
 ‘I started meeting people who work on Wikipedia in real life’

b. *Ça dépend qui est mon interlocuteur ou mon interlocutrice* (Paris 36)
 ‘That depends on who my interlocutor is’

(9) a. *je parle beaucoup avec les copains copines francophones et notamment français de France* (Montreal 37)
 ‘I talk a lot with francophone friends, and notably French people from France’

b. *Tsé parce qu'un ou une allié e va se faire rattacher tsé aux luttes plus grandes* (Montreal 34)
 ‘You know because an ally will become attached, you know, to larger fights’

Subparts of nouns, particularly the endings, may also be doubled, via coordination (10a), repetition (10b), or by the formation of “long form” nouns incorporating both endings (10c).

(10) a. *je sais que si j'utilise certains mots ou certains arguments ça va braquer mon interlocuteur ou trice* (Paris 10)
 ‘I know that if I use certain words or certain arguments it's going to turn off my interlocutor’

b. *pis les formes abrégées style euh constructeur -trice* (Montreal 01)
 ‘and abbreviated forms like uh uh builder’

c. *on co-écris avec euh mon amie dont je te parlais euh qui est travailleuse sociale* (Montreal 09)
 ‘I co-wrote with my friend who I was telling you about, who is a social worker’

4.2 Overall descriptive results

We extracted 29,990 noun phrases referring to humans from the corpus. We found 942 metalinguistic uses and 158 noun phrases in languages other than French, that were excluded from the statistical analyses. Of the 28,890 remaining noun phrases, 10,858 (37.58 %) are feminine only, 6,367 (22.04 %) are epicenes, 4,921 (17.03 %) are gender-neutralized, 5,647 (19.55 %) are masculine, 762 are collectives (2.64 %), and 335 (1.16 %) are doublets. Crucially, it turns out that most of the masculine-marked noun phrases ($n = 3,985$ or 70.58 % of all masculine noun phrases) are used to refer to a man or to men only. This distribution gives us the first main result of our paper: so-called generic masculine noun phrases account for only 5.75 % ($n = 1,661$) of all noun phrases referring to humans in the corpus. If we add up all the relevant expressions that can be used to accomplish generic, gender-neutral, or gender-mixed reference (so-called generic masculines, epicene noun phrases, neutralizations, and doublets), the percentage rises only to 12.5 %. For information, Table 2 shows the ten most frequent uses of masculine-marked noun phrases in the corpus to possibly refer to people other than men. Note that our coding scheme is very generous to so-called generic masculines: as soon as non-male reference did not create a contradiction in the context, we coded a masculine as “so-called generic”. It seems probable to us that most of the occurrences of *violeur* and *agresseur* are made with men in mind; however, since the context in which they occurred did not clearly specify this, we coded them as “so-called generic”. In this way, the result that only 12.5 % of noun phrases with generic or gender-neutral/mixed reference are masculine should be taken as an upper-bound: in reality, there are probably even fewer truly generic or gender-neutral/mixed masculines in the corpus.

This result is noteworthy because, as discussed in Section 2, the question of what to do with masculines is generally treated as one of the fundamental, hotly debated issues in French feminist linguistic activism. To better understand why the proportion of so-called generic masculines is so small in our spoken corpus, we will now examine how the distribution of gender-marked expressions referring to humans varies according to city.

Figure 1 represents the proportion of noun phrases per grammatical gender category in the three subcorpora: Montreal (10,466 noun phrases), Paris (12,959 noun phrases), and Marseille (5,465 noun phrases). It shows that the proportion of masculine and feminine noun phrases is lower in Montreal than in other cities, but that the proportion of epicene noun phrases is the highest in Montreal (27.77 % of noun phrases, versus 19.03 % in Marseille, and 18.68 % in Paris).

As described in Section 3, the main topic of the interviews is feminist and queer activism, so it is not surprising that all the speakers frequently refer to women and

Table 2: Ten most frequent nouns in generic masculine noun phrases.

Montreal			Paris			Marseille		
Noun	Freq.	%	Noun	Freq.	%	Noun	Freq.	%
<i>étudiant</i>	42	8.96	<i>étudiant</i>	63	7.38	<i>étudiant</i>	54	15.98
‘student’			‘student’			‘student’		
<i>enfant</i>	21	4.48	<i>violeur</i>	29	3.40	<i>sourd</i>	15	4.44
‘child’			‘rapist’			‘deaf person’		
<i>expert</i>	20	4.26	<i>militant</i>	28	3.28	<i>enfant</i>	12	3.55
‘expert’			‘activist’			‘child’		
<i>policier</i>	18	3.84	<i>enfant</i>	25	2.93	<i>expert</i>	10	2.96
‘police officer’			‘child’			‘expert’		
<i>ennemi</i>	16	3.41	<i>travailleur</i>	23	2.69	<i>ennemi</i>	9	2.66
‘enemy’			‘worker’			‘enemy’		
<i>blanc</i>	14	2.99	<i>expert</i>	19	2.22	<i>agresseur</i>	8	2.37
‘white person’			‘expert’			‘assaulter’		
<i>enseignant</i>	14	2.99	<i>agresseur</i>	18	2.11	<i>militant</i>	8	2.37
‘teacher’			‘assaulter’			‘activist’		
<i>québecois</i>	13	2.77	<i>avocat</i>	17	1.99	<i>bourgeois</i>	7	2.07
‘quebecker’			‘lawyer’			‘bourgeois’		
<i>imposteur</i>	12	2.56	<i>acteur</i>	13	1.52	<i>copain</i>	7	2.07
‘impostor’			‘actor’			‘friend’		
<i>militant</i>	10	2.13	<i>client</i>	13	1.52	<i>blanc</i>	6	1.78
‘activist’			‘client’			‘white person’		

therefore have a high rate of feminine-marked noun phrases. What is a bit more unexpected is that both epicene and gender-neutralized noun phrases appear more frequently than both gender-specific masculines and so-called generic ones. In the next sections, we examine both of these categories in detail, starting with the neutralizations.

4.3 The role of pronunciation in neutralizations

As discussed in Section 2, common gender nouns (like *journaliste* ‘journalist’, *artiste* ‘artist’, *astronome* ‘astronomer’, *professeur* ‘professor’, etc.) often give rise to noun phrases in which grammatical gender is neutralized – i.e., in which it is impossible to tell whether the NP is masculine or feminine. In the corpus, there are 6,561 noun phrases formed with “common gender” nouns. As expected, the vast majority of them (4,909, or 74.82 %) are gender neutralized; 815 (12.42 %) are feminine, and 837 (12.76 %) are masculine. Most gender-neutralized common gender noun phrases (60.43 %) are in the plural form, which is, again, expected because French plural determiners/

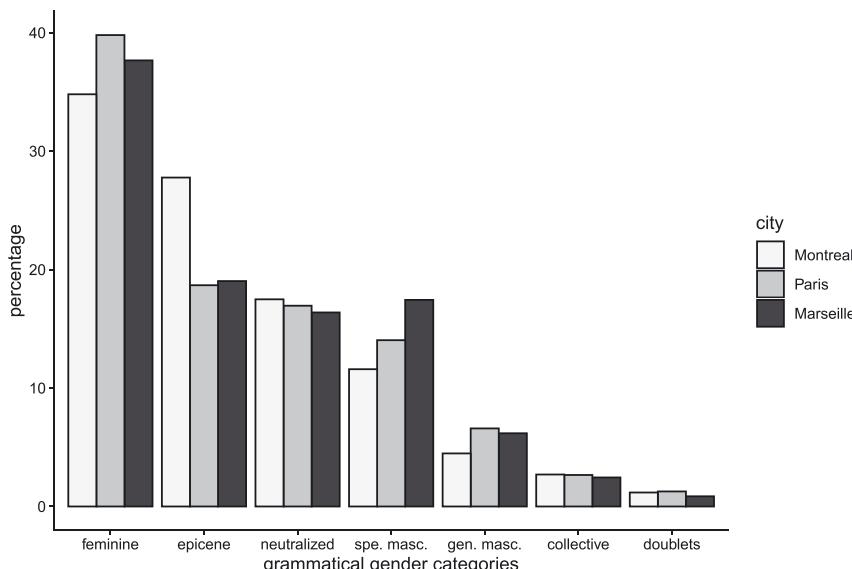


Figure 1: Distribution of the types of noun phrases referring to humans in CaFé.

demonstratives (*les, des, ces, mes*, etc.) are not marked for gender. Neutralization also seems to be correlated with the structure of the first syllable of common gender nouns. Of the 1,153 singular nouns beginning with a vowel, 62.27 % are gender neutralized, versus 45.11 % of 1,443 singular nouns beginning with a consonant. This is due to, among other things, the deletion of the vowel of the definite determiner before vowel-initial nouns or adjectives (i.e., *le + artiste* = *l'artiste* = *la + artiste*). In the plural, 93.71 % of the 1876 noun phrases with common gender nouns starting with a vowel are neutralized, versus 85.30 % of the 2,089 nouns that start with a consonant.

As we have described, written French has a robust class of common gender nouns; however, this class is much larger in the spoken language. For example, many nouns that distinguish masculine from feminine by a final *-e* in writing are now part of the common gender noun class in speech. Examples include *ami* – *amie* ‘friend’ and *ennemi* – *ennemie* ‘enemy’, *professeur* – *professeure* ‘professor’, and *auteur* – *auteure* ‘author’, among many others.

In CaFé, we find that 33.88 % (n = 1,667) gender-neutralized noun phrases have expressed gender marking in writing but not in speech; they are formed most frequently with *amie* (n = 508), *ennemie* (n = 287), and *alliée* (n = 247). When looking at the referents of these 1,667 noun phrases, we find that, in writing and without the use of gender-inclusive spelling and taking into account the discursive context in which

the NP was uttered, 84.58 % (n = 1,410) would be written as (generic) masculines, 0.18 % (n = 3) as (specific) masculines, and 15.24 % (n = 254) as feminine noun phrases. It means that if our corpus were written in non-inclusive writing, the proportion of generic masculines in the corpus would thus be 10.63 %, almost double the 5.75 % we found by only taking into account spoken forms.

4.4 The role of person-centered language in epicenes

As discussed in Section 4.1, speakers in the Montreal corpus have both the lowest rates of so-called generic masculines and the highest rates of epicenes. Table 3, which shows the list of epicene nouns in each subcorpus, gives some insight into how speakers in Montreal differ from their French counterparts.

Although there is much overlap between the lists from Paris, Montreal, and Marseille, one difference immediately stands out: speakers in Montreal use the noun *personne* much more often (58.26 % of epicene nouns) than speakers in Paris and Marseille, who use *personne* about as often as *gens*. Rather than being a coincidence, we hypothesize that part of the reason that the Montreal speakers are using fewer so-called generic masculines than the speakers from France is that they have a higher rate of use of a construction that is a variant of a different sociolinguistic variable, unrelated to gender: *person nouns* versus *property nouns*. In particular, we find examples in which, to describe a person or a group of people, speakers in Montreal construct a NP composed of the noun *personne* (and to a lesser extent *gens*) and a modifier describing a property, while speakers in France use noun phrases which directly describe the property. Some examples of *person noun* versus *property noun* pairs in the corpus are shown in (11)–(13).

Table 3: Five most frequent epicene nouns, by city (*personne* ‘person’, *gens* ‘people’, *parent* ‘parent’, *personnage* ‘character’, *victime* ‘victim’, *grand-parent* ‘grandparent’).

Montreal			Paris			Marseille		
Noun	Freq.	%	Noun	Freq.	%	Noun	Freq.	%
<i>personne</i>	1,693	58.26	<i>gens</i>	1,213	50.1	<i>personne</i>	485	46.63
<i>gens</i>	1,012	34.82	<i>personne</i>	990	40.89	<i>gens</i>	454	43.65
<i>parent</i>	152	5.23	<i>parent</i>	147	6.07	<i>parent</i>	67	6.44
<i>personnage</i>	12	0.41	<i>victime</i>	35	1.45	<i>personnage</i>	18	1.73
<i>victime</i>	11	0.38	<i>grand-parent</i>	13	0.54	<i>grand-parent</i>	6	0.58

(11) a. *homo ça devrait être pour qualifier et les personnes gays et les personnes lesbiennes* (Montreal 09)
“Homo” should be to qualify gay people and lesbian people’

b. *enfin le truc c'est que les lesbiennes j'ai l'impression que c'est toujours les meilleures alliées des gays et les gays ils nous laissent tomber tout le temps* (Paris 17)
'Actually the thing is I have the impression that lesbians are always the best allies of gay people, and gay people let us down all the time'

(12) a. *en plus des personnes juives ou identifiées comme telles c'est euh les personnes tsiganes et roms et les personnes musulmanes ou identifiées* (Paris 32)
'More and more Jewish or identified people, it's uh Romani and Gypsy people and Muslim people'

b. *je compare euh les chrétiens euh les musulmans euh les juifs euh les confin Confucius, Bouddha, euh voilà* (Marseille 10)
'I compare Christians, Muslims and Jews, euh the Con- well Confucious, Buddha, there you go'

(13) a. *la police n'est pas une réponse pour les surtout les personnes immigrantes et racisées* (Montreal 39)
'The police is not an answer, especially for immigrant and racialized people'

b. *de plus en plus les collectifs de féministes qui sont pas nécessairement racisés ou qui ou où les racisés sont minoritaires* (Paris 36)
'More and more feminist collectives that are not necessarily racialized or where racialized people are in the minority'

The opposition between a description of a person (or a group of people) based around a property that they have and one which describes them first as a *person* and then states the property is reminiscent of the oppositions that are at play in the *Person-centered* (or *person-first*) *language* movement. In the 1990s, in the wake of the passage of the *Americans with Disabilities Act* (1990) in the United States, a number of important publications, including one from the American Psychological Association (1992), argued that “using labels [property nouns] to define people, which had long been used and was widely accepted, resulted in increased stigma in the medical, legal, and social realms” (Granello and Gibbs 2016: 33). The proposed solution was to form an NP around *person* or *people*, and describe their disability with a post nominal modifier, such as *the disabled* instead of *people with disabilities*. In English, there is an additional variant which contains a *person* noun: *disabled people*. This variant is known as the *identity-centered language* option because, according to its

advocates, the prenominal adjective communicates that the property described in the adjective is part of the person's identity in a way that the “person first” variant (*people with disabilities*) does not (see Andrews 2019 for a review). Proponents of the *identity-centered* variant draw a parallel between noun phrases describing people with disabilities and noun phrases describing people with other cultural or social properties: in the same way that no one finds *French people* offensive, the same holds for *disabled people*, under the disability-as-cultural-experience view. Although the *property noun* variant (*the disabled*) is universally disfavored by disability activists, whether the person first (*people with disabilities*) or identity first (*disabled people*) variants are favored depends on a variety of factors including age, gender, country of residence, and, above all, which particular disability community is being studied (Sharif et al. 2022; Vivanti 2020; among others). The person- and identity-centered language movements have not been very influential in France; however, they have become mainstream in French Canada. For example, the OQLF has both *langage centré sur la personne* and *langage centré sur l'identité* in its *Grand dictionnaire terminologique*,³ and the use of one of these variants (or both) is frequently recommended in guidelines for institutions and organizations, such as the Centrale des syndicats du Québec.

Although person/identity-centered language comes from disability activism, the strategy of avoiding using a property noun to describe people in minorities or oppressed groups has become very general, albeit variable in the CaFé corpus, as shown in (11)–(13).

In order to test the hypothesis that Montreal speakers use more person/identity-centered language, and that this affects the distribution of gender marking in their speech, we conducted a quantitative study of variation in the use of person nouns versus property nouns. In the corpus, we found 785 occurrences of epicene nouns followed by an adjective or by a prepositional phrase, which could be considered direct alternatives to nouns (regardless of their type: feminine, masculine, or neutralized). *Les personnes blanches* ‘white people’ can be considered a direct alternative to *les blancs* or *les blanches*, for example. Likewise, *une personne en situation d'itinérance* ‘person experiencing homelessness’ can be considered an alternative to *un itinerant* ‘a homeless person’. The vast majority of these people noun phrases are formed with *personne(s)* (n = 751); others are formed with *gens* (n = 32) and *individu(s)* (n = 2). These people noun NPs are distributed into 103 different types; the 20 most frequent are displayed in Table 4, with their frequency in the corpus and their proportion in all people noun NPs. Many of these people noun NPs are related to gender and sexuality (*personne trans, non-binaire, queer, lgbtq*,

³ <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26558380/langage-centre-sur-lidentite>;
<https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26558379/langage-centre-sur-la-personne>.

Table 4: Twenty most frequent “people noun” NPs in the corpus.

NP	Freq.	%	NP	Freq.	%
personne(s) trans	183	23.31	personne(s) handicapée(s)	10	1.27
personne(s) non-binaire(s)	115	14.65	personne(s) immigrante(s)	10	1.27
personne(s) racisée(s)	99	12.61	personne(s) autochtone(s)	9	1.15
personne(s) queer	45	5.73	personne(s) homosexuelle(s)	9	1.15
personne(s) blanche(s)	37	4.71	personne(s) militante(s)	9	1.15
personne(s) noire(s)	33	4.2	personne(s) cis	8	1.02
personne(s) lgbtq	14	1.78	personne(s) lesbienne(s)	8	1.02
personne(s) féministe(s)	13	1.66	personne(s) intersexé(e)s	7	0.89
personne(s) hétéro	13	1.66	personne(s) marginalisée(s)	7	0.89
personne(s) bisexuelle(s)	10	1.27	personne(s) gay	6	0.76

hétéro, etc.), to race (*personne racisée, blanche, noire, autochtone*), and to disability (*personne sourde, handicapée, autiste, malentendant(e)*). Most people nouns (78.73 %, $n = 618$) are in the plural form; 21.27 % ($n = 167$) are in the singular form.

In order to explore social variation in the use of people nouns versus property nouns, we created a subset based on all “properties” (*trans, non-binaire, racisé*, etc.) expressed by both people nouns and property nouns in the corpus. It contains 64 properties and 4,714 noun phrases: 703 people nouns and 4,011 property nouns (including 2,020 neutralizations, 1,433 feminine noun phrases, 517 masculine noun phrases, and 41 doublets). Most nouns are in the plural form (2,920, with 2,368 property nouns and 552 people nouns). There are 1,794 nouns in the singular form (1,643 property nouns and 151 people nouns). Figure 2 shows the observed probabilities of people nouns versus property nouns in each subcorpus and for each age group. It suggests an effect of both factors. The probability of producing people nouns is higher in Montreal than in Marseille and Paris, especially for the two younger groups. Moreover, age seems to be negatively correlated with the probability of people nouns, with a decrease in probability from one age group to another.

We created a mixed-effects logistic regression model with the lme4 R package (Bates et al. 2015), with speaker and property as random effects, and city, age, education, engagement, and number as fixed effects. The model predicts the production of people nouns. The reference levels of predictors are Montreal, older speakers, high school graduates, speakers with associative activities, and the singular form of nouns. The results are displayed in Table 5. The model shows a significant effect of city: the probability of using people nouns is lower in Marseille ($OR = 0.48, p = 0.018$) and Paris ($OR = 0.43, p = 0.001$) than in Montreal. Age also has a partial effect: the two youngest groups use more people nouns versus property nouns than older speakers

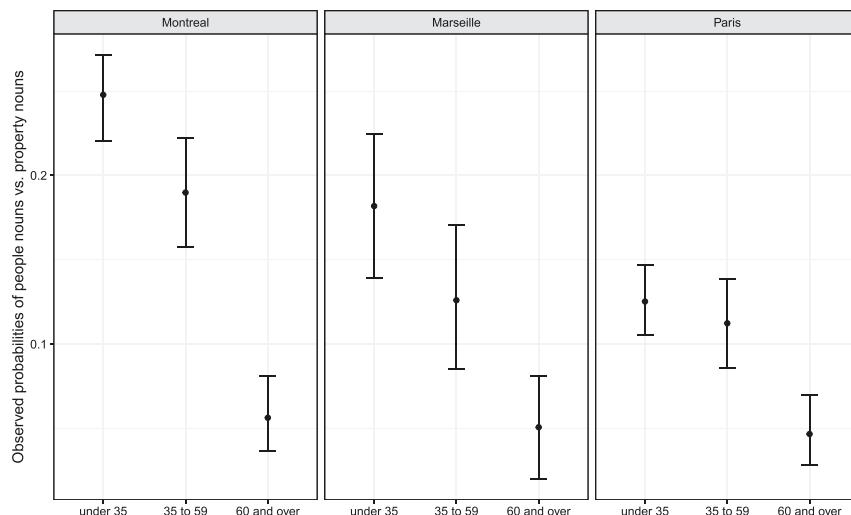


Figure 2: Observed probabilities of people nouns versus property nouns, by city and age.

Table 5: Mixed-effects logistic regression model 1 (boldface indicates statistically significant p-values).

Predictors	Odds ratios	CI	p
(Intercept)	0.05	0.02–0.17	<0.001
City: Marseille	0.48	0.26–0.88	0.018
City: Paris	0.43	0.26–0.72	0.001
Age: 19 to 34 y.o.	7.21	3.71–14.02	<0.001
Age: 35 to 59 y.o.	4.94	2.51–9.71	<0.001
Education: Undergraduate	1.05	0.45–2.44	0.916
Education: Master's	0.85	0.37–1.96	0.703
Education: PhD	0.41	0.15–1.14	0.087
Engagement: Collective	0.46	0.23–0.93	0.032
Engagement: Media	0.56	0.29–1.08	0.086
Engagement: Professional	0.95	0.48–1.90	0.895
Engagement: Academic	0.92	0.45–1.87	0.815
Number: plural	2.53	1.89–3.40	<0.001

($OR = 7.21$ and $p < 0.001$ for 19–34 y.o.; $OR = 4.94$ and $p < 0.001$ for 35–59 y.o.). For the engagement factor, there is a significant difference between the associative groups and the collective group ($OR = 0.46$, $p = 0.032$). Finally, the model shows that the probability of using people nouns is higher for plural noun phrases than for singular noun phrases ($OR = 2.53$, $p < 0.001$).

A priori, politically, the use of a person noun instead of a property noun in order to emphasize the referent's "personhood" is orthogonal to gender-inclusive language; however, grammatically, the fact that *personne* is feminine means that switching to a noun phrase built around this noun from a non-common gender property noun will have the effect of decreasing the so-called generic masculines. If all epicene people nouns in the subset ($n = 702$) were replaced by generic masculines, the overall proportion of generic masculines in the corpus would be 8.18 % ($n = 2,364$), instead of 5.75 %.

We therefore conclude that one of the main reasons that Montreal speakers use fewer so-called generic masculines than their French counterparts is that they are greater users of person- and identity-centered language.

4.5 The role of feminist engagement in doublets

As mentioned in Section 2, much of the discourse surrounding current feminist linguistic practices focuses on doublets and composed forms, which are pronounced as doublets in speech. Although integrating doublets into their speech is perceived as requiring little effort for some speakers, other speakers in CaFé describe having to train themselves to use these forms systematically (14).

(14) a. *ça vient vite il y a quand même forcément effectivement au départ à faire un tout p -un petit effort c'est-à-dire quel que s- même si vous êtes convaincue à cent mille pour cent qu'il faut utiliser le féminin et le masculin ce qui est mon cas euh il y a des telles habitudes voire habitus langagiers euh qu'il faut cognitivement euh défaire certaines euh réflexes de de parole pour en installer euh pour en installer d'autres mais ça vient très moi je trouve que ça vient très vite* (Paris 14)

'It comes quickly. There is indeed necessarily at the beginning a small effort, that is, even if you're 100 000% convinced that you need to use the feminine and the masculine, which is my case, there are so many linguistic habits, or habitus, that we must cognitively deconstruct certain speech reflexes to put in other, but it comes very – I find that it comes very quickly'

b. *du coup je m'efforce de le faire à l'oral je le fais pas tout le temps parce que j'oublie* (Paris 24)

'So I force myself to do it in speech. I don't do it all the time because I forget'

In order to investigate whether these impressions translate into production patterns, we conducted a quantitative study of variation between so-called generic masculines

Table 6: Ten most frequent pairs in doublets in CaFé.

Nouns	Frequency	%
<i>étudiant/étudiante</i> ‘student’	58	17.31
<i>travailleur/travailleuse</i> ‘worker’	37	11.04
<i>auteur/autrice</i> ‘author’	22	6.57
<i>copain/copine</i> ‘buddy’	20	5.97
<i>chercheur/chercheuse</i> ‘researcher’	17	5.07
<i>militant/militante</i> ‘activist’	17	5.07
<i>politicien/politicienne</i> ‘politician’	9	2.69
<i>colleur/colleuse</i> ‘street activist’	6	1.79
<i>expert/experte</i> ‘expert’	6	1.79
<i>interlocuteur/interlocutrice</i> ‘interlocutor’	6	1.79

and doublets. The 335 doublets found in the corpus are formed with 86 different pairs of nouns (or in some case a single noun). The ten most frequent pairs are shown in Table 6. Most doublets are formed with nouns that have a different form in the masculine and in the feminine in speech (*travailleurs et travailleuses* ‘workers’, *des spectatrices et des spectateurs* ‘spectators’), but some are formed with neutralized nouns, with a determiner or adjective (*un ou une allié* ‘ally’, *certain certaines amies* ‘certain friends’). Since the doublets are presented as ways to directly replace so-called generic masculines, we study the probability of using doublets versus these masculines.

The dataset used in this analysis contains all generic masculine NPs ($n = 1,661$) and all doublets found in the corpus ($n = 335$). The probabilities of using a doublet compared to a so-called generic masculine according feminist engagement is shown in Figure 3. It suggests that academics use more doublets than other people in other categories of feminist engagement.

We created the model with the same random effects and predictors as model 1. Reference levels of predictors are Montreal, older speakers, high school graduates, academics, and singular NPs. The model (Table 7) predicts the probability of doublets. There is an effect of city, with the probability of producing doublets versus generic masculine NPs being lower in Marseille than in Montreal ($OR = 0.30, p = 0.007$). The effect of education is partial, with people with only undergraduate diplomas producing more doublets than those with only high-school diplomas ($OR = 5.90, p = 0.023$). The type of engagement is also significantly linked to the probability of doublets, which is lower in three groups (association, media and professional) than in the academic group. The effect of age is partially significant: younger speakers (19–34 y.o.) use more doublets than the oldest group ($OR = 2.38, p = 0.44$). There is no significant effect of grammatical number.

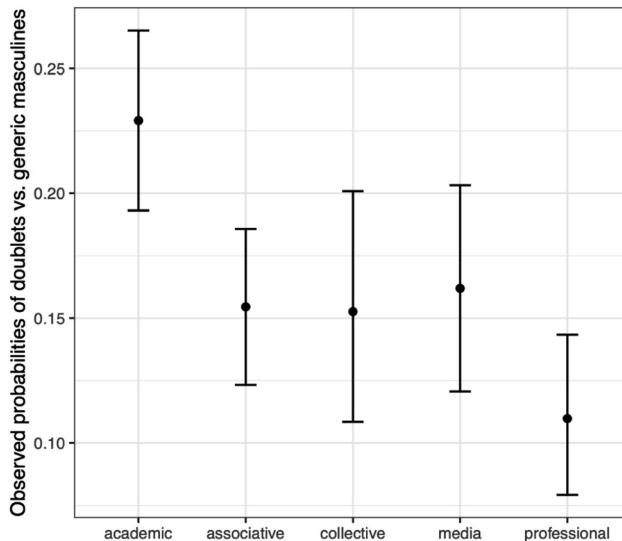


Figure 3: Observed probabilities of doublets versus generic masculine, per type of engagement.

Table 7: Mixed-effects logistic regression model 2. (boldface indicates statistically significant p-values)

Predictors	Odds ratios	CI	<i>p</i>
(Intercept)	0.03	0.00–0.18	<0.001
City: Marseille	0.30	0.12–0.72	0.007
City: Paris	0.50	0.25–1.01	0.053
Age: 19 to 34 y.o.	2.38	1.02–5.55	0.044
Age: 35 to 59 y.o.	2.03	0.87–4.72	0.101
Education: Undergraduate	5.90	1.28–27.31	0.023
Education: Master's	3.05	0.67–14.00	0.151
Education: PhD	1.90	0.35–10.29	0.458
Engagement: Association	0.26	0.10–0.71	0.008
Engagement: Collective	0.33	0.10–1.05	0.061
Engagement: Media	0.30	0.11–0.86	0.025
Engagement: Professional	0.24	0.08–0.71	0.010
Number: Plural	1.00	0.64–1.57	0.992

Once again, we find that Montreal activists and younger speakers are leaders when it comes to avoiding so-called generic masculines (although there is no significant difference between speakers aged 35–59 and the oldest speakers).⁴ Feminist

⁴ A reviewer points out that younger people use more doublets because they are more interested in queer issues than older speakers. While this may be true, the question is to what extent do queer

engagement also has a significant effect, with speakers engaged in academic feminism showing a higher rate of doublets than everyone else. We hypothesize that this effect is the result of the “technical” nature of the doublets. As described above, many of the speakers in CaFé describe having to “train” themselves over time to use them, and some mention still failing, despite their firmly held conviction that this feminist linguistic practice is important. Public speaking in classes (educational contexts) and conferences (scientific contexts) is part of an academic’s job, and so people in this profession have many opportunities to practice using doublets, which ends up making them more proficient. They are also engaged in training and knowledge transfer, so they may have a greater desire to model the linguistic behavior that they believe will have a positive effect on gender equality. Indeed, a number of speakers in CaFé bring up actively teaching their students how to use doublets (15), or having university professors model this linguistic behavior for them (16).

(15) *c'est ce que j'expliquais aux aux étudiants et aux étudiantes que j'avais en face de moi [...] il faut au départ en prendre conscience qu'on ne parle qu'au masculin et et de se dire, "là, allez masculin et féminin" puis une habitude va en chasser une autre* (Paris 14)

‘It's what I was explaining to the students I had in front of me or the people that I might have in training [...] one first needs to become aware that we only speak in the masculine and to tell oneself “now, let's go masculine, feminine”, and then one habit will chase away another’

(16) *on a une professeure qui à chaque début de cours quand elle commence un nouveau cours elle conseille à ses étudiants étudiantes d'utiliser l'écriture inclusive* (Montreal 20)

‘At university, [...] we have a teacher that, at the beginning of each class when she started a new class, she advised her students to use inclusive writing’

In this vein, we also consider it significant that the most commonly doubled noun is *étudiant(e)* ‘student’, a noun that has been commonly featured in written doublets in university publications for many years (see Burnett and Pozniak 2021).

5 Conclusions

In this paper, we presented a quantitative study of noun phrases in the spoken French corpus *Cartographie linguistique des féminismes* (CaFé). We argued that CaFé

issues relate to doublets: doublets could simply be used to visibilize women, which is what they have traditionally been used for, and not necessarily non-binary people.

is a particularly interesting corpus for studying reference to human beings because of how big a part grammatical gender plays in French feminist linguistic activism and how intensely the question of how/whether to avoid so-called generic masculines is debated among activists and the general public. We show that, although many of the CaFé speakers express the idea that gender-inclusive communication is easier in writing than in speech, reflecting the overwhelming focus on written French in linguistic guides and public debates, their spontaneous use of masculines to refer to people other than men is extremely low. Crucially, however, the rate of so-called generic masculines is not so low because speakers are following the recommendations of many guides and using masculine-feminine doublets systematically. In fact, the doublets represent only 1.4 % of the noun phrases in the corpus, and are predominantly used by academics who need to train themselves and others to use them. Rather, one of the reasons so-called generic masculines are so low in the CaFé corpus is the fact that the spoken language generates far many more morpho-phonological contexts in which grammatical gender is neutralized than the written one. With our corpus study, we were able to quantify these contexts and showed that, were the corpus written, the percentage of so-called generic masculines would double, going from around 5–10 % of the noun phrases. Further research is needed to see if this gap between the spoken French language and the written one is also present to that extent in other oral corpora; however, since this gap has nothing to do with feminism or LGBT activism, we hypothesize that, were we to replicate this study on an oral corpus with non-activist speakers, we would find a similar result in which neutralizations are much more frequent than so-called generic masculines.

We did, however, identify another sociolinguistic phenomenon that interacts with gender-inclusive speaking and that we do expect to be more common in activist speech: person/identity-centered language. We argued that linguistic practices originating in 1990s North American anglophone disability activism have come into French and been expanded to cover a wide range of property nouns, not just ones referring to disabilities. French Canada being more aligned with English Canada and the United States on these topics than France, it is not surprising that we find a much higher rate of person noun phrases in Montreal than in Paris or Marseille. We showed that the more widespread use of person/identity-centered language in Montreal significantly decreases so-called generic masculines in this city's sub-corpus, since the person noun *personne* is grammatically feminine, while many of the property nouns are gender marked even in speech.

Overall, we have argued that activist and scientific descriptions of French should give the spoken language more consideration. From the point of view of descriptive and theoretical linguistics, our paper makes a novel contribution to the growing literature on how spoken French looks very different from written French. Other researchers have highlighted how the grammatical systems of varieties of spoken

French differ greatly in the (pro)nominal and verbal domains, and the conclusion that emerges from our study is that this is also the case when it comes to the types of noun phrases used: while masculines are one of the main ways to accomplish generic and gender-mixed reference in writing, our results suggest that so-called generic masculines may be marginal in speech, at least compared to neutralizations. From the point of view of gender-inclusive language policy, rather than emphasizing how complicated doublets are to spontaneously produce, guides could be emphasizing how French phonology and morphophonology make gender-inclusive speaking easier, and the synergies that exist with person-centered language. More generally, our study shows that francophones' speech can be highly gender inclusive, even if little effort is made to double masculines with feminines. We therefore conclude that, at least for some activists, the problem of how to avoid so-called generic masculine noun phrases has already been mostly solved.

Acknowledgments: The authors would like to thank Irvine Descout for his work on this project.

Research funding: This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 850539).

References

Abbou, Julie. 2023. Inclusive writing: Tracing the transnational history of a French controversy. *Gender and Language* 17(2). 148–173.

Abbou, J. & H. Burnett. 2025. Devenir féministe et queer à Paris et Montréal: Récits de vie dans le corpus CaFé. In Wim Remysen & Hélène Blondeau (eds.), *(Re)donner la parole aux corpus montréalais. Regards rétrospectif et prospectif*, 187–208. Montréal: Presses Universitaires de Montréal.

Abbou, Julie, Aron Arnold, Maria Candea & Noémie Marignier. 2018. Qui a peur de l'écriture inclusive? Entre délire eschatologique et peur d'émasculation Entretien. *Semen. Revue de sémio-linguistique des textes et discours* 44. <https://doi.org/10.4000/semen.10800>.

American Psychological Association. 1992. Guidelines for nonhandicapping language in APA journals.

Andrews, Erin E. 2019. *Disability as diversity: Developing cultural competence*. New York: Oxford University Press.

Arbour, Marie-Ève, Hélène de Nayves & Ariane Royer. 2014. Féminisation linguistique: Étude comparative de l'implantation de variantes féminines marquées au Canada et en Europe. *Langage et Société* 2. 31–51.

Atkinson, E. 2012. Gender features on n & the root. *Romance Linguistics*. 229–244.

Auger, Julie. 1994. *Pronomial clitics in Quebec colloquial French: A morphological analysis*. Philadelphia: University of Pennsylvania PhD dissertation.

Baron, Dennis. 2020. *What's your pronoun? Beyond he and she*. New York & London: Liveright Publishing.

Barra-Jover, Mario. 2004. Interrogatives, négatives et évolution des traits formels du verbe en français parlé. *Langue Française* 1. 110–125.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Berrendonner, Alain. 2004. Grammaire de l'écrit vs grammaire de l'oral: Le jeu des composantes micro- et macro-syntaxiques. In Alain Rabaté (ed.), *Interactions orales en contexte didactique: Mieux (se) comprendre pour mieux (se) parler et pour mieux (s') apprendre*, 249–264. Lyon: Presses Universitaires de Lyon.

Blanche-Benveniste, Claire. 1986. La syntaxe et le français parlé. *Études de linguistique appliquée* 63. 16–22.

Blanche-Benveniste, Claire. 2007. Corpus de langue parlée et description grammaticale de la langue. *Langage et Société* 3–4(121–122). 129–141.

Blanche-Benveniste, Claire, José Deulofeu, Jean Stéfanini & Karel Van den Eynde. 1987. *Pronom et syntaxe: L'approche pronomiale et son application au français*. Paris: SELAF.

Blondeau, Hélène. 2001. Real-time changes in the paradigm of personal pronouns in Montreal French. *Journal of Sociolinguistics* 5(4). 453–474.

Bousquet, Danielle & Gaëlle Abily. 2015/2022. *Guide pratique pour une communication publique sans stéréotype de sexe*. Paris: Haut conseil à l'égalité entre les femmes et les hommes.

Brauer, Markus & Michaël Landry. 2008. Un ministre peut-il tomber enceinte? L'impact du générique masculin sur les représentations mentales. *L'année psychologique* 108(2). 243–272.

Burnett, Heather & Céline Pozniak. 2021. Political dimensions of gender inclusive writing in Parisian universities. *Journal of Sociolinguistics* 25(5). 808–831.

Burr, Elisabeth. 2003. Gender and language politics in France. In M. Hellinger & H. Bußmann (eds.), *Gender across languages: The linguistic representation of women and men*, vol. 3, 119–139. Amsterdam & Philadelphia: John Benjamins.

Bußmann, Haudmod & Marlis Hellinger. 2003. Engendering female visibility in German. In M. Hellinger & H. Bußmann (eds.), *Gender across languages: The linguistic representation of women and men*, vol. 3, 141–174. Amsterdam & Philadelphia: John Benjamins.

Cameron, Deborah. 1995. *Verbal hygiene*. London: Routledge.

Conrod, Kirby. 2018. Pronouns and gender in language. In Kira Hall and Rusty Barrett (eds.), *The Oxford handbook of language and sexuality*, online ed. Oxford Academic, 2018. <https://doi.org/10.1093/oxfordhb/9780190212926.013.63> (accessed 28 February 2024).

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Crémier, Loïs. 2023. *Ce qu'ils font au neutre: Analyse sémiotique des guides de communication inclusive au Québec francophone actuel*. Montreal: Université du Québec à Montréal PhD thesis.

Culbertson, Jennifer. 2010. Convergent evidence for categorial change in French: From subject clitic to agreement marker. *Language* 86(1). 85–132.

Deulofeu, Henri José. 2001. La notion de construction corrélative en français: Typologie et limites. *Recherches sur le Français Parle* 16. 103–124.

Díaz Colmenares, Yarubi. 2021. Un regard sur le français inclusif canadien dans une journée de Twitter. *Toronto Working Papers in Linguistics*, 43. Toronto.

Díaz Colmenares, Yarubi & David Heap. 2020. Variation dans les accords du français inclusif. In *Proceedings of the 2020 annual conference of the Canadian Linguistic association*, vol. 29.

Dubois, Jean & René Lagane. 1973. *La nouvelle grammaire du français*. Paris: Larousse.

Elmiger, Daniel. 2021a. Y a-t-il un guide dans la rédaction? Les genres récrits: Chronique n° 8. *GLAD!. Revue sur le langage, le genre, les sexualités* 10. 144–149.

Elmiger, Daniel. 2021b. Toutes pour une, une pour toutes? Ou: que faire du masculin à valeur générique? Les genres récrits: chronique n° 9. *GLAD!. Revue sur le langage, le genre, les sexualités* 11. 174–181.

Flesch, Marie & Éléonore de Beaumont. 2023. Usages informels du français inclusif: Étude des doubles abrégés et complets sur Twitter, Reddit et YouTube. *Langue Française* 220(3). 59–78.

Granello, Darcy Haag & Todd A. Gibbs. 2016. The power of language and labels: “The mentally ill” versus “people with mental illnesses”. *Journal of Counseling and Development* 94(1). 31–40.

Grevisse, Maurice & André Goosse. 2016. *Le bon usage*. Bruxelles: De Boeck Supérieur.

Gygax, Pascal, Ute Gabriel, Oriane Sarrasin, Jane Oakhill & Alan Garnham. 2008. Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes* 23(3). 464–485.

Haddad, Raphaël (ed.), 2019. *Manuel d’écriture inclusive*. Paris: Mots-Clés.

Heiden, Serge, Jean-Philippe Magué & Bénédicte Pincemin. 2010. TXM: Une plateforme logicielle open-source pour la textométrie—Conception et développement. In *10th international conference on the statistical analysis of textual data, JADT 2010*, 1021–1032. Rome.

Hellinger, Marlis & Hadumod Bußmann. 2003. The linguistic representation of women and men. In M. Hellinger & H. Bußmann (eds.), *Gender across languages: The linguistic representation of women and men*, vol. 3, 1–26. Amsterdam & Philadelphia: John Benjamins.

Hornscheidt, Lann. 2003. Linguistic and public attitudes towards gender in Swedish. In M. Hellinger & H. Bußmann (eds.), *Gender across languages: The linguistic representation of women and men*, vol. 3, 339–368. Amsterdam & Philadelphia: John Benjamins.

Houdebine-Gravaud, Anne-Marie. 1998. *La féminisation des noms de métiers: En français et dans d’autres langues*. Paris: L’Harmattan.

Ihsane, Tabea & Petra Sleeman. 2016. Gender agreement with animate nouns in French. In *Selected proceedings of the 43rd linguistic symposium on romance languages*, 159–176. Amsterdam: Benjamins.

Kastronic, Laura. 2016. *A comparative variationist approach to morphosyntactic variation in hexagonal and Quebec French*. Ottawa: Université d’Ottawa/University of Ottawa Doctoral dissertation.

Laberge, Suzanne & Gillian Sankoff. 1979. Anything you can do. In Talmy Givón (ed.), *Discourse and syntax*, 417–440. Leiden: Brill.

Lamothe, Jacqueline, Anne-Marie Benoit, Fernande Dupuis & Sonia Lafond. 1992. *Guide de féminisation ou la représentation des femmes dans les textes*. Montreal: Comité institutionnel de féminisation.

Lessard, Michaël & Suzanne Zaccour. 2018. *Manuel de grammaire non sexiste et inclusive*. Paris: Syllepse.

Liang, Yiming. 2023. *Quantitative syntax, formal syntax and information theory: Bridging gaps by studying French variation*. Paris: Université Paris Cité PhD thesis.

Liénardy, Cyril, Julia Tibblin, Pascal Gygax & Anne-Catherine Simon. 2023. Écriture inclusive, lisibilité textuelle et représentations mentales. *Discours. Revue de linguistique, psycholinguistique et informatique. A Journal of Linguistics, Psycholinguistics and Computational Linguistics* 33. <https://doi.org/10.4000/discours.12636>.

Massot, Benjamin. 2010. Le patron diglossique de variation grammaticale en français. *Langue Française* 4. 87–106.

Michard, Claire. 1996. Genre et sexe en linguistique: Les analyses du masculin générique. *Mots. Les langages du Politique* 49(1). 29–47.

Michard, Claire. 1999. Humain/femelle: Deux poids deux mesures dans la catégorisation de sexe en français. *Nouvelles Questions Feministes* 20(1). 53–95.

Michel, Lucy. 2016. *La relation entre genre grammatical et dénomination de la personne en langue française: Approches sémantiques*. Dijon: Université de Bourgogne PhD Thesis.

Müller-Spitzer, Carolin, Samira Ochs, Alexander Koplenig, Jan Oliver Rüdiger & Sascha Wolfer. 2024. Less than one percent of words would be affected by gender-inclusive language in German press texts. *Humanities and Social Sciences Communications* 11(1). 1–13.

Office québécois de la langue française. 2020. Autoformation sur la rédaction épicène. <https://www.oqlf.gouv.qc.ca/redaction-epicene/formation-redaction-epicene.pdf>.

Palasis, Katerina. 2013. The case for diglossia: Describing the emergence of two grammars in the early acquisition of metropolitan French. *Journal of French Language Studies* 23(1). 17–35.

Pauwels, Anne. 1998. *Women changing language*. London: Longman.

Poplack, Shana. 1993. The inherent variability of the French subjunctive. In Christiane Laeufer & Terrell A. Morgan (eds.), *Theoretical analyses in romance linguistics*, 235–265. Amsterdam & Philadelphia: John Benjamins.

Pozniak, Céline, Emma Corbeau & Heather Burnett. 2023. Contextual dilution in French gender inclusive writing: An experimental investigation. *Journal of French Language Studies* 34(2). 273–292.

Richy, Célia & Heather Burnett. 2021. Démêler les effets des stéréotypes et le genre grammatical dans le biais masculin: Une approche expérimentale. *GLAD! Revue sur le langage, le genre, les sexualités* 10. 96–126.

Riegel, Martin, Jean-Christophe Pellat & René Rioul. 1994. *Grammaire méthodique du français*. Paris: Presses Universitaires de France.

Roberge, Yves. 1990. *Syntactic recoverability of null arguments*. Kingston, ON and Montreal: McGill-Queen's Press.

Rowlett, Paul A. 2011. Syntactic variation and diglossia in French. In *Salford working papers in linguistics and applied linguistics*, vol. 1.

Schafroth, Elmar. 2003. Gender in French. In M. Hellinger & H. Bußmann (eds.), *Gender across languages: The linguistic representation of women and men*, vol. 3, 87–117. Amsterdam & Philadelphia: John Benjamins.

Schmid, Helmut. n.d. TreeTagger: A part-of-speech tagger for many languages. <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> (accessed 21 February 2024).

Sharif, Ather, Adeal Liam McCall & Kianna Roces Bolante. 2022. Should I say “disabled people” or “people with disabilities”? Language preferences of disabled people between identity-and person-first language. In *Proceedings of the 24th international ACM SIGACCESS conference on computers and accessibility*, 1–18.

Simon, Anne-Catherine & Clémence Vanhal. 2022. Renforcement de la féminisation et écriture inclusive: Étude sur un corpus de presse et de textes politiques. *Langue Française* 3. 81–102.

Spinelli, Elsa, Jean-Pierre Chevrot & Varnet Léo. 2023. Neutral is not fair enough: Testing the efficiency of different language gender-fair strategies. *Frontiers in Psychology* 14. 1256779.

Toma, M. 2021. *Étude quantitative sur les dénominations des pratiques langagières féministes dans la presse francophone de 1984 à nos jours*. Paris: Université de Paris Master's thesis.

Vachon-L'Heureux, Pierrette. 1992. Quinze ans de féminisation au Québec: De 1976 à 1991. *Recherches Féministes* 5(1). 139–142.

Vivanti, Giacomo. 2020. Ask the editor: What is the most appropriate way to talk about individuals with a diagnosis of autism? *Journal of Autism and Developmental Disorders* 50(2). 691–693.

Wilkens, Rodrigo, Bruno Oberle, Frédéric Landragin & Amalia Todirascu. 2020. French coreference for spoken and written language. In *Proceedings of the twelfth language resources and evaluation conference*, 80–89.

Xiao, Hualin, Brent Strickland & Sharon Peperkamp. 2023. How fair is gender-fair language? Insights from gender ratio estimations in French. *Journal of Language and Social Psychology* 42(1). 82–106.

Yaguello, Marina. 1979. Pronoun envy ou la querelle du masculin générique. *Cahiers Charles* V1(1). 151–159.