

Evolution of the gender of “COVID-19” in traditional and social media in the French of three continents

Michael Dow & Patrick Drouin
Université de Montréal

Preprint submitted for peer review (January 2021)*

Abstract

In this paper, we trace the evolution from February to June 2020 of the gender of the noun “COVID-19” in French with respect to two databases: first, a Twitter database of nearly 78,000 tweets and second, a database of traditional media of approximately 500,000 articles. Each database considers only gender-marked instances of the term “COVID-19,” and our corpora are tagged for geographical origin in order to compare and contrast varieties of French as spoken in three continents, namely Africa, (North) America and Europe. We find that American media comply categorically and immediately with the recommendations of the feminine by the World Health Organization and various local (Canadian) linguistic authorities in early March 2020. More than 50% of tweets in the American data follow suit soon after. African media is similar in that a large number of articles and tweets adopt the feminine, but only coinciding with the recommendation of the feminine by the Académie Française in early May 2020. Finally, we find negligible use of the feminine in the European data. We argue that several factors are likely at play in these results, namely, dialect-specific tendencies in loanword adaptation and the French gender system, the relationship between linguistic authorities and local media, public attitudes towards linguistic authority and the relative time of the recommendations made by these linguistic authorities.

1. Introduction

The introduction of a novel noun to French, whether by borrowing, spontaneous neology or any other process of word formation, necessarily brings with it the question of the word’s grammatical gender. Despite a noted preference in the literature for the masculine as the “default” gender in French, many factors come into play in determining the ultimate gender of a given noun, such as lexical, semantic and phonological factors, as well as diatopic and diastratic variation (*grosso modo*, geographic and socioeconomic factors, respectively) and finally the attitudes of the speech community towards linguistic authority.

In this paper, we perform a quantitative study to trace the evolution of the gender of the word “COVID-19”¹ in French as spoken in three continents, that is, Africa, America and Europe. The time-sensitive nature of the database allows us to correlate trends with the publication of gender-specific recommendations by the press and linguistic authorities such as the Office québécois de la langue française and the Académie française. In this paper, we use two databases to synthesize two different perspectives: First, we employ the COVID-19-Tweet-IDs repository to generate a database of nearly 78,000 unique, French-language tweets from February to June 2020 with geographic data and unambiguous gender cues for the word “COVID.” This database allows us to hone in on less formal French as used by more everyday speakers and is unique in that it allows us to estimate the

*Acknowledgements redacted. Comments and suggestions are welcome: michael.dow@umontreal.ca, patrick.drouin@umontreal.ca

¹Hereafter, with the exception of the history of the term in §2.4, we abbreviate this term to “COVID.”

usage of either gender according to social status, as gauged by Twitter follower count. Second, we use the Eureka.cc database² to study the same question in traditional francophone media in the same three continents.

The word “COVID” provides an interesting case study in the morphosyntactic incorporation of neologisms in French and effectively serves as a microcosm for observing both the establishment of norms in borrowing and the real-time influence of previously noted factors in loanword adaptation. And while much ink has been spilled in the public sphere over the question of the gender of “COVID” in French, the present study is unique in (1) our two-pronged approach, (2) our approximation of social status and (3) our inclusion of African French, which is frequently neglected from such discussions.

A note on the terminology about the varieties of French studied is in order before continuing. Note, first of all, that the term “American” technically refers to both North and South America, though the number of tweets from South America meeting our criteria are negligible,³ and no South American media outlets are included in the Eureka.cc study. Both datasets (Twitter and traditional media) contain a variety of communities within each continent. Our searches were necessarily limited to the level of continent, seeing as the number of observations was not sufficient to extend the analysis to the level of country or region/province. Thus in referring to the data, we speak in terms of continents. However, much of the pertinent literature focuses on Hexagonal French and Québécois French. Seeing as these communities impressionistically provide the majority our European and American results, we permit ourselves some liberty of extrapolation in our analysis. We have made an explicit effort, however, to misrepresent neither our results nor the literature by mixing or switching degrees of specificity.

Furthermore, the situation of French in Africa is immensely complex, given the diversity of communities in which French is spoken and/or an official language. In addition, the usage and features of a given community’s French are stratified along both linguistic and extralinguistic factors which are absent from Europe and North America. The notion of “African results” should therefore be met with a healthy degree of skepticism in the absence of certain, more pointed studies in the literature (cf. §2.3). However, given that African varieties of French are frequently neglected in the discussion of phenomena like the gender of “COVID-19,” we find it important to include these results in the present study and interpret them to the best of our ability given the literature at our disposal.

The rest of our paper is structured as follows: In §2, we discuss gender in the French lexicon, with an emphasis on borrowings, as well as regional-specific differences. We also discuss the history of the word “COVID,” with specific reference to French. Section 3 outlines the methodology of both studies, and §4 presents our results. We synthesize and analyze our results in §5 and conclude.

2. Literature Review

2.1 Gender in the French lexicon

French nouns obligatorily fall into two morphosyntactic groups, traditionally called *masculine* and *feminine* genders, which can readily be observed in prenominal determiners (e.g., definite article *le* vs. *la*, masculine and feminine, respectively) and in adjective agreement. Unlike certain languages in which gender is highly predictable according to phonological factors (e.g., Qafar; Parker and Hayward, 1985) or a combination of phonological factors and declensional classes (e.g., Russian;

²<http://eureka.cc>

³For example, only 0.18% of unique users in our Twitter database mentioned “Brazil” or “Brasil” as their location.

Corbett, 1991), French gender is seen as fairly opaque (e.g., Bloomfield, 1933), though a certain number of regularities—often interacting and sometimes competing—can be observed in the lexicon.

Setting aside animate nouns,⁴ both phonological and derivational factors contribute to gender certainty (i.e., the degree to which a given form can be reliably predicted as having a certain gender) in French nouns. As Tucker et al. (1977) show, certain word-final strings, whether simple or complex, demonstrate high degrees of gender regularity in the lexicon. For instance, more than 99% of words ending in [ã] are masculine (e.g., *un accent*), while only 12% of [ad]-final nouns are (e.g., *un grade*). Seeing as nominal suffixes contribute a categorical gender, the segmentability of these endings (that is, whether or not a given substring constitutes or belongs to a separable morpheme) must also be considered. For instance, whereas the endings [aʒ] and [ɛʒ] are both predominantly masculine in the lexicon, only the former is a productive suffix (e.g., the *-age* in *lavage* ‘washing’). Meanwhile, words ending in [ɛʒ] are predominantly feminine and monomorphemic (e.g., *auberge* ‘hostel’). Beyond these observations on the lexicon, the evidence is robust that French speakers pay attention to these cues in processing lexical information and in assigning gender to nonce or novel words, both independently of each other and conjointly (e.g., Tucker et al., 1977; Karmiloff-Smith, 1979; Desrochers et al., 1989; Taft and Meunier, 1998; Holmes and de la Bâtie, 1999; Holmes and Segui, 2004; Author, 2013), and gender errors are strikingly uncommon in L1 French acquisition (Carroll, 1989).

2.2 Gender in anglicisms and borrowings

Just as in the French lexicon in general, the attribution of gender to borrowed words in French has been described as arbitrary and mysterious (Pergnier, 1989, p. 39); however, more recent studies again reveal the existence of several complex and competing factors.

First of all, nouns borrowed from languages with a grammatical gender system tend to preserve their gender in the source language (Poplack et al., 1982; Roché, 1992), insofar as that system’s categories aligns with those of French. While general and more theoretical discussions of borrowings and gender in French do not specifically consider African varieties of French, this tendency is independently confirmed for Arabic loans in Algerian (Derradji, 1999; Smaali, 1994) and Moroccan (Benzakour, 1995; Gaadi, 1995) French, as well as Italian loans in the French of Cameroonian internet users (Cutrì, 2014). The adaptation of genders other than masculine and feminine (e.g., neuter) have been shown to be subject to the same forces as those driving borrowings from languages without gender (cf. Beardsmore, 1971 for the adaptation of Flemish neuter nouns in Brussels French), to which we now turn our attention.⁵

The French lexicon has a relatively equal number of nouns of each gender, if slightly biased towards the masculine (56% vs. 44%); however, the vast majority of contemporary borrowings from languages without gender are masculine, at 85% (Roché, 1992). While the general equilibrium of genders is noted as far back as Old French, the disparity in borrowings at that stage is reversed (only 36% masculine), with a steady rise in masculine borrowings over time (idem). This reversal in trends can be explained in part by a change in source languages. Borrowings in Old French were most prominently technical or learned vocabulary from Latin, which skews heavily feminine

⁴As in many languages, the gender of animate, especially human referents in French is often determined by sex and/or gender identity, though not necessarily (e.g., *une victime* is always feminine). As these factors are extralinguistic, in keeping with the sources cited in this discussion, we consider only inanimate nouns.

⁵It is unclear to what degree the bilingual proficiency of a community plays a role in these phenomena. This factor does not prove to be significant for rate of borrowing in Poplack et al. (1988), though their study concerns English borrowings. Beardsmore (1971) finds, somewhat unsurprisingly, that more monolingual French speakers pay more attention to cues internal to French rather than the source gender of Flemish borrowings.

due to its derivational suffixes. After a rise in borrowings from Romance languages (with their own gender systems, see above) in Middle French, English quickly became the dominant source language, originating nearly 2.5% of the modern French lexicon (Rey-Debove, 1987). This coincided with, if not contributed to the rise of, an increasing and self-reinforcing productivity of the masculine, to the point where scholars consider it the “default” or “unmarked” gender in French (see in particular (Roché, 1992, 114-116)),⁶ and currently only 10 to 12.4 percent of borrowings from English are feminine (Hanon, 1970; Humbley, 1974; Surridge, 1984; Soubrier, 1985; Johnson, 1986).

A major factor in determining the gender of an English borrowing in French is the attraction of pre-existing words in the lexicon. This is typically discussed in the literature in terms of *parasynonyms* and/or *quasi-homonyms*. That is, English words often receive the gender of their French calques or translations, whether based on orthographic, phonetic or semantic analogy (Haden and Joliat, 1940; Nymansson, 1995; Lupu, 2005). Examples of this for the feminine include *une love affair* (based on *une affaire*) and *une backroom* (based on the correspondence of *room* with the French *une pièce*). Belleau (2016) also notes the importance of “paradigmatic integration,” by which borrowings in a certain semantic field (e.g., types of sausages and cured meats *pepperoni*, *mortadelle*, *chorizo*, and so on) tend to pattern together within a variety of French, presumably based on analogy with a more frequent and/or established borrowing within that field.

Another, somewhat more opaque factor in the determination of a borrowing’s gender is via ellipsis with a syntactically higher and often unexpressed French noun (Haden and Joliat, 1940; Nymansson, 1995; Lupu, 2005). This is the argument for words such as *une Ford* and *une start-up*, in that they receive the feminine gender based on understood nouns *une voiture* and *une entreprise* (or *une firme*), respectively. Note that this will be the argument put forward by several linguistic authorities for “COVID” (cf. §2.4).

Phonetic factors play a role, though diminished (Belleau, 2016) in determining a borrowing’s gender. These may be based on analogy with the lexicon or may be unique to borrowings. Concerning the former, English word endings may be associated with certain word endings in French and their gender. For instance, English *-y* (as in *party*) is frequently associated with French *-ie* [i], which skews feminine in the lexicon (Haden and Joliat, 1940; Nymansson, 1995). These factors may conflict with those discussed above, yielding variation. For instance, the *new beat* genre of music may be either feminine by analogy with *la musique* or masculine due to the phonetic factors (Nymansson, 1995).⁷ It is crucial to recall, however, that these forms in variation are not the norm, as discussed above.

Other phonetic factors are documented but must be considered in light of regional differences, which we will turn our attention to in §2.3. Before doing so, given the potential influence of the aforementioned phonetic factors, we find it opportune to present some basic statistics on the gender of the ending [id] and its various orthographic representations in the French lexicon. A survey of the [id]-final singular nouns of Lexique (Gimenes et al., 2020) yields 21 entries, 17 of which are masculine and 4 feminine. In nearly all words, the [id] rhyme corresponds to the orthographic sequence *-ide* or

⁶This is not to suggest that the notion of the masculine and feminine as “unmarked” and “marked” genders, respectively, is universally accepted or without controversy (see, for example, Coady (2018) and references therein). What is important for our purposes is the strong tendency for English borrowings to receive the masculine gender.

⁷Nymansson (1995) cites Tucker et al. (1977) as finding only 5 feminine forms in 120 words ending in [t]. We are not sure where this assertion comes from. Tucker et al. (1970) find that the simple ending [t] is ambiguous, being roughly 51% masculine. Expanding to the [it] ending, they find that 100% of those written as *-it* are masculine, in comparison with only 28% of *-itte* forms. In our own survey of [it]-final nouns in Lexique (Gimenes et al., 2020), we find 16 of 43 (37%) singular nouns with listed gender to be masculine. The uncertainty of this assertion should not detract from the existence of conflict between lexical and phonetic factors and the documented variability in the gender of certain borrowings.

-ide. The only exceptions are the masculine nouns *oxyde* and *caïd*, an Arabic loanword. Otherwise, words ending in orthographic *-id* are not pronounced as [id], rather [a] (as in *froid*) and [ɛ] (as in *laid*), and 100% of gender-bearing entries are masculine. As such, we find very little evidence in the lexicon for the attribution of the feminine to “COVID” based on phonetic factors (setting aside the “19” for the sake of argument).

2.3 Regional differences in gender and borrowing

We start with European and American varieties of French, which are more extensively studied with respect to gender and often contrasted with each other. These varieties do not show significant differences with respect to the gender of common, native French lexemes. A small number of exceptions are noted, especially in vowel-initial words, but are not necessarily specific to a region, rather being a property of oral, vernacular French (cf. Belleau, 2016, 62-67, for example).

Quantitatively speaking, patterns in gender assignment to English borrowings are noted to be quite similar between Canadian and European varieties of French (Haden and Joliat, 1940; Nyman, 1995; Belleau, 2016), with a few exceptions. First, the gender of specific words may show differences, a famous example being *party*, which is feminine in European French but masculine in Canadian French (e.g., Belleau, 2016). Additionally, the gender of specific morphemes has historically differed between the two regions. English *-ing* has occasionally yielded feminine nouns in Canadian French, whether phonetically adapted or not, for instance, *la réguine* vs. *la siding*, respectively (Haden and Joliat, 1940), while this ending is categorically masculine in European French. It should be noted, though, that these examples are dated from the perspective of modern Québécois French.

Both of these previous examples (i.e., *party* and *-ing*) illustrate two purported larger differences between the two regions with respect to phonetic factors. Vowel-final English words tend to be masculine and consonant-final ones feminine in Canadian French, unlike European French (Meney, 2018). Finally, monosyllabic words (e.g., *job*) tend to be masculine in European French but feminine in Canadian French (Belleau, 2016).

Findings on gender in African varieties of French generally fall into two categories: First, a general difficulty in acquiring and consistently applying the gender distinctions of French is noted among language learners and in certain lects of French in certain countries, regardless of the existence of nominal class systems in co-existing, vernacular languages. (Biloua, 2003, a.o.) notes this for Cameroonian French, going so far to state that “en l’état actuel des études portant sur le français du Cameroun, il est difficile de systématiser l’emploi du genre en français du Cameroun, sans courir le risque de se tromper à chaque fois” (pp. 144-145). Holtzer (2004) and Calvet and Dumont (1969) make similar observations for Guinean and Senegalese French, respectively. Ndjerassem (2005) mentions that certain words in Chadian French have a different gender as in normative French (e.g., *cafétéria* being masculine instead of the normative feminine).

A second theme arising in the literature is the omission of gender-signaling determiners. This is noted for French as spoken in Cote d’Ivoire (Jabet, 2006; Boutin, 2007) as well as in the French of Ivorian students (Herault and Vonrosbach, 1967; N’Guessan, 1982), a phenomenon which leads to general confusion over the use of the masculine and feminine (Ayewa, 2009). Such determiner dropping is a noted commonality between Ivorian French and the popular French of Montréal (e.g., *quand j’ai lâché [l]’école*), though less common in the latter (Hattiger and Simard, 1982, citing Sankoff and Cedergren, 1971). Omission of gender agreement is a documented feature of “Camfranglais⁸,” both spoken (de Féral, 2006) and written on the internet (Telep, 2014).

⁸Camfranglais is a mixed language featuring indigenous languages of Cameroon, French and English.

While English loanwords in African varieties of French are extensively documented (e.g., Schmidt, 1990 and what can be extracted from Blondé, 1983), we were not able to find a detailed discussion or synthesis on the attribution of gender to these nouns, especially inanimate ones. A superficial survey of the lexicons of Gabonese (Boucher and Lafage, 2000), Chadian (Ndjerassem, 2005) and Cameroonian (Nzesse, 2009) French reveals a very small number of feminine loans from English (Gabonese *la blaze* ‘showoff’, *la mangrove*, *la shoes* ‘pair of shoes’; Chadian *l’antilope* (f.); Cameroonian *la dream-team* and *la shoes* ‘pair of shoes’) but not enough to derive any significant trends about any one variety. A search of Tunesian French (Naffati and Queffélec, 2004) yielded no feminine English loans.

2.4 COVID-19

On February 11, 2020, the International Committee on Taxonomy of Viruses officially named the novel coronavirus detected a few months prior “severe acute respiratory syndrome coronavirus 2” (abbreviated SARS-CoV-2). The same day, the World Health Organization (WHO) gave the disease caused by SARS-CoV-2 the abbreviated name “COVID-19,” for “coronavirus disease 2019” (World Health Organization, 2020a). Also on February 11, both Radio-Canada (R-C) and the Office Québécois de la langue française (OQLF) created terminological records for the term. On the one hand, Nathalie Bonsaint, a linguistic consultant at Radio-Canada (p.c.) reported that the R-C record classified that “COVID-19” was masculine. On the other had, Xavier Darras (p.c.), a language production coordinator at the OQLF, indicated that their record did not, at that time, include a gender. As observed by Nathalie Bonsaint, and corroborated by our corpora and by the analysis published for the news site *The Conversation* in May by Mathieu Avanzi (Avanzi, 2020), the term “COVID-19” was generally employed in the masculine until early March, with the exception of earlier WHO publications on the subject which variably used the feminine form. A statement on the web page for one of their online courses reflects that fact (World Health Organization, 2020b):

“Suite à la révision par l’OMS de l’appellation de la maladie et du virus qui la cause, ‘COVID-19’ est considérée comme une locution féminine. Nous vous prions ainsi de noter que toute mention ‘le COVID-19’ fait donc référence à la COVID-19.”

Bonsaint reported that by March 6 the WHO had updated its web site in order to use the feminine and that she took the same action on Radio-Canada’s internal terminological record in keeping with the French publications by the WHO (Radio-Canada, 2020). We are not aware of a press release by the WHO specifically recommending the feminine apart from sporadic mention on web pages dealing with the disease. According to Xavier Darras, the OQLF also updated its terminological record on the same day classifying the term “COVID-19” as feminine (Office québécois de la langue française, 2020). The Académie Française finally published an official recommendation of the feminine on May 7, 2020 (Académie française, 2020)⁹. As far we know, at the time of writing this paper, the Délégation générale à la langue française et aux langues de France (DGLFLF) has not put forward any formal recommendation on the subject.

The reasoning for the classification of “COVID-19” as feminine in all three sources (i.e., the Bonsaint memo and the recommendations of the OQLF and Académie Française) is the same, that is, the base referent of the term is the feminine word *maladie* ‘disease’, whether directly expressed

⁹During an interview for November 1st 2020 edition of France Culture’s podcast *Soft Power*, Hélène Carrière d’Encausse, Perpetual Secretary of the Académie française states that she took this decision by herself (without consulting other members of the Académie): <https://www.franceculture.fr/emissions/soft-power/soft-power-le-magazine-des-internets-emission-du-dimanche-01-novembre-2020>.

or not. That is, regardless of whether one acknowledges the ‘D’ for the English ‘disease’ in the acronym, these sources argue that *la COVID-19* should be interpreted as an ellipsis for *la maladie COVID-19* or similar.

The question of the gender of “COVID-19” proved contentious in the public sphere, and one can find polemics on the matter in francophone media up through December 2020 (e.g., Meteyer, 2020). This debate is outside the scope of this paper. That is, the merits of arguments for or against the feminine usage of “COVID-19” are not of interest to us, nor do we make a recommendation one way or the other. Indeed, such an activity would be antithetical to the descriptivist principles of modern linguistics. Rather, we seek only to document usage of either gender by the public and the media over time as a function of variety of French, and to elucidate potential causes for and/or explanations of these trends.

3. Methodology

3.1 Twitter study

The COVID-19-TweetIDs repository (Chen et al., 2020) served as the starting point for the current study’s database. This repository provides the unique identification numbers (hereafter, “tweet IDs”) of all publicly available tweets since January 21, 2020 containing any of a list of keywords such as “coronavirus,” “COVID-19,” and so on.¹⁰ According to the June 23 version of the project’s documentation, French-language tweets comprised roughly 3% of the corpus, numbering over 5.5 million tweets.

The tweet IDs from the months of January to June inclusive were then “hydrated.” The process of hydration essentially consists of downloading all available information for a given unique tweet identifier, provided by Twitter; the amount of information varies from tweet to tweet. This was performed using a Python script provided by the authors of the repository. The data were then standardized, subsetted and analyzed for gender in the R language (R Core Team, 2020) along the following lines.

3.1.1 Text processing

First, all non-French-language tweets were discarded. Here and throughout this paper, “French-language tweets” refers to tweets whose language is automatically identified as such by Twitter’s proprietary algorithm.¹¹ In accordance with the findings that geolocation is a useful metric in gauging the accuracy of Twitter’s automatic language detection (e.g., Williams and Dagli 2017; Graham et al. 2014), we also extracted geographical data from the user profiles of our database. By focusing on continents with large French-speaking populations, we were able to limit our dataset to more probable true positive identifications. As indicated in §3.1.2, samples of potentially questionable tokens (e.g., those originating from Spain) were manually verified and confirm the proper functioning of automatic language detection to a high degree of accuracy.

¹⁰A complete list can be found on the project’s GitHub repository at <https://github.com/echen102/COVID-19-TweetIDs/blob/master/keywords.txt>. Note that several permutations of “COVID-19” are included, to account for letter case and the presence or absence of the dash (or any variation thereof, such as en dashes).

¹¹See Trampus (2015) for more details. Some imperfections in the algorithm aside (e.g., Zubiaga et al. 2016; Lui and Baldwin 2012; Bergsma et al. 2012), Twitter’s automatic language identification is adequate for our purposes. For example, Lui and Baldwin (2014, p. 24) note that “the accuracy of the Twitter API is not substantially better than the best off-the-shelf language identifiers” employed in their paper. While certain language pairs such as Italian and Romanian do pose problems, we were not able to find specific reference in the literature to poor performance on French.

Tweets were then limited to those whose text contains gender-marked instances of the string “covid” in a case-insensitive search, regardless of the presence of “-19” (or any permutation thereof). Gender marking was identified by the presence of the following words in the immediately preceding word: *le*, *au*, *du* and *ce* for the masculine and *la* and *cette* for the feminine.

Tweet text was then cleaned up as follows. In order to later eliminate duplicate tweets, entry-initial “RT @[username]” was eliminated. URLs and Unicode characters were also removed. Apostrophes were standardized, and all punctuation (including the hash character) was then removed, except for apostrophes, commas and periods. Once line breaks and unnecessary whitespaces are finally cleaned up, each duplicate tweet was then reduced to a single instance.

The number of masculine and feminine occurrences in each entry of the database was then tabulated. Meanwhile, the timestamps provided by Twitter (expressed in Coordinated Universal Time) were converted to a POSIX date/time class interpretable by R, and the month and day were retained. The total of masculine and feminine occurrences of the word “COVID” was then calculated for each day.

Finally, user follower count was used to serve as a “brute-strength” proxy for social status, and accounts within each continent were separated into three bins of “small,” “medium” and “large,” each containing a roughly equal number of points. These ranges are reported in §4.1.

3.1.2 Geographical information

The remaining tweets were then processed for geographical information. While Twitter allows for users to tag their tweets for location, unfortunately, this information was present in only approximately 1% of the data at this stage of processing. In order to fill this gap, we processed the `user.location` field (non-empty for nearly than 63% of the dataset) for relevant information, after Unicode characters had been removed.

Two initial issues presented themselves with this field: First, the formatting is non-standard, in that people can include various information such as city, country, both or neither. Second, country names may be in either French or English (among others). To counteract these issues, we made a bilingual database of cities and regions (equivalent to French *régions* and Canadian provinces) with their respective countries and continents using the `maps` (Brownrigg, 2018), `countrycode` (Arel-Bundock et al., 2018) and `raster` (Hijmans, 2020) R packages. Names in this database were limited to those found on the European, African and American continents, in order to reduce mismatches.

After standardizing names between the packages, we removed from the user-provided information all words unattested in our custom place-name database. Words in `user.location` were then matched for cities in our database and their corresponding continent. This process was repeated separately for regional and country names. Finally, a subset of the 1,000 most common unmatched user-provided locations were manually assigned a continent. Subsets were also verified throughout the procedure, and certain manual corrections were implemented in the algorithm. For instance, North American cities beginning with “San” matched both America and Africa due to the San commune in Mali; this was corrected. Geotagged users’ country information was also extracted `place.country` field and matched with its continent. In the rare occurrence of mismatches between sources of information, or of multiple returns (typically because place names spanning two or more continents were provided by the user), if present, manually provided and geotagged information were taken as authoritative. Otherwise, the first continent was arbitrarily chosen.

A subset of 450 users, 50 per continent per follower number group, was randomly selected for verification of the accuracy of continent identification. We found 93.8% of the subset to be correctly identified and thus within the limits of acceptability. Africa had the lowest accuracy of the three continents at 87.3%, versus America at 95.3% and Europe at 98.7%.

All in all, this procedure resulted in a final database of 76,054 unique French-language tweets which, in summary, contained unambiguous gender information about the word “COVID” and from which geographical information could be ascertained.

3.2 Media study

The Eureka.cc database, essentially an aggregator of the world’s newspapers and other forms of media, was used in order to trace the evolution in usage of both genders for the term “COVID(-19)” in francophone media. The same masculine and feminine forms of “COVID” detailed above were entered separately in week-long intervals beginning with February 11, 2020 and ending June 30, 2020. Omission of “-19” did not preclude the full form “COVID-19” from appearing in the results. Each week’s search was performed separately for all French-language media in the database for each continent. The number of sources for each continent at the time of data collection were the following: 653 (North America), 825 (Europe) and 78 (Africa).

The number of articles corresponding to each gender (again, by week and continent) was then entered into a database.¹² While syndicated articles are present in the database, they could not be eliminated, nor do we believe they should be. Not only do we strongly doubt the gender of the term “COVID-19” to be a deciding factor on which articles are syndicated, but also we believe that the proliferation of certain articles containing one gender or the other reflects a certain zeitgeist as well as consumers’ experience.

4. Results

4.1 Twitter results

Table 1 presents the number of tweets in our final database by continent and month. The number of distinct users for each continent for the entire database are the following: 6649 for Africa, 4712 for America and 32767 for Europe. Given the sum of tweets per continent reported in Table 1, this users contributed on average the following number of tweets: 1.8 (Africa), 1.97 (America) and 1.67 (Europe). Follower size groups are defined in the following way: Small accounts (abbreviated “S” in

	Africa	America	Europe	Total
Feb.	47	64	417	528
Mar.	904	728	4636	6268
Apr.	2270	1460	9367	13097
May	2385	1817	10169	14371
Jun.	6349	5197	30244	41790
Total	11955	9266	54833	76054

Table 1: Number of tweets in database by continent and month

certain tables and figures) range from 0 to 213 followers in Africa, 0 to 285 in America and 0 to 196 in Europe. Medium (M) accounts consist of 214 to 1558 followers in Africa, 286 to 1595 followers in America and 197 to 1017 followers in Europe. Finally, large (L) accounts have minimally 1559

¹²The nature of the search engine was prohibitive to counting tokens, and furthermore, such an approach is undesirable. We do not assume authors to vary in usage within a single article, direct quotations notwithstanding. In addition, whereas tweets are severely limited in their length, newspaper articles are not. Lengthier articles could then skew individual counts for a given gender. Finally, while articles debating the gender of the term are likely to be present in the database, such articles will provide only one count for each gender, essentially nullifying each other.

		Feb.	Mar.	Apr.	May	Jun.
Africa	S	0/13 (0%)	5/323 (1.55%)	17/790 (2.15%)	109/774 (14.08%)	748/2142 (34.92%)
	M	0/14 (0%)	15/346 (4.34%)	35/830 (4.22%)	135/834 (16.19%)	749/2035 (36.81%)
	L	0/20 (0%)	6/239 (2.51%)	18/677 (2.66%)	156/838 (18.62%)	932/2259 (41.35%)
America	S	0/9 (0%)	73/207 (35.27%)	188/466 (42.15%)	290/626 (46.33%)	943/1862 (50.64%)
	M	0/22 (0%)	109/267 (40.82%)	280/522 (53.64%)	349/612 (57.03%)	923/1732 (53.29%)
	L	0/33 (0%)	140/261 (53.64%)	340/509 (66.8%)	423/611 (69.23%)	1186/1720 (68.95%)
Europe	S	0/91 (0%)	16/1640 (0.98%)	41/3124 (1.31%)	204/3405 (5.99%)	839/10293 (8.15%)
	M	0/160 (0%)	22/1602 (1.37%)	39/3048 (1.28%)	169/3374 (5.01%)	894/10263 (8.71%)
	L	2/166 (1.2%)	23/1455 (1.58%)	85/3257 (2.61%)	190/3525 (5.39%)	872/10032 (8.69%)

Table 2: Feminine over total gendered uses of “COVID” by continent, follower size and month, Twitter data

followers in Africa, 1596 in America and 1018 in Europe. Table 2 presents the number of feminine uses of “COVID” and its percentage of total gendered uses (masculine or feminine) per month within each continent’s group. Counts of masculine and feminine uses (as indicated by colour) per day are graphed over time, using X-splines, in Figure 1, according to continent and follower size. Note that the x- and y-axis limits are technically unique to each pane. To our knowledge, the marked spike in activity in early June in all types of accounts can in part be attributed to discussion about anti-racism demonstrations in the wake of the killing of George Floyd.

The package `EnvCpt` (Killick et al., 2020) was used to detect the date of the maximum-likelihood estimates of change points in the percentage of feminine uses for each subgroup (continent by follower size). In our case, this corresponds to any day on which the percentage of feminine uses of the word “COVID” rises to an important degree. The dates identified by this procedure were then compared manually with the percentages themselves to eliminate negligible or ephemeral switchpoints.¹³ These results are plotted in Figure 2. The y-axis of each plot corresponds to the percent feminine per day. Red lines indicate the mean percent of each period identified by the model; a switchpoint is then the date at which the mean changes. From this data, we attempted to identify a single, crucial date for each type of account, which are the following (presented in the order of small, medium and large within each continent): May 12, 11 and 10 for African accounts; March 7, 6 and 8 for American accounts; and May 9, 8 and 7 for European accounts. Note, of course, that the degree of change is not comparable from one group to the next, especially at the level of continent, as can be seen in Figure 2.

¹³These were limited to negligible spikes in early March in African and European accounts of all follower sizes, at approximately 10 and 5 percent, respectively. Additionally, the increase of overall activity in June saw with it an even greater increase of the percent feminine in all African accounts as well as small American accounts. No veritable dates in June were identified by this procedure for European accounts.

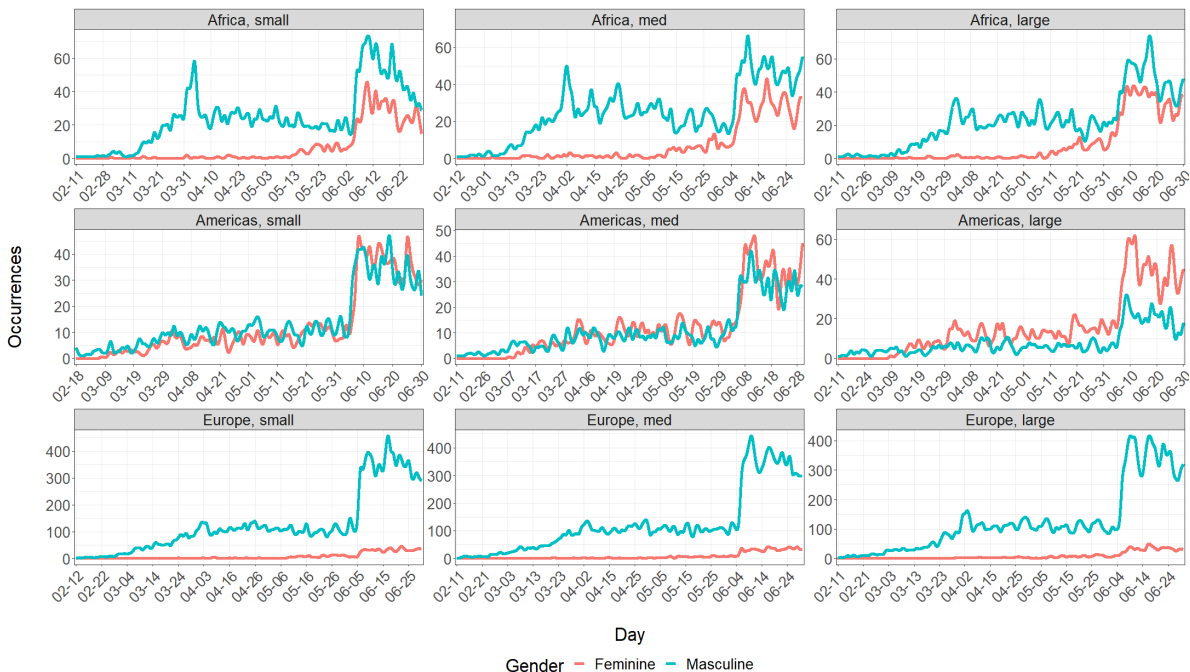


Figure 1: Masculine and feminine occurrences of “COVID” over time, Twitter data

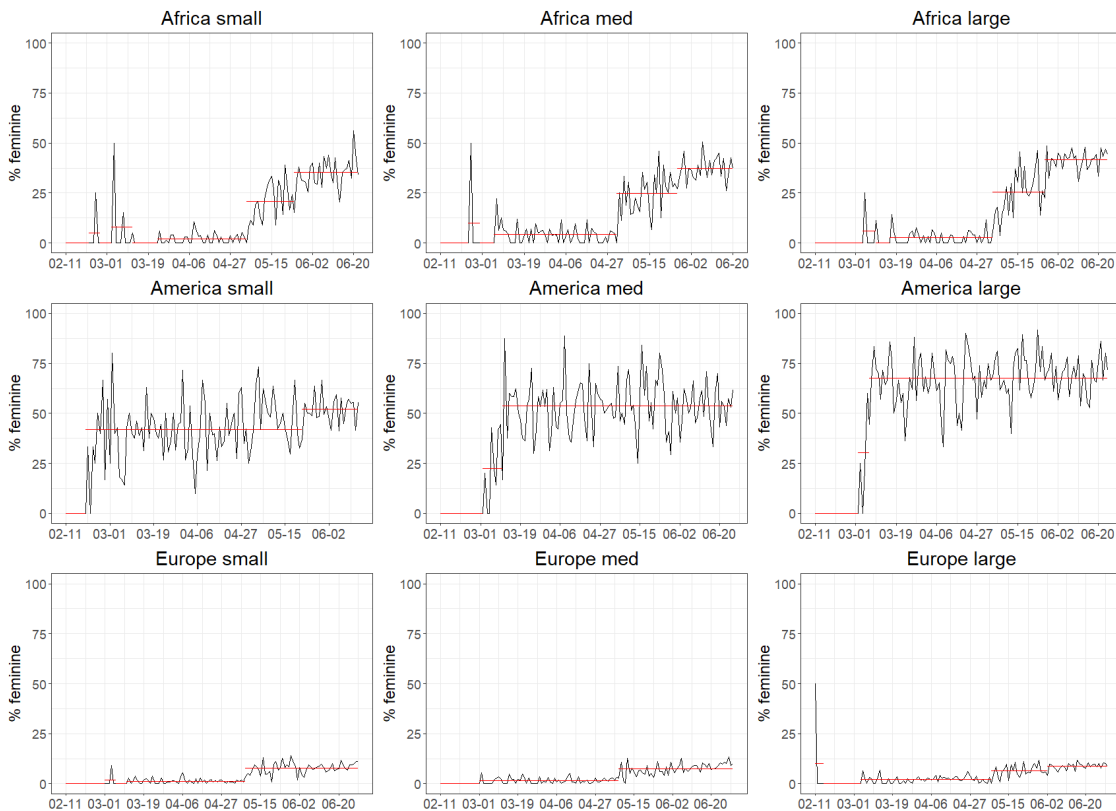


Figure 2: Percent feminine over time with switchpoints, Twitter data

4.2 Media results

The number and proportion of feminine instances of “COVID” are provided in Table 3 by week, as indicated by the starting day of the 7-day period, and continent. These numbers are plotted in Figure 3. While it would be of interest to see how individual outlets have varied over time, the

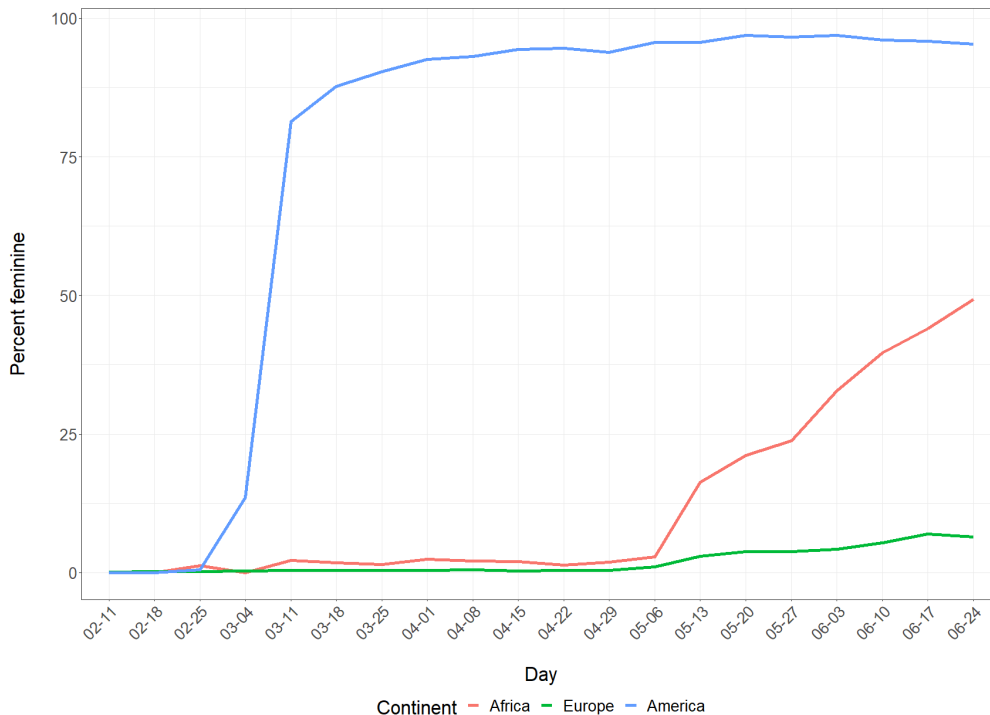


Figure 3: Percent of feminine use by week in traditional media, by continent

Eureka.cc search engine does not easily lend itself to this inquiry.

5. Discussion & conclusion

5.1 Summary

Both our Twitter data and our media data show important differences among the three continents studied, concerning the mean usage of the feminine gender of “COVID” and variation therein. We first discuss what we see as the two extremes, America and Europe, and then Africa.

The American Twitter data show an immediate and important increase in the feminine coinciding with the events detailed in §2.4 (in particular, the Radio-Canada memo and the related publication). This effect, however, is stratified by number of followers. Small and medium accounts converge on 50% feminine usage towards June, while large accounts show a steeper increase in March and a higher convergence, at 70%. Meanwhile, the media study shows a near-categorical passage to the feminine in early March. All together, the higher percentage of the large account Twitter group is likely due to the presence of Twitter accounts associated with these media organizations.

In stark contrast, the European Twitter data demonstrate both negligible usage of the feminine and little stratification between account sizes. While all account types see a rise in feminine instances of “COVID” coinciding with the recommendation of the Académie Française in early May, the difference between April and May is approximately 2 to 4 percent, or from 1 or 2 percent to 5 or

Date	Africa	America	Europe
02-11	0/1 (0%)	0/313 (0%)	1/817 (0.12%)
02-18	0/17 (0%)	0/460 (0%)	3/1510 (0.2%)
02-25	2/157 (1.27%)	6/1176 (0.51%)	10/4631 (0.22%)
03-04	0/111 (0%)	324/2401 (13.49%)	20/6543 (0.31%)
03-11	10/445 (2.25%)	8183/10062 (81.33%)	59/15237 (0.39%)
03-18	19/1071 (1.77%)	10134/11558 (87.68%)	84/22729 (0.37%)
03-25	20/1373 (1.46%)	11155/12342 (90.38%)	96/25619 (0.37%)
04-01	37/1510 (2.45%)	11651/12585 (92.58%)	93/26620 (0.35%)
04-08	27/1290 (2.09%)	10376/11138 (93.16%)	101/22302 (0.45%)
04-15	25/1247 (2%)	11393/12072 (94.38%)	79/23546 (0.34%)
04-22	16/1220 (1.31%)	10478/11076 (94.6%)	93/21943 (0.42%)
04-29	21/1131 (1.86%)	10151/10811 (93.9%)	82/19520 (0.42%)
05-06	31/1095 (2.83%)	10071/10534 (95.6%)	191/18878 (1.01%)
05-13	180/1105 (16.29%)	8207/8579 (95.66%)	583/19727 (2.96%)
05-20	170/804 (21.14%)	8561/8829 (96.96%)	633/16933 (3.74%)
05-27	223/936 (23.82%)	7552/7815 (96.63%)	588/15746 (3.73%)
06-03	360/1097 (32.82%)	6742/6954 (96.95%)	648/15215 (4.26%)
06-10	368/927 (39.7%)	6653/6922 (96.11%)	845/15670 (5.39%)
06-17	407/926 (43.95%)	6390/6662 (95.92%)	1120/16194 (6.92%)
06-24	559/1135 (49.25%)	5284/5540 (95.38%)	1033/16038 (6.44%)

Table 3: Feminine uses of “COVID” in traditional media, by week and continent

6 percent. While June saw a similar rise from May, the average percent of feminine use did not meet 9 percent. European media outlets are even slower and less consistent in their use of the feminine, ranging from 1 to 3 percent around the publication of the Académie Française, and reaching only a maximum of 6.9 percent in mid-June.

The African data can be seen as situated between the other two continents. Just like the European Twitter data, African accounts are not stratified in the same way as the American data are. They do, however, show a more important increase in the use of feminine in May (an increase of approximately 12 to 16 percent), a trend which continues into July. African media sources mirror this behaviour, showing a stark rise in the feminine in early May and continuing to rise. While the increase in May is similar to that of Twitter accounts, the end result is slightly higher at 49%.

Using the Eureka.cc database, we can extrapolate the media trends into the future. For the month of December 2020, North American media remains stable at more than 95 percent feminine (32,021 feminine vs. 1613 masculine). Otherwise, we see increases in both African and European media, more prominently in the latter. European outlets amount to slightly higher than 15 percent feminine (12,354 feminine vs. 68,732 masculine), while African outlets rise to over 61 percent (3,441 feminine vs. 2191 masculine). It remains to be seen whether African media will converge on near-categorical use of the feminine, as in America, or will continue to show variation (as American Twitter accounts do towards the end of the Twitter database).

5.2 Analysis

Without direct input from speakers (e.g., survey data), we can only speculate on the reasons behind these trends. However, we see three originating causes for the differences among continents, one linguistic and two extralinguistic.

First, we consider the linguistic variable of dialect-specific practices in morphosyntactic adaptation of loanwords (especially English loanwords), as well as community-specific differences in functional load of gender. We noted in §2.3 that Québécois French has a tendency to feminize consonant-final (English) loanwords. This is one factor which may favour attribution of feminine to “COVID.” The reader is reminded, however, that this is only one of several differences from European French, and that Québécois French does not categorically feminize English loans (recall, for instance, the word *party*). It is unclear whether this case study of “COVID” suggests that generalization across word-final strings in the lexicon (i.e., [id] being a predominantly masculine ending) is less important to speakers of Québécois French; we leave this matter open to future research.

Concerning African French, less discussion was available in the literature, but we saw that in certain lects and/or certain geographically-specific varieties, gender distinctions proved less important. This was manifested by omission of gender markers and variation in the gender of native French words. While we believe that these observations may account for some of the variation, we are skeptical as to whether it is the impetus for the tendencies observed in either the Twitter or media results. Much more research needs to be done in this area before stronger conclusions can be drawn, with respect to both shared and novel vocabulary as well as to loans of various sources.

The second potential explaining factor is the unique relationship between media outlets and linguistic authorities in Canada. Specifically, the OQLF offers a service of linguistic consultation to Québécois media outlets with respect to terminological and neological questions, as does Radio-Canada for its own journalists across Canada. With respect to the term “COVID” and its gender, both the OQLF and Radio-Canada recounted to us to have consulted with journalists, and the recommendations of the feminine detailed in §2.4 were met with little resistance on the part of Canadian journalists (Darras, p.c.; Bonsaint, p.c.). We are not aware of similar services offered by the Académie, and while the Délégation générale à la langue française et aux langues de France

(DGLFLF) does offer linguistic consultation to French journalists, the DGLFLF has not published a recommendation for either gender for the word “COVID.”¹⁴ Meanwhile, we are not aware of governmental agencies in African countries specific to the French language, though some countries have agencies in matters of the *Francophonie* or in affairs of national languages.¹⁵

The influence of these institutions on the North American francophone media landscape and the evident (but voluntary) compliance of journalists to these authorities are no doubt a crucial factor in the propagation of the feminine there and eventually beyond its borders. This is in stark contrast with the perseverance of European media in the use of the masculine after the recommendations both March and May. Meanwhile, judging from the increase in the feminine in early May in African media (traditional and social), it would appear, at least in the case “COVID-19,” that a non-negligible sector of African francophone media defers to the Académie Française for matters of terminology and neology, although it may certainly be the case that local intermediaries played a role in encouraging the feminine.

Finally, related to this second point are the attitudes of the public with respect to linguistic authorities and their recommendations. In Kim’s (2017) study, Québécois participants responded positively to the statements (1) that French should be regulated in line with the societal norm and (2) that the government’s work in promoting French is helpful. In comparison, French, Belgian and Swiss participants responded negatively to these questions. Similarly, Tremblay (1994) finds in a survey of Québécois speakers that, while they generally prefer endogenous terms (that is, terms organically or spontaneously arising in Québec) to those created by the OQLF, they respect the work of the OQLF and hold a positive attitude towards the French spoken in Québec. To our knowledge, little has been written on the attitudes of African French speakers towards the Académie Française, although language policy has largely proven ineffectual, according to Spolsky (2018, p. 71):

“After independence (whether it was seized or granted), the French-speaking elite replaced the colonial rulers, applying much the same language policy in most cases or attempting to establish hegemony for a local variety. . . . [C]entralized language policy failed to change the widespread traditional language practices. . . . Assuming that the answers [to language problems] are linguistic and that central language management will work appears, from the French colonial experience, to be a mistake.”

Indeed, this failure has created an environment in African countries for innovation and the creation of local norms, as Francine Quérémer of the Organisation internationale de la Francophonie: “The French language is not going to wait in all these [African] countries for the Académie to decide before it evolves” (O’Mahony, 2019). It would appear from our results that a sizeable cross-section of African media outlets and Twitter users do indeed defer to the Académie, but it is unclear to what degree this deference is sustainable or representative of the future.

The American data also touch upon the question of a local language norm in Québécois society and the role that the media, most specifically Radio-Canada, plays therein. As Bigot (2017) notes, Radio-Canada presenters regularly receive linguistic training (Bertrand, 2005), and their French is largely considered as the reference variety for Québécois French, citing the results of Bouchard and Maurais (1999). The immediate acceptance of the feminine by a large number

¹⁴Additionally, we are not aware of other, similar European agencies (e.g., the Service de la Langue française of the Fédération Wallonie-Bruxelles) which offer such services to the press and have put forward a judgment with respect to the gender of “COVID-19.”

¹⁵This estimation is based on a survey of the lists of ministers and governmental positions of the 10 most populous African countries with French as an official language, as offered by the website of the Ministère de l’Europe et des affaires étrangères of France.

of American (Québécois) may thus speak to the complex interplay between homegrown, implicit community norms and the explicit norms of language authorities, be they the OQLF or the media by proxy. More specifically, in the absence of competition of a more spontaneous, informal and widely accepted in-group (Québécois) variant, the important rise in the feminine in spring 2020 may in part be attributed to the public’s trust in and cooperation (though incomplete) with entities like Radio-Canada and the OQLF. This is, of course, assuming most uses were made with direct knowledge of these recommendations and that linguistic variables (in particular, feminization of consonant-final words) were not the sole cause of early results in the Twitter corpus.

Additionally, it is worth noting that this acceptance and propagation of the feminine in America was made despite the persistence of the masculine in European media (both traditional and social), as well as the silence of the Académie Française until May. This may be taken as a sign for the codification of a norm for Québécois French independently of an international standard. However, the mere act of this codification and the public’s acceptance may also speak to a persistent pressure on Québécois French to justify its features to outside parties. It may be the case that speakers of the more prestigious European, especially Hexagonal and Parisian, variety of French do not feel such pressures to defer to linguistic authorities, going so far as to equate “la COVID” with snobbery (Meteyer, 2020).

It is crucial to note, however, that the relative lateness of the Académie to recommend the feminine and the lack of action from the DGLFLF gave ample time for the masculine to take root in European usage. As Poplack et al. (1982) and Poplack et al. (1988) note, the gender of loanwords in French once “established” is essentially invariable. It would appear, then, from the European data that the period of February to May proved sufficient for the masculine gender to become fixed. (This interpretation hinges, however, on an ignorance of or disregard for the February recommendation of the WHO.) Meanwhile, the variation still present (at least in June) in American Twitter accounts may speak to the difficulty in switching from the masculine to the feminine after only a month of exposure. Only time will tell if the feminine prevails in the French of everyday American (as well as African) speakers, although—with a little optimism—we can only hope that the circumstances give us little occasion to speak of COVID-19 in the future.

5.3 Conclusion

Our goal with this paper was to follow the evolution of gender for the noun “COVID-19” in French. Being a sudden but globally used neologism, this word provides an unparalleled testing ground for the factors influencing the morphosyntactic incorporation of novel words in various varieties of French. We processed data from a corpus of social media (Twitter) and a newspaper corpus to identify the geographical origin of the tweets and newspaper articles in order to compare and contrast the varieties of French spoken in three continents: Africa, (North) America and Europe. Overall, we found that American media passed overwhelmingly to the feminine in March 2020, following recommendations by Canadian (and more specifically, Québécois) instances of linguistic authority, while usage in American Twitter plateaued off to 50-70% by June. Meanwhile, African media and users increased dramatically in their use of the feminine, but only after the recommendation of the Académie Française in May. Finally, use of the masculine is essentially negligible in both European datasets. We proposed an interplay of several factors to explain these results, both linguistic and extralinguistic. First, varieties of French differ somewhat with respect to their gender systems, particularly in English loanword adaptation. In addition, we noted differing roles of and attitudes towards language authorities. Finally, the relative tardiness of European (French) institutions likely played a role in solidifying those trends (despite a similar recommendation by the WHO months prior), allowing the masculine to become the community norm.

References

- Académie française. 2020. Le covid 19 ou la covid 19. Fiche terminologique.
<http://www.academie-francaise.fr/le-covid-19-ou-la-covid-19>.
- Arel-Bundock, Vincent, Nils Enevoldsen, and CJ Yetman. 2018. countrycode: An R package to convert country names and country codes. *Journal of Open Source Software* 3(28): 848. URL
<https://doi.org/10.21105/joss.00848>.
- Author. 2013.
- Avanzi, Mathieu. 2020. Le/la covid ? Réouvrir ou rouvrir ? Les leçons de grammaire du coronavirus. *The Conversation* URL <https://theconversation.com/le-la-covid-reouvrir-ou-rouvrir-les-lecons-de-grammaire-du-coronavirus-138633>.
- Ayewa, Kouassi Noël. 2009. Une enquête linguistique: le français, une langue ivoirienne. *Le français en Afrique* (25): 117–134.
- Beardsmore, H. Baetens. 1971. A gender problem in a language contact situation. *Lingua* 27: 141–159.
- Belleau, Rémi. 2016. *Attribution et variation du genre d'emprunts à l'anglais, à l'italien, au japonais et à l'arabe dans le lexique du français*. Master's thesis, Université Laval.
- Benzakour, Fouzia. 1995. Le français au Maroc: Processus néologique et problèmes d'intégration. In *Le français au Maghreb (Aix-en-Provence, 2-4 septembre 1994)*. 61–76.
- Bergsma, Shane, Paul McNamee, Mossaab Bagdouri, Clayton Fink, and Theresa Wilson. 2012. Language identification for creating language-specific Twitter collections. In *Proceedings of the second workshop on language in social media*. 65–74.
- Bertrand, Guy. 2005. La radio et la télévision: modèles linguistiques ou miroirs de société? In *Le français au Québec: les nouveaux défis*, ed. Alexandre Stefanescu and Pierre Georgeault. Québec: Fides, 445–460.
- Bigot, Davy. 2017. Regard rétrospectif sur la norme du français québécois oral. *Arborescences* (7): 17–32.
- Biloua, Edmond. 2003. *La langue française au Cameroun: analyse linguistique et didactique*. Peter Lang.
- Blondé, Jacques. 1983. *Inventaire des particularités lexicales du français en Afrique noire*, vol. 1. Paris: Agence de coopération culturelle et technique.
- Bloomfield, Leonard. 1933. *Language*. Henry Holt.
- Bouchard, Pierre and Jacques Maurais. 1999. La norme et l'école: l'opinion des Québécois. *Terminogramme* (91-92): 91–116.
- Boucher, Karine and Suzanne Lafage. 2000. *Le lexique français du Gabon: entre tradition et modernité*. Institut de linguistique française.
- Boutin, Akissi. 2007. Déterminant zéro ou omission du déterminant en français de Côte d'Ivoire. *Le français en Afrique* 22: 161–182.
- Brownrigg, Ray. 2018. *maps: Draw geographical maps*. URL <https://CRAN.R-project.org/package=maps>, original S code by Richard A. Becker and Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka and Alex Deckmyn. R package version 3.3.0.
- Calvet, Maurice and Pierre Dumont. 1969. Le français au Sénégal: interférences du wolof dans le français des élèves sénégalais. *Collection IDERIC* 7(1): 71–90.
- Carroll, Susanne. 1989. Second-language acquisition and the computational paradigm. *Language Learning* 39(4): 535–594.
- Chen, Emily, Kristina Lerman, and Emilio Ferrara. 2020. Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. *JMIR Public Health and Surveillance* 6(2): e19273.
- Coady, Ann. 2018. *The non-sexist language debate in French and English*. Ph.D. thesis, Sheffield Hallam University.
- Corbett, Greville G. 1991. *Gender*. Cambridge University Press.
- Cutri, Guiseppina. 2014. Quelques notes sur le phénomène de l'emprunt chez les internautes camerounais italophones. *Le français en Afrique* : 147–157.
- Derradji, Yacine. 1999. Le français en algérie: langue emprunteuse et empruntée. *Le français en Afrique* 13: 71–82.
- Desrochers, Alain, Allan Paivio, and Sylvie Desrochers. 1989. L'effet de la fréquence d'usage des noms

- inanimés et de la valeur prédictive de leur terminaison sur l'identification du genre grammatical. *Canadian Journal of Psychology/Revue canadienne de psychologie* 43(1): 62.
- de Féral, Carole. 2006. Étudier le camfranglais: recueil des données et transcription. *Le français en Afrique* 21: 211–218.
- Gaadi, Driss. 1995. Le français au maroc: L'emprunt à l'arabe et les processus d'intégration. In *Le français au Maghreb (Aix-en-Provence, 2-4 septembre 1994)*. 131–151.
- Gimenes, Manuel, Cyril Perret, and Boris New. 2020. Lexique-infra: Grapheme-phoneme, phoneme-grapheme regularity, consistency, and other sublexical statistics for 137,717 polysyllabic French words. *Behavior Research Methods* .
- Graham, Mark, Scott A Hale, and Devin Gaffney. 2014. Where in the world are you? Geolocation and language identification in Twitter. *The Professional Geographer* 66(4): 568–578.
- Haden, Ernest F and Eugene A Joliat. 1940. Le genre grammatical des substantifs en franco-canadien empruntés à l'anglais. *Publications of the Modern Language Association of America* 55(3): 839–854.
- Hanon, Suzanne. 1970. *Anglicismes en français contemporain*. Ph.D. thesis, Université Aarhus.
- Hattiger, Jean-Louis and Yves Simard. 1982. Deux exemples de transformation du français contemporain: Le français populaire d'Abidjan et le français populaire de Montréal. *Bulletin de l'Observatoire du français contemporain en Afrique noire Abidjan* 3: 67–81.
- Herault, Georges and Jean-Paul Vonrospach. 1967. *Étude phonétique et syntaxique du français d'élèves de cours préparatoire de la région d'Abidjan*, vol. 1. Institut de linguistique appliquée d'Abidjan.
- Hijmans, Robert J. 2020. *raster: Geographic data analysis and modeling*. URL <https://CRAN.R-project.org/package=raster>, r package version 3.3-13.
- Holmes, Virginia M and B Dejean de la Bâtie. 1999. Assignment of grammatical gender by native speakers and foreign learners of French. *Applied Psycholinguistics* 20(4): 479–506.
- Holmes, Virginia M and Juan Segui. 2004. Sublexical and lexical influences on gender assignment in French. *Journal of Psycholinguistic Research* 33(6): 425–457.
- Holtzer, Gisèle. 2004. Savoirs et compétences en français écrit d'élèves guinéens: les enquêtes campus (1998-2001). *Le français en Afrique* 19: 35–73.
- Humbley, John. 1974. Vers une typologie de l'emprunt linguistique. *Cahiers de Lexicologie* 25: 46–70.
- Jabet, Marita. 2006. Noms sans déterminant en français abidjanais: trait sociolinguistique, sémantique et/ou pragmatique? *Le français en Afrique* 21: 325–337.
- Johnson, Micheline. 1986. *Les mots anglais dans un magazine de jeunes: Hit-magazine, 1972-1979*, vol. 18. Frankfurt-am-Main: Peter Lang GmbH, Internationaler Verlag Der Wissenschaften.
- Karmiloff-Smith, Anette. 1979. *A Functional Approach to Language Acquisition*. Cambridge: Cambridge University Press.
- Killick, Rebecca, Claudie Beaulieu, Simon Taylor, and Harjit Hullait. 2020. *Envcpt: Detection of structural changes in climate and environment time series*. URL <https://CRAN.R-project.org/package=EnvCpt>, r package version 1.1.2.
- Kim, Minchai. 2017. *Variation terminologique en francophonie : Élaboration d'un modèle d'analyse des facteurs d'implantation terminologique*. Ph.D. thesis, Université Paris Sorbonne.
- Lui, Marco and Timothy Baldwin. 2012. langid.py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 system demonstrations*. 25–30.
- Lui, Marco and Timothy Baldwin. 2014. Accurate language identification of Twitter messages. In *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)*. 17–25.
- Lupu, Mihaela. 2005. La masculinisation du lexique français: le rôle catalyseur des anglicismes. *Analele Universității „Alexandru Ioan Cuza” din Iași. Secțiunea IIIe. Lingvistică* .
- Meney, Lionel. 2018. *Le français québécois entre réalité et idéologie: un autre regard sur la langue: étude sociolinguistique*. Presses de l'Université Laval.
- Meteyer, Madeleine. 2020. Les gens qui disent LA covid sont-ils seulement des snobs? *Le Figaro* .
- Naffati, Habiba and Ambroise Queffélec. 2004. *Le français en Tunisie*. UFR Lettres, Arts et sciences humaines.
- Ndjerassam, Mbai-Yelmia Ngabo. 2005. *Le français au tchad*. UFR Lettres, Arts et sciences humaines.
- N'Guessan, Kouadio. 1982. Nombre et spécification du nom en baoulé et en français : Étude comparative illustrée par des exemples pris dans le français produit par des élèves baoulé. *Le français en Afrique* 3: 159–165.

- Nymansson, Karin. 1995. Le genre grammatical des anglicismes contemporains en français. *Cahiers de lexicologie* 66(1): 95–113.
- Nzesse, Ladislas. 2009. *Le français au cameroun: d'une crise sociopolitique à la vitalité de la langue française (1990-2008)*. UFR Lettres, Arts et sciences humaines.
- Office québécois de la langue française. 2020. Covid-19. Fiche terminologique. http://gdt.oqlf.gouv.qc.ca/fiche0qlf.aspx?Id_Fiche=26557671.
- O'Mahony, Jennifer. 2019. Why the future of French is African. *BBC News* URL <https://www.bbc.com/news/world-africa-47790128>.
- Parker, Enid M. and Richard J. Hayward. 1985. *An Afar-English-French dictionary: With grammatical notes in English*. School of Oriental & African Studies University of London.
- Pergnier, Maurice. 1989. *Les anglicismes: danger ou enrichissement pour la langue française?* Presses universitaires de France.
- Poplack, Shana, Alicia Pousada, and David Sankoff. 1982. Competing influences on gender assignment: Variable process, stable outcome. *Lingua* 57(1): 1–28.
- Poplack, Shana, David Sankoff, and Christopher Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 57: 1–28.
- R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Radio-Canada. 2020. Covid-19 est un terme féminin, et voici pourquoi on vous a dit le contraire. *Radio-Canada* URL <https://ici.radio-canada.ca/nouvelle/1657532/covid-linguistique-virus-epidemie-wuhan-usage>.
- Rey-Debove, Josette. 1987. Effet des anglicismes lexicaux sur le système du français. *Cahiers de lexicologie* 51: 265–273.
- Roché, Michel. 1992. Le masculin est-il plus productif que le féminin? *Langue française* (96): 113–124.
- Sankoff, Gillian and Henrietta Cedergren. 1971. Some results of a sociolinguistic study of Montreal French. *Linguistic diversity in Canadian society* : 61–87.
- Schmidt, Jean. 1990. Panorama des emprunts à l'anglais dans le français d'Afrique. *Bulletin du réseau des observatoires du français contemporain en Afrique noire* (7): 1987–88.
- Smaali, Dalila. 1994. *Les particularités lexicales du français dans la presse algérienne actuelle*. Master's thesis, Université de Provence.
- Soubrier, Jean. 1985. *Le français économique et commercial: ambiguïté d'une langue parallèle*. Ph.D. thesis, Université Lyon 2.
- Spolsky, Bernard. 2018. Language policy in French colonies and after independence. *Current Issues in Language Planning* 19(3): 231–315.
- Surridge, Marie E. 1984. Le genre grammatical des emprunts anglais en français: la perspective diachronique. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 29(1): 58–72.
- Taft, Marcus and Fanny Meunier. 1998. Lexical representation of gender: A quasisregular domain. *Journal of Psycholinguistic research* 27(1): 23–45.
- Telep, Suzie. 2014. Le camfranglais sur internet: pratiques et représentations. *Le français en Afrique* (28): pp–27.
- Trampus, Mitja. 2015. Evaluating language identification performance. URL https://blog.twitter.com/engineering/en_us/a/2015/evaluating-language-identification-performance.html.
- Tremblay, Louis. 1994. *Convergence et divergence dans l'emploi de termes communs recommandés par l'office de la langue française*. Master's thesis, Université Laval.
- Tucker, G. Richard, Wallace E. Lambert, and André A. Rigault. 1977. *The French speaker's skill with grammatical gender: An example of rule-governed behavior*, vol. 8. Walter de Gruyter GmbH & Co KG.
- Tucker, G. Richard, Andro A. Rigault, and Wallace E. Lambert. 1970. Le genre grammatical des substantifs en français: analyse statistique et étude psycholinguistique. *Actes du X^e Congrès International des Linguistes* : 279–290.
- Williams, Jennifer and Charlie Dagli. 2017. Twitter language identification of similar languages and dialects without ground truth. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*. 73–83.
- World Health Organization. 2020a. Novel Coronavirus (2019-nCoV) Situation Report - 22. Press Release.

<https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf>.

World Health Organization. 2020b. Prévention et contrôle des infections (PCI) pour le virus de la COVID-19. <https://openwho.org/courses/COVID-19-IPC-FR>.

Zubiaga, Arkaitz, Inaki San Vicente, Pablo Gamallo, José Ramon Pichel, Inaki Alegria, Nora Aranberri, Aitzol Ezeiza, and Víctor Fresno. 2016. TweetLID: A benchmark for tweet language identification. *Language Resources and Evaluation* 50(4): 729–766.