Characterizing Non-binary French:
A First Step towards Debiasing Gender Inference

Marie Flesch and **Heather Burnett**
Laboratoire de Linguistique Formelle
Université Paris Cité
marie.flesch@u-paris.fr
heather.burnett@cnrs.fr

Abstract

This paper addresses a bias of gender inference systems: their binary nature. Based on the observation that, for French, systems based on pattern-matching of grammatical gender markers in "I am" expressions perform better than machine-learning approaches (Ciot et al., 2013), we examine the use of grammatical gender by non-binary individuals. We describe the construction of a corpus of texts produced by non-binary authors on Reddit, (formely) Twitter and three forums. Our linguistic analysis shows three main patterns of use: authors who use non-binary markers, authors who consistently use one grammatical gender, and authors who use both feminine and masculine markers. Using this knowledge, we make proposals for the improvements of existing gender inference systems based on grammatical gender.

1 Introduction

Gender inference constitutes an important domain of NLP research and applications. Being able to identify a user's social gender can have many benefits, ranging from commercial (eg. capitalizing on gender-based consumption patterns, (Wachter, 2020), to civic (e.g. ensuring that people of all genders have equal access to public services and platforms (Pareek, 2019; Küchler et al., 2023), and scientific (e.g. properly understanding how gender influences computer-mediated communication, Argamon et al., 2007; Bamman et al., 2014; Schler et al., 2006). Despite researchers and public/private actors becoming more conscious of the complexities of social gender, and a stated emerging desire to include users whose gender does not align with the male/female binary (non-binary, genderfluid, etc., Dev et al., 2021; Ovalle et al., 2023), in NLP, gender inference is almost always based on a binary conception of social gender. This situation is the result of many factors, among them the fact that many gender inference papers in NLP still adopt an essentialist view of gender (i.e. one in which linguistic patterns are directly attributed to biological aspects of sex/gender), and, as observed by Larson (2017), training and testing datasets with non-binary (or other) users are lacking. In this way, individuals whose genders do not correspond to *male* or *female* are made invisible by current NLP gender inference systems. We consider that this invisibilization constitutes bias against non-binary users (BIAS STATEMENT).

Building on one of the principles of data feminism ("what gets counted, counts" (D'ignazio and Klein, 2023), we argue that this bias creates both representational and allocational harms (HARM STATEMENT). The fact that most (if not all) systems fail to recognize the existence of individuals whose gender exists outside the male/female binary is, by definition, a representational harm, (see Blodgett et al., 2020, p. 5455-5456), and this misrepresentation of the gender distribution of online spaces hinders research in social science devoted to studying them (Pareek, 2019). The extent to which binary gender inference systems create allocational harms will depend on their applications: systems that use gender inference to propose beneficial services or products will exclude non-binary users, which could adversely impact their material (social, political and economic) conditions.

As discussed above, if we want to debinarize, and therefore debias, gender inference and other NLP systems, a crucial first step is to create datasets composed of contributions from people with a varied set of gender identities. However, creating these datasets is challenging for several reasons: because "non-binary" is a label that encompasses diverse gender identities, because current labeling practices in gender inference datasets are not adequate, and because, in the age of LLMs, large datasets are favored – and non-binary individuals make up a small (although possibly growing) portion of the population (Brown, 2022).

160

Another possible reason for the lack of non-binary representation in NLP datasets may be the focus on the English language (O'Connor et al., 2024). Deep learning techniques, which have been increasingly used in the field of gender inference since 2017, need large amount of data to be accurate. However, in French, and for other languages that have grammatical gender, small datasets and simple methods may be enough to create reliable gender inference systems (i.e., more reliable than most gender inference machine learning/deep learning techniques). This is because in French, speakers have to gender themselves when talking/writing about themselves, providing information about their gender identity. Using a deterministic pattern-matching technique based on grammatical gender in "I am" type statements, Ciot et al. (2013) reaches a higher accuracy in gender inference than "traditional" machine learning techniques. This type of system may be also used to infer gender beyond the binary, as non-binary speakers may use gender-neutral grammatical markers, which can combine the masculine and the feminine (*français.e*) or use neomorphemes (*françaiz*). It implies a shift from machine learning gender inference systems based on "sociolinguistic features" or on less interpretable features (character ngrams, bleached features), to systems based on grammatical gender. These systems rely on linguistic knowledge of grammatical gender. However, to this date, we do not know much about the way non-binary individuals use grammatical gender, because most linguistic studies rely on survey and questionnaire data, which may not reflect actual use of language (for example Kaplan, 2022 or Hord, 2016).

Thus, we present the first corpus study of the ways non-binary individuals use of grammatical gender when writing about themselves, in a corpus of computer-mediated communication. We automatically extract grammatical gender in *je+être* ("I+to be") expressions expressions and classify gender markers into four categories: feminine, masculine, non-binary (use of neomorphemes or combined use of a masculine and a feminine marker), and neutralization (adjectives or nouns that bear no grammatical gender information). We then examine inter-platform and inter-author variation. Results show that there is considerable variation between individuals, some sticking to one grammatical gender, and others switching between grammatical genders. We also consider the question of whether there are non-binary markers; one of the

candidates is the period (*français.e*). However, we find that the use of the period is largely limited to a single platform; we conclude that constructing a gendered inference system based on the period would lead to misgendering.

Our contributions are:

- A description of the construction of a corpus of non-binary French using various data sources (Reddit, Twitter, and three forums).

- A description of a methodology for the automatic extraction of grammatical gender (presented in more detail in Flesch and Burnett, 2025), which can be used as a basis for systems for other languages with grammatical gender.

- A description of the use of grammatical gender by non-binary French speakers, which may also be of use for other NLP tasks such as machine translation, text generation, etc., by providing authentic uses of grammatical gender by non-binary individuals.

- Suggestions for the creation of a non-binary gender inference system for French based on grammatical gender.

2 Including Non-binary Individuals in Gender Inference Datasets: Balancing Harms and Benefits

2.1 Benefits

The first benefit of including non-binary individuals in gender-inference datasets is the fact that it may help reduce the impact of misclassification (or misgendering) by systems, which can create discrimination (Pareek, 2019). Hamidi et al. (2018) explored the impact of misgendering by interviewing transgender and non-binary individuals about their perceptions and attitudes about automatic gender recognition systems that infer gender from video, pictures or voice. Among the harms listed are the increase in dysphoria, and the fact that gender inference can be used as a tool for oppression that invalidates non-binary identities (p. 7). For some participants, being misgendered by a machine was seen as worse than being misgendered by humans; one pointed out that "Programmatic misgendering [...] just adds to the ocean we all swim in of constant small comments ... [Misgendering] is death by a thousand paper cuts" (p. 5). It is necessary to underscore here that misgendering does not

only affect non-binary individuals, but also transgender and cisgender women and men. This was explored by Fosch-Villaronga et al. (2021), who asked Twitter users if they had been misgendered by the company's gendering algorithm. 19% of the 109 respondents had been misgendered. Interestingly (and maybe not surprisingly), gay and bisexual men, non-binary individuals and women were more likely to be misgendered than straight men. The second benefit is the social impact that such research can have. As D'ignazio and Klein (2023) point out, "what is counted—like being a man or a woman—often becomes the basis for policymaking and resource allocation. By contrast, what is not counted — like being nonbinary — becomes invisible" (p. 97). Gathering quantitative data is essential for social change: it backs advocacy efforts aimed at policy reform and highlights structural inequalities on a large scale, without reducing marginalization to anecdotal stories (Tandon, 2018). The third benefit is that looking at the way non-binary individuals use language, with quantitative methods, may help us improve our understanding of the relationship between language and gender, and highlight its complexity - forcing us to rethink the premise of gender inference, i.e. that gender identity can be inferred from the way people write or speak.

2.2 Harms

Creating more inclusive datasets is not without dangers, however. To create gender inference systems, we need to label authors; and in order to do this, we need categories. When it comes to non-binary individuals, what are the right categories? Non-binary identities resist categorization (Pareek, 2019). The label "non-binary" itself can be seen as problematic, and there is no consensus as to its definition; it is used to refer to a broad spectrum of identities, that can sometimes be fluid (Pareek, 2019). Any attempt to represent more than two gender categories can thus be problematic: it "must also be viewed critically because all category models tend to create exclusions and develop normative discourses" (Motschenbacher, 2010, p. 40). Beyond the issue of reproducing a limiting view of gender, one might wonder if creating non-binary gender inference system is desirable at all. The answer may depend on the intended use of the systems. Commercial systems, by contrast to the systems used to produce purely scientific knowledge, have an impact on people's lives. They are usually used

for profit, for marketing purposes, and to support decision-making processes in recruiting or credit applications, for example (Fosch-Villaronga et al., 2021). As is widely known now, many of these systems are biased and will disadvantage women (Hall and Ellis, 2023). Including transgender or non-binary categories may thus lead to an increase of stigma for populations that are already dealing with discrimination and oppression. Moreover, in states or territories where anti-trans or anti-non-binary policies are enacted, gender inference systems could be used to identify, target and persecute gender-diverse individuals.

2.3 The need for an ethical labeling process

In our view, a more ethical and diverse approach to gender inference starts with the labeling process. To infer gender from corpora, NLP systems need what is sometimes referred to as "ground truth", or labels that reflect the "known" gender identity of individuals. However, this "ground truth" often seems shaky. Studies do not always report how it was obtained (Larson, 2017). When they do, it becomes clear that obtaining these labels is a gender inference task in itself, as opposed to, for example, using preexisting metadata about people based on self-declarations. This task is generally performed by humans who rely on one or several clues. For example, since Twitter does not provide structured sociodemographic metadata about its users, annotators may rely on profile pictures to generate the "ground truth" (for ex., in Ciot et al., 2013). Other datasets are annotated by also looking at user names, user descriptions, and grammatical gender markers if available (for ex., Verhoeven et al., 2017). These types of approaches are questionable, because they are likely to classify non-binary individuals as "men" or "women". Indeed, in the absence of self-declarations such as "I'm a man" or "I'm a woman", how can one decide that a person is *not* non-binary ? Non-binariness is not reflected in first names or appearance. Moreover, this method may also misgender a number of women and men who do not have conventional gender expressions, or who have ambiguous/uncommon first names (or first names which association to gender varies from one culture to another, such as "Nicola"). While some studies acknowledge the bias inherent in binary gender inference, few address the limitations of the "ground truth" labels themselves.

3 Non-binary French

3.1 Non-binary French and grammatical gender

In this study, we attempt to characterize non-binary French by focusing on the use of grammatical gender by non-binary individuals when they talk about themselves, in *je+être* ("I+to be") statements. We think that if there is a linguistic "signature" of non-binary French, it is the context where it may be the most visible. To understand the choices non-binary individuals make, it is important to know the constraints they are faced with, when talking about themselves. When *être* ("to be") is an auxiliary, it is followed by a past participle which is always gendered in written French (*je suis allé$_M$ je suis allée$_F$*, I went). When *être* is an attributive verb, it can be followed (among other things) by an adjective, a noun or a noun phrase. Some adjectives are gendered (*intelligent$_M$, intelligente$_F$*, "smart"), and some are not (*triste* "sad", *jeune* "young"). Determiners are always gendered (*la$_F$, le$_M$* "the"; *un$_M$ une$_F$* "a"). For nouns, there is a variety of cases: gendered nouns (*client$_M$, cliente$_F$*), common gender nouns (Corbett, 1991) which are gendered but have the same form in the masculine and the feminine (*un / une artiste* "artist", *un / une collègue* "colleague"); epicene nouns which are gendered but can refer to people of all genders (*une personne* "a person", *un parent* "a parent"); and so-called generic masculines, which can be used for people of all genders, even when a feminine version exists, especially used for titles and functions (*avocat* "lawyer", *professeur* "professor"). Francophone speakers who wish to find alternatives to masculine or feminine forms can use two main types of solutions. The first one invisibilizes gender; it consists in using epicene nouns (*je suis une personne française* instead of *je suis français* "I'm French"); clippings (*ingé* instead of *ingénieur*, "engineer"); anglicisms (*je suis happy* instead of *je suis content/contente*); locutions (*je suis à sec* instead of *je suis fauchée* "I'm broke"), etc. Bypassing binary gender this way may require some effort, but it generally stays "under the radar", as the linguistic resources used are not specifically non-binary. However, in some cases, like with past participles, this approach is near impossible to implement in written French (as opposed to spoken French, where most gender markers in past participles are not audible). The second type of solution aims to make non-binary gender visible, and thus requires an intervention on the French grammatical gender system. Various solutions have been proposed; they were described by Kaplan (2022), who makes the distinction between three approaches. The "Compounding" approach combines masculine and feminine suffixes in either order, often, but not always, using a typographical sign; (*content·e* "happy", *acteurice* "actor", *joueureuse* "player", etc.); these forms emerged in the context of feminist linguistic activism, and are used both to provide alternatives to so-called generic masculines, giving more visibility to women in language, and to refer to non-binary individuals. The "Invariable" approach uses a single non-binary suffix (*amiz* "friend", *acteuz* "actor"). Finally, in the "Systematic" approach, more complex grammatical systems are created, taking into account the morpho-phonology of French. The most famous is probably the Alpheratz system (Alpheratz, 2018) which proposes various neutral morphemes; for example, the *-ix* morpheme for words that end in [i] (*amix* "friend") or the *-ae* morpheme for words that end in [e] (*députae* "deputy").

3.2 Related work

Studies that have investigated the way non-binary individuals make use of the various solutions they have at their disposal are generally based on interviews or questionnaires, mostly conducted in Quebec. Some are small scale studies, such as Kaplan (2022), who asked six non-binary individuals about their attitudes, preferences, and knowledge of non-binary/gender-neutral French gender systems, showing that non-binary French is a site of significant instability. Jack-Monroe (2021) examined how seven non-binary bilingual (French-English) individuals navigate the French grammatical gender system; the participants' responses shed light on the diversity of practices and attitudes towards grammatical gender, a person stating for example that the binary nature of French grammatical gender allows them to express themselves with more nuance than English, by switching between masculine and feminine markers. Studies on a larger scale, such as Hord (2016) and LaVieEnQueer (2017), asked participants about their preferred pronouns, terms of address, or their preferred practices in writing. Dumais (2021) is one of the very few corpus studies of non-binary French; it looks at the way eight non-binary individuals from Quebec use grammatical gender when referring to other people in a corpus of sociolinguistic interviews, showing

that some are "superneutralizers" who use few gendered words when talking about non-binary referents. Another corpus study (Flesch and De Beaumont, 2023) examined inclusive language on Twitter, Reddit, and YouTube comments, finding that non-binary individuals use inclusive markers more frequently than women and men. However, despite the multiple proposals made by non-binary grammars and the current debate around inclusive and gender-neutral language, no study, to our knowledge, has specifically investigated the use of grammatical gender by non-binary individuals in self-reference.

4 Dataset

To create the corpus, we used five platforms: Twitter (scraped in 2022 and 2023, before it became X), Reddit, and three online forums: betolerant.fr (a forum dedicated to queer identities), forum.asso-contact.org, and forums.madmoizelle.com[1]. The data collection approach was different for each platform, depending on their structure and affordances. For example, for Reddit, we used a large (preexisting) corpus containing 16,480,376 comments from 21 subreddits; for forums.madmoizelle.com, we extracted data from a single discussion thread dedicated to non-binary identities. Table 1 describes the methods used to create each subcorpus, and the corpus size. The initial corpus contains a total of 16,818,576 texts, mostly originating from Reddit and Twitter. Even if the three forums account for a small part of the original dataset, we considered it was important to include them, as two of them (betolerant and non-binary discussion thread on Madmoizelle) are queer spaces, where the likelihood of non-binary individuals interacting seemed higher than on Reddit or Twitter.

We only included individuals who explicitly identified as being non-binary. For forums, we searched at *je suis non-binaire* statements ("I'm non-binary") in posts and comments, using a list of non-binary gender terms (Appendix A.1) compiled using various sources (Wilfried, 2021; Wikipédia, 2024; Espineira; Klutz and Wallis; Rézo; lgbtqia.fandom). For Twitter, we searched for the "naked" gender identifiers in users' profile descriptions (*agenre* "agender" instead of *je suis agenre* "I'm agender", for example). In addition, for Twitter users, we considered the presence of the non-

binary flag in a profile description as a non-binary identifier. As the non-binary emoji does not exist, this flag is represented in our corpus by a sequence of yellow, white, purple and black heart or circle emoji. After extraction of non-binary identifiers, each post, comment, and Twitter profile description was manually inspected to remove false positives due to the use of reported speech, for example (*il a dit je suis non-binaire* "he said I'm non-binary") or uncertainty about one's gender identity (*je sais pas si je suis non-binaire* "I don't know if I'm non-binary"). Then, a subset of the initial corpus was created by retaining only the users who explicitly identified as non-binary. The resulting corpus contains 18,662 texts (878,250 words) by 398 unique accounts (Table 2). Even though it is possible that a Twitter user and a betolerant participant (for example) are the same person using different screen names, we will refer to these accounts as "users".

5 Grammatical Gender Analysis

5.1 Extraction of grammatical gender in *je+être* expressions

Using the R package Quanteda (Benoit et al., 2018), we generated concordance lines using as keywords various *je+être* ("I+be") expressions, including spelling variants of *je suis* "I am" (*j'suis*, *chuis*, *jsuis*, *ch'uis*), and the verb conjugated in various tenses (*j'étais*, *j'ai été*, *je serai*, *j'aurai été*, *j'avais été*, *je fus*, *j'eus été*, *je serais*, *j'aurais été*), together with their negative forms (*je ne suis*, *je n'étais*, etc.). We then extracted grammatical gender from adjectives, past participles and nouns that come directly after these expressions (*je suis grande*) or after an adverb (*je suis très grande*) using pattern-matching with an ad-hoc lexicon, created by combining several lexicons: the GLÀFF (Sajous et al., 2013), the Lefff (Sagot, 2010), a subset of the Flexique lexicon (Bonami et al., 2013) annotated with animacy information (Chlebowski and Bonami, 2015), and two lists of titles and functions (Cerquiligni et al., 1999; Otto-Bruc, 2022). The lexicon contains past participles, adjectives, and only nouns that can refer to human beings. Tokens are annotated with one of three grammatical gender labels: feminine, masculine, and neutralization (common gender nouns such as *élève* "student", gender-neutral adjectives such as *triste* "sad", epicene nouns such as *personne* "person"). As common gender nouns and gender-neutral adjectives can be part of a gendered noun phrase, when a determiner is used for example

Platform	Source of data	Scraping method	Date of data collection	Texts
Twitter	tweets in French, geolocalized in France, Québec, Morocco and Belgium	Twitter API	2022–2023	333,721
Reddit	21 subreddits: AskFrance, AskMec, AskMeuf, besoindeparler, conseiljuridique, Elles, france, FranceDetendue, FranceLibre, jeuxvideo, LgbtqiEtPlus, lgbtfrance, Lyon, NonBinairesFR, ParentingFR, paris, Québec, questionsante, SexualiteFR, vosfinances	PMAW python function (Podolak) and Apify (Rudiger, 2022)	2022–2025	16,480,376
betolerant	non-binary forum ("Forum non binaire")	custom R script with rvest package (Wickham, 2024)	2025	1547
madmoizelle	discussion thread "Pirates du genre"	custom R script	2025	2443
asso-contact	use of keywords to find threads discussing non-binary identities	custom R script	2025	489

Table 1: Description of the methods used to create the corpus.

Platform	Users	Texts	Words
Twitter	360	7,139	156,872
Reddit	22	10,993	653,059
madmoizelle	7	417	38,298
betolerant	5	88	24,647
asso-contact	4	25	5,374
Total	398	18,662	878,250

Table 2: Subset used for the analyses.

(*je suis un$_M$ jeune* "I'm a young person", vs. *je suis jeune* "I'm young"), the system takes into account the context of these words to detect gender. When the system is unable to detect grammatical gender (as referring to the author), either because *être* "be" is used as a localization verb (*je suis à la maison*, "I'm home"), or because the noun/adjective/past participle is not in the lexicon (non-binary gender, slang terms, neologisms, mispellings), it labels the concordance as "NA". After this automatic extraction, both authors of the paper and an intern manually checked the labels, adding missing labels (including non-binary labels) when needed and correcting labeling errors due to reported speech or conditional statements. The anonymized and annotated of *je+être* expressions, with concordance lines shortened to a 4-word window to protect the authors' privacy, is available on OSF [2].

5.2 Variation across platforms

1564 expressions containing grammatical gender produced by 137 authors remained. Among these expressions, 177 were feminine (*je suis un peu paumée* "I'm a little lost"; *je suis pansexuelle* "I'm pansexual"); 885 were masculine (*je suis vraiment soulagé* "I'm really relieved"; *je suis un idiot* "I'm an idiot"), 95 were non-binary (*je suis très curieuxe* "I'm very curious"; *je suis plutôt content.e* "I'm quite happy"); and 405 were neutralizations (*je suis pas vraiment fan* "I'm not really a fan", *je suis allergique* "I'm allergic") (Table 3).

	f	m	nb	neutr.	total
betolerant	3	44	9	15	71
forum asso	3	10	0	13	26
madmoizelle	23	19	57	63	162
reddit	94	684	16	246	1040
twitter	54	128	14	69	265

Table 3: Grammatical gender markers in *je+être* expressions. (f = feminine; m = masculine; nb = non-binary; neutr. = neutralization)

The most frequent gender marker, overall, is the masculine, which accounts for 61.97% of all

[2]https://osf.io/8wzg3/

165

markers in the betolerant subcorpus, 65.77% in the Reddit subcorpus, and 48.30% in the Twitter subcorpus. The masculine is less present than neutralizations in the forum asso contact subcorpus (but note that this subcorpus only contains 26 markers in total). The madmoizelle subcorpus stands out in two ways: first, the masculine is the least frequent grammatical gender marker (11.73% of all markers, vs. 35.19% for non-binary markers and 38.89% for neutralizations); and second, it features the most non-binary markers (59.38% of all non-binary markers in the corpus). The betolerant subcorpus comes next in terms of non-binary markers (12.68% of all markers). Finally, feminine forms are much less frequent than masculine forms and neutralizations, especially on Reddit (9.04% of all forms) and betolerant (4.23% of all forms).

5.3 Variation across authors

When it comes to the use of grammatical gender, platforms seem to have a linguistic profile; but what about individuals? To answer this question, we focus on the 21 authors who have used at least 10 grammatical gender markers. There is quite a bit of dispersion in the dataset, with a median number of 31 markers per author, and an interquartile range of 34. Figure 1 shows the breakdown of the grammatical gender markers the 21 authors used (arranged in a way that shows authors that used the most non-binary markers to the left of the graph; authors present in the betolerant, madmoizelle and forum asso contact subcorpora are grouped in the "forum" category). All authors, except for one, used at least one masculine marker or one neutralization. Nine authors used at least one non-binary marker. Four authors exclusively used masculine markers; nine authors used two types of grammatical gender markers (a combination of masculine and feminine for eight of them), and eight authors used all three types of gender markers. Setting aside the neutralizations (which make gender disappear and seem relatively evenly distributed among authors), we grouped authors using k-means clustering on the relative frequency of feminine, masculine, and non-binary gender markers. Adding new clusters does not help reduce within-cluster sum of squares very significantly after n=3. The mean values (Table 4) show that cluster 1 (2 authors, both from the madmoizelle subcorpus) is characterized by a high frequency of non-binary markers; cluster 2 (13 authors) shows a very high proportion o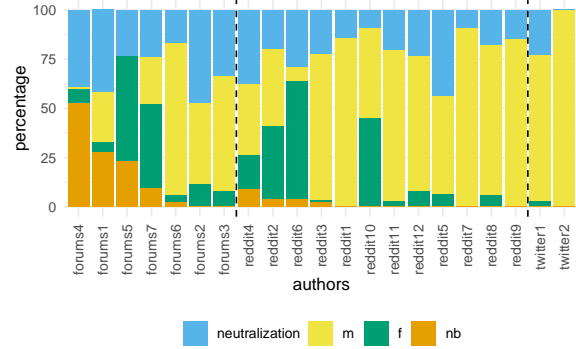f masculine markers, while cluster 3 (6 authors) is more balanced, with a higher count of feminine markers, followed by masculine and non-binary markers.

Figure 1: Proportion of gender markers used by authors in the subset, by platform

clusters	f_prop	m_prop	nb_prop
cluster 1	0.104	0.226	0.670
cluster 2	0.060	0.935	0.004
cluster 3	0.559	0.330	0.111

Table 4: Mean values per cluster (k-means results)

5.4 Comparison with the use of grammatical gender by women and men

In Flesch and Burnett (2025), we used the grammatical gender extraction system described in section 5.1 to infer binary gender in a corpus of Reddit comments. We found that a number of women use masculine markers when in *je+être* expressions, but also that some men do use feminine markers (mostly in frequent expressions such as *je suis sûr/e* or *je suis désolé/e*). In either case (omission or addition of a feminine marker), the pronunciation stays the same, and the variation in usage can be interpreted as a product of the complexity of the French spelling system, which has retained feminine markers when they no longer exist in speech. Thus, variation in grammatical gender usage is not unique to non-binary individuals. Here, we compare the variation in the use of feminine and masculine markers in our non-binary corpus to the test set in our previous corpus study, using subsamples of authors who used at least six grammatical gender markers (23 non-binary individuals, 19 women, and 38 men). Overall, in these samples, the ratio of masculine to feminine markers is much higher among non-binary individuals (781 masculine markers / 127 feminine markers, or ratio of 6.15) than for

women (34/63, ratio of 0.093). It is however lower than for men (2077/37, ratio of 56.13). To have a more precise idea of individual variation, we computed Shannon entropy scores with the R package vegan (Oksanen et al., 2025). We then classified users into three groups; low entropy (up to 0.3), medium entropy (0.3 to 0.7), and high entropy (0.7 to 1). In the high entropy group, there are five non-binary individuals (22% of the non-binary sample), two women (11% of women), and two men (5% of men). Non-binary individuals are a minority in the low entropy group (n=11, or 48%), which comprises 84% of women and 76% of men.

5.5 A look at non-binary markers

The 96 non-binary grammatical gender markers extracted from the corpus were produced by 24 people (or 17.52% of authors) (see Appendix A.2). All of them were created by combining a masculine and a feminine marker (generally in that order). The vast majority (n=91) were formed by using a punctuation sign. Among these, 72 were formed with a full stop; 7 with an interpunct (·), 11 with a hyphen, and one with parentheses. The preference of non-binary individuals for the full stop as a component of these markers echoes findings of other studies (LaVieEnQueer, 2017; Flesch and De Beaumont, 2023). Only five words were formed without a punctuation sign: the adjectives *heureuxse* and *curieuxe*, and the determiners *lea* and *lae* (the only form placing the feminine before the masculine, used twice by the same author). We thus see no trace, in the *je+être* expressions in our corpus, of morphemes proposed by non-binary grammars. This may be due to the fact that forms that compound the feminine and the masculine have gained visibility over the past decade following the debate on gender-inclusive language in France, and have entered the linguistic repertoire of Francophones.

6 Conclusions

We believe that gender inference can serve as a valuable methodological tool in scientific research, particularly when used to shed light on structural inequalities. However, this task should be conducted ethically, with a clear understanding that gender is not binary, fixed, or always externally discernible from texts. We emphasized the need for transparency in the labeling process and for inclusion of gender identities outside the binary.

In this paper, we tried to determine whether there

is a linguistic signature to being non-binary in written French, in order to assess the possibility of creating a pattern-matching NLP system. Our corpus study of grammatical gender in *je+être* expressions shows that there is not a single distinctive signature that would allow us to infer non-binary gender, but, instead, multiple patterns. This diversity of patterns could be due to (among other things) the instability of non-binary forms, which have emerged in the 21st century; the fact that the non-binary label encompasses diverse and fluid gender identities; the contexts in which internet users interact; but also, their attitude towards gender and language.

One finding was that some non-binary individuals use one grammatical gender fairly or very consistently; thus, there is no way to differentiate them from women and men, based on grammatical gender. Other authors use grammatical gender in ways that seem distinct to what women and men do. The first possible linguistic signature we have uncovered is a high amount of variation in the use of feminine and masculine markers; such variation seems much more frequent among non-binary individuals than it is for women and men. Adding a measure of entropy to a pattern-matching system would be a way to identify some non-binary authors, but not all. The most distinctive (i.e., distinct from what women and men usually do with grammatical gender in French) is the use of non-binary gender markers: it seems safe to say that the vast majority of people who use them in "I am" statements are non-binary. There could be some exceptions, however, such as people who wish to conceal their gender identity online, or people using non-binary markers in reported speech or ironically/mockingly. Our study provides valuable insight into what these markers look like; creating a gender inference system that extracts these markers in *je+être* statements using regular expressions would be fairly simple, and it could help debinarize gender inference in French corpora. However, it would be far from an ideal solution, as non-binary gender markers were used by a minority of authors in our corpus. This type of system would thus misgender most non-binary individuals, by classifying them as women or men. Furthermore, the productivity of this method would depend greatly on the type of corpus used; texts produced in queer/feminist spaces (such as the madmoizelle forum) seem the most likely to feature this type of grammatical markers.

We contend that, using this knowledge, it may

be possible to create a system that infers gender beyond the binary in French datasets: it would extract grammatical gender in *je+être* expressions with a lexicon, using regular expressions to extract non-binary markers, and integrate measures of diversity. To limit misgendering, we propose the following steps: combining extraction of grammatical gender with extraction of gender-identity statements; manual inspection of samples; and creation of an "unknown gender" category in cases of ambiguity. In our view, the most ethical way to produce a non-binary gender inference system (or any gender inference system) is to emphasize robustness rather than maximizing recall.

7 Limitations

The first limitation of our study is the fact that our dataset may not reflect broader non-binary French usage, for several reasons: the corpus is small (137 authors); we only included users who explicitly identified as being non-binary; and our analysis focuses on *je+être* expressions, which may not capture the full range of ways non-binary individuals express gender through language. Moreover, the lack of additional sociodemographic information, such as age or region, limits our analysis of sociolinguistic variation. Finally, because of the sensitive content of the corpus and concerns surrounding the privacy of internet users, we have decided not to share the full dataset, instead only making available the *je+être* expressions analyzed in section 5. While we understand this considerably limits the reproducibility of our study, we consider this solution to be a reasonable compromise, which illustrates the tension between the principles of open science and the need to protect the marginalized participants to our research.

8 Bias statement

Gender inference systems are almost always based on a binary conception of social gender. This situation is the result of many factors, among them the fact that many gender inference papers in NLP still adopt an essentialist view of gender (i.e. one in which linguistic patterns are directly attributed to biological aspects of sex/gender), and, as observed by Larson (2017), training and testing datasets with non-binary (or other) users are lacking. In this way, individuals whose genders do not correspond to *male* or *female* are made invisible by current NLP gender inference systems. We consider that this invisibilization constitutes bias against non-binary users (BIAS STATEMENT).

This bias creates both representational and allocational harms (HARM STATEMENT). The fact that most (if not all) systems fail to recognize the existence of individuals whose gender exists outside the male/female binary is, by definition, a representational harm, (see Blodgett et al., 2020, p. 5455-5456), and this misrepresentation of the gender distribution of online spaces hinders research in social science devoted to studying them (Pareek, 2019). The extent to which binary gender inference systems create allocational harms will depend on their applications: systems that use gender inference to propose beneficial services or products will exclude non-binary users, which could adversely impact their material (social, political and economic) conditions.

9 Acknowledgments

References

Alpheratz. 2018. *Grammaire du français inclusif*. Vent Solars.

Shlomo Argamon, Moshe Koppel, James W Pennebaker, and Jonathan Schler. 2007. Mining the blogosphere: Age, gender and the varieties of self-expression. *First Monday*.

David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. 2014. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2):135–160.

Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. quanteda: An r package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30):774.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (technology) is power: A critical survey of "bias" in nlp. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Olivier Bonami, Gauthier Caron, and Clément Plancq. 2013. Flexique: An inflectional lexicon for spoken french. In *Actes du quatrième Congrès Mondial de Linguistique Française*, page 2583-2596.

Anna Brown. 2022. About 5% of young adults in the u.s. say their gender is different from their sex assigned at birth.

Cerquiligni, Anne-Marie Becquer, Nicole Cholewka, Martine Coutier, Marie-Josèphe Mathieu, and Josette Frécher. 1999. Femme, j'écris ton nom... guide d'aide à la féminisation des noms de métiers, titres, grades et fonctions. Technical report, CNRS - Institut National de la Langue Française.

Aurélie Chlebowski and Olivier Bonami. 2015. *Annotation sémantique des noms de Flexique*. [rapport de stage de master, Laboratoire de Linguistique Formelle.

Morgane Ciot, Morgan Sonderegger, and Derek Ruths. 2013. Gender inference of twitter users in non-english contexts. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, page 1136–1145.

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

Sunipa Dev, Masoud Monajatipoor, Anaelia Ovalle, Arjun Subramonian, Jeff M Phillips, and Kai-Wei Chang. 2021. Harms of gender exclusivity and challenges in non-binary representation in language technologies. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, volume 1968–1994.

Gabrielle Dumais. 2021. "what do they say in quebec?": Non-binary gender expression in informal spoken quebec french. *Toronto Working Papers in Linguistics*, 43(1).

Catherine D'ignazio and Lauren F Klein. 2023. *Data feminism*. MIT press.

Karine Solène Espineira. Les genres non-binaires sur internet et facebook.

Marie Flesch and Heather Burnett. 2025. Mesurer les inégalités de genre en ligne avec le genre grammatical: Une étude du subreddit r/france. In *CORIA-TALN 2025*, Marseille.

Marie Flesch and Éléonore De Beaumont. 2023. Usages informels du français inclusif: étude des doublets abrégés et complets sur twitter, reddit et youtube. *Langue française*, 220(3):59–78.

Eduard Fosch-Villaronga, Adam Poulsen, Roger A Søraa, and Bart H.M Custers. 2021. A little bird told me your gender: Gender inferences in social media. *Information Processing Management*, 58(3):102541.

Paula Hall and Debbie Ellis. 2023. A systematic review of socio-technical gender bias in ai algorithms. *Online Information Review*, 47(7):1264–1279.

Foad Hamidi, Morgan Klaus Scheuerman, and Stacy M Branham. 2018. Gender recognition or gender reductionism? the social implications of embedded gender recognition systems. In *Proceedings of the 2018 chi conference on human factors in computing systems*, page 1–13.

Levi CR Hord. 2016. Bucking the linguistic binary: Gender neutral language in english, swedish, french, and german. *Western Papers in Linguistics*, 3(1).

Maxen Jack-Monroe. 2021. *Il, elle, on... iel est: queer (socio) linguistics and identity amongst franco-anglophone gender non-binary young people in Montreal*. Master's thesis, McGill University (Canada).

Jennifer Kaplan. 2022. Binary-constrained code-switching among non-binary french-english bilinguals. *Proceedings of the Linguistic Society of America*, 7(1):5279–5279.

Billy Klutz and Estelle Wallis. Lexique de la communaute lgbt.

Constanze Küchler, Anke Stoll, Marc Ziegele, and Teresa K. Naab. 2023. Gender-related differences in online comment sections: findings from a large-scale content analysis of commenting behavior. 41(3):728–747.

Brian Larson. 2017. Gender as a variable in natural-language processing: ethical considerations. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, page 1–11, Valencia, Spain. Association for Computational Linguistics.

LaVieEnQueer. 2017. *Le langage dans la communauté non-binaire*.

lgbtqia.fandom. Wiki lgbtqia fr.

Heiko Motschenbacher. 2010. *Language, gender and sexual identity. Poststructuralist perspectives*. John Benjamins Publishing Company, Amsterdam Philadelphia.

Jari Oksanen, Gavin L. Simpson, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R.B. O'Hara, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner, Matt Barbour, Michael Bedward, Ben Bolker, Daniel Borcard, Tuomas Borman, Gustavo Carvalho, Michael Chirico, Miquel De Caceres, Sebastien Durand, Heloisa Beatriz Antoniazi Evangelista, Rich FitzJohn, Michael Friendly, Brendan Furneaux, Geoffrey Hannigan, Mark O. Hill, Leo Lahti, Dan McGlinn, Marie-Helene Ouellette, Eduardo Ribeiro Cunha, Tyler Smith, Adrian Stier, Cajo J.F. Ter Braak, and James Weedon. 2025. *vegan: Community ecology package*.

Clémentine Otto-Bruc. 2022. Les noms de métiers, titres honorifiques et grades au féminin et au masculin.

169

Anaelia Ovalle, Palash Goyal, Jwala Dhamala, Zachary Jaggers, Kai-Wei Chang, Aram Galstyan, Richard Zemel, and Rahul Gupta. 2023. "i'm fully who i am": Towards centering transgender and non-binary voices to measure biases in open language generation. In *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency*, page 1246–1266.

Karen O'Connor, Su Golder, Davy Weissenbacher, Ari Z Klein, Arjun Magge, and Graciela Gonzalez-Hernandez. 2024. Methods and annotated data sets used to predict the gender and age of twitter users: Scoping review. *Journal of Medical Internet Research*, 26:e47923.

Vedika Pareek. 2019. *Non-binary gender and data*. PubPub.

Matthew Podolak. Pmaw: Pushshift multithread api wrapper.

Gustavo Rudiger. 2022. Reddit scraper.

Rézo. Lexique lgbtq+.

Benoıt Sagot. 2010. The lefff, a freely available and large-coverage morphological and syntactic lexicon for french. In *7th international conference on language resources and evaluation (LREC 2010)*.

Franck Sajous, Nabil Hathout, and Basilio Calderone. 2013. GlÀff, un gros lexique À tout faire du français. In *Actes de la 20e conférence sur le Traitement Automatique des Langues Naturelles (TALN'2013)*, page 285–298, Les Sables d'Olonne, France.

Jonathan Schler, Moshe Koppel, Shlomo Argamon, and James W Pennebaker. 2006. Effects of age and gender on blogging. In *AAAI spring symposium: Computational approaches to analyzing weblogs*, volume 6, page 199–205.

Ambika Tandon. 2018. Feminist methodology in technology research. *The Centre for Internet and Society*.

Ben Verhoeven, Iza Škrjanec, and Senja Pollak. 2017. Gender profiling for slovene twitter communication: The influence of gender marking, content and style. In *Proceedings of the 6th workshop on balto-slavic natural language processing*, page 119–125.

Sandra Wachter. 2020. Affinity profiling and discrimination by association in online behavioral advertising. *Berkeley Technology Law Journal*, 35(2):367–430.

Hadley Wickham. 2024. rvest: Easily harvest (scrape) web pages.

Wikipédia. 2024. Non-binarité. Page Version ID: 219993262.

Carbonell Wilfried. 2021. Qu'est-ce qu'une personne non binaire, et la non binarité?

A Appendix

A.1 List of non-binary gender identifiers

A.2 Non-binary forms

Non-binary gender identifiers
a-binaire, abinaire, agender, agenre, agenré, agenrée, androgyne, aporagenre, bigenre, demi boy, demi genre, demi girl, demi-boy, demi-genre, demi-girl, demiboy, demigenre, demigirl, enbien, enby, emby, fluide, fluide de genre, ft*, ftn, ftu, ftx, gender [+any word], gender[+any word], genre fluide, genre-fluide, genrefluide, genreflux, inter-genre, libragenre, maverique, mt*, mtn, mtu, mtx, multigenre, nb, neutrois, non binaire, non genré, non genrée, non-binaire, non-genré, non-genrée, nonbinaire, pangenre, paragenre, polygenre

Table 5: List of non-binary terms used to identify non-binary internet users.

Non-binary forms
désolé·e (1); *dévasté·e* (1); *développeur·se* (1); *doué.e* (1); *embêté.e* (1); *étonné.e* (1); *fâché-e* (1); *fatigué.e* (3); *fauché.e* (1); *gamin.e* (1); *genré.e* (1); *gentil.le* (1); *heureux.se* (2); *heureuxse* (1); *lae seul.e* (2); *lea seul.e* (1); *maladroit.e* (1); *maquillé-e* (1); *marqué-e* (1); *ménopausé.e* (1); *mis.e* (1); *né.e* (2); *noir·e* (1); *nul-le* (1); *obsédé.e* (1) *orienté·e* (1); *over-blindé.e* (1); *pansexuel.le* (1); *partant.e* (1); *passé.e* (1); *perçu.e* (1); *persuadé.e* (1); *poussé·e* (1); *ravi.e* (1); *reconnaissant.e* (1) *renseigné.e* (1); *représentatif-ve* (1); *resté.e* (1); *soigné.e* (1); *sorti.e* (1); *sûr-e* (1); *sûr.e* (2); *tatoueur.euse* (1); *terrifié.e* (2); *tombé.e* (2); *un.e* (1); *venu.e* (1)

Table 6: Non binary forms in *je+être expressions*, with their raw frequency.