

# L'évolution du genre du mot "COVID-19" dans le français de trois continents

Une étude de deux corpus médiatiques

---

Michael Dow et Patrick Drouin  
Université de Montréal

Congrès annuel de l'ACL 2021  
6 juin 2021

<https://ling.auf.net/lingbuzz/005711>



# Introduction

---

## Quel est le problème?

- En français, la néologie, par emprunt ou par création, soulève souvent la question du genre grammatical (désormais *genre*).

## Quel est le problème ?

- En français, la néologie, par emprunt ou par création, soulève souvent la question du genre grammatical (désormais *genre*).
- Malgré un biais dans certains aspects du français pour le masculin, plusieurs facteurs peuvent entrer en jeu : lexicaux, sémantiques, phonologiques, socioéconomiques, diatopiques, diastratiques, etc.

## Quel est le problème ?

- En français, la néologie, par emprunt ou par création, soulève souvent la question du genre grammatical (désormais *genre*).
- Malgré un biais dans certains aspects du français pour le masculin, plusieurs facteurs peuvent entrer en jeu : lexicaux, sémantiques, phonologiques, socioéconomiques, diatopiques, diastratiques, etc.
- De plus, la présence d'un organisme exerçant une certaine influence sur la langue peut modifier l'usage (OQLF, Académie française).

## Quel est le problème ?

- En français, la néologie, par emprunt ou par création, soulève souvent la question du genre grammatical (désormais *genre*).
- Malgré un biais dans certains aspects du français pour le masculin, plusieurs facteurs peuvent entrer en jeu : lexicaux, sémantiques, phonologiques, socioéconomiques, diatopiques, diastratiques, etc.
- De plus, la présence d'un organisme exerçant une certaine influence sur la langue peut modifier l'usage (OQLF, Académie française).
- L'émergence rapide et largement documentée du nouveau terme *COVID-19* nous permet d'étudier de près plusieurs de ces phénomènes.

## Objectifs

- Observer l'évolution du genre du terme *COVID-19* en français sur 3 continents : l'Afrique, l'Amérique et l'Europe.

## Objectifs

- Observer l'évolution du genre du terme *COVID-19* en français sur 3 continents : l'Afrique, l'Amérique et l'Europe.
- Constituer un corpus à partir de données de Twitter pour procéder aux observations.



## Objectifs

- Observer l'évolution du genre du terme *COVID-19* en français sur 3 continents : l'Afrique, l'Amérique et l'Europe.
- Constituer un corpus à partir de données de Twitter pour procéder aux observations.
- Comparer les données issues de Twitter à celle de la presse francophone pour la même période.

## Objectifs

- Observer l'évolution du genre du terme *COVID-19* en français sur 3 continents : l'Afrique, l'Amérique et l'Europe.
- Constituer un corpus à partir de données de Twitter pour procéder aux observations.
- Comparer les données issues de Twitter à celle de la presse francophone pour la même période.
- Estimer l'influence/importance du *statut social* (modélisé à partir du nombre d'abonnés).

## Objectifs

- Observer l'évolution du genre du terme *COVID-19* en français sur 3 continents : l'Afrique, l'Amérique et l'Europe.
- Constituer un corpus à partir de données de Twitter pour procéder aux observations.
- Comparer les données issues de Twitter à celle de la presse francophone pour la même période.
- Estimer l'influence/importance du *statut social* (modélisé à partir du nombre d'abonnés).
- Mettre en relation les résultats avec la présence ou l'absence de recommandation de la part d'organismes officiels.

## Intérêt de l'étude

Notre étude se démarque par les éléments suivants :

- La nature du terme COVID-19.

## Intérêt de l'étude

Notre étude se démarque par les éléments suivants :

- La nature du terme COVID-19.
- L'exploration quasi *en temps réel* des données.

## Intérêt de l'étude

Notre étude se démarque par les éléments suivants :

- La nature du terme COVID-19.
- L'exploration quasi *en temps réel* des données.
- Les deux volets étudiés : réseau social et journaux.

## Intérêt de l'étude

Notre étude se démarque par les éléments suivants :

- La nature du terme COVID-19.
- L'exploration quasi *en temps réel* des données.
- Les deux volets étudiés : réseau social et journaux.
- La prise en compte du français africain.

## Intérêt de l'étude

Notre étude se démarque par les éléments suivants :

- La nature du terme COVID-19.
- L'exploration quasi *en temps réel* des données.
- Les deux volets étudiés : réseau social et journaux.
- La prise en compte du français africain.
- L'estimation de l'influence/importance des locuteurs.



## *Caveat auditor*

- La question du genre de “COVID-19” s’est avérée litigieuse dans la sphère publique (par ex., Meteyer, 2020) et suscite même un certain prescriptivisme humoristique chez les linguistes...

# Introduction

## *Caveat auditor*

- La question de la liberté d'expression dans la sphère publique est un sujet qui a certain presc



le litigieuse dans  
discite même un  
nguistes...

## *Caveat auditor*

- La question du genre de “COVID-19” s’est avérée litigieuse dans la sphère publique (par ex., Meteyer, 2020) et suscite même un certain prescriptivisme humoristique chez les linguistes...
- Ce débat n’entre pas dans le cadre de cette présentation et, de plus, les mérites des arguments pour ou contre l’usage féminin de COVID-19 ne nous intéressent pas.

## *Caveat auditor*

- La question du genre de “COVID-19” s’est avérée litigieuse dans la sphère publique (par ex., Meteyer, 2020) et suscite même un certain prescriptivisme humoristique chez les linguistes...
- Ce débat n’entre pas dans le cadre de cette présentation et, de plus, les mérites des arguments pour ou contre l’usage féminin de COVID-19 ne nous intéressent pas.
- Nous ne faisons donc pas de recommandation dans un sens ou dans l’autre.

## *Caveat auditor*

- La question du genre de “COVID-19” s’est avérée litigieuse dans la sphère publique (par ex., Meteyer, 2020) et suscite même un certain prescriptivisme humoristique chez les linguistes...
- Ce débat n’entre pas dans le cadre de cette présentation et, de plus, les mérites des arguments pour ou contre l’usage féminin de COVID-19 ne nous intéressent pas.
- Nous ne faisons donc pas de recommandation dans un sens ou dans l’autre.
- Nous cherchons plutôt à documenter l’usage par le public et les médias et à élucider les causes et/ou les explications potentielles des tendances observées.

## Travaux antérieurs

---

## Attribution de genre aux noms inanimés

- La catégorisation des noms en genres masculin et féminin est opaque comparativement à d'autres langues (par ex., Corbett, 1991) mais suit des régularités phonologiques et morphologiques dans l'ensemble du lexique.

## Attribution de genre aux noms inanimés

- La catégorisation des noms en genres masculin et féminin est opaque comparativement à d'autres langues (par ex., Corbett, 1991) mais suit des régularités phonologiques et morphologiques dans l'ensemble du lexique.
- Plusieurs séquences finales de mots sont prévisibles pour le genre (par ex., [ã] = masculin à 99% vs. [ad] à 12%; Tucker et al., 1977). (NB : [id] est largement masculin dans le lexique.)



## Attribution de genre aux noms inanimés

- La catégorisation des noms en genres masculin et féminin est opaque comparativement à d'autres langues (par ex., Corbett, 1991) mais suit des régularités phonologiques et morphologiques dans l'ensemble du lexique.
- Plusieurs séquences finales de mots sont prévisibles pour le genre (par ex., [ã] = masculin à 99% vs. [ad] à 12%; Tucker et al., 1977). (NB : [id] est largement masculin dans le lexique.)
- Cependant, le statut morphologique doit être considéré, puisque le genre apporté par un suffixe nominal est systématique.

## Attribution de genre aux noms inanimés

- La catégorisation des noms en genres masculin et féminin est opaque comparativement à d'autres langues (par ex., Corbett, 1991) mais suit des régularités phonologiques et morphologiques dans l'ensemble du lexique.
- Plusieurs séquences finales de mots sont prévisibles pour le genre (par ex., [ã] = masculin à 99% vs. [ad] à 12%; Tucker et al., 1977). (NB : [id] est largement masculin dans le lexique.)
- Cependant, le statut morphologique doit être considéré, puisque le genre apporté par un suffixe nominal est systématique.
- Les locuteurs prêtent attention à ces indices dans le traitement d'informations lexicales et dans l'attribution de genre à des néologismes (par ex., Karmiloff-Smith, 1979; Holmes et Segui, 2004).

## Les emprunts contemporains à des langues “sans genre”

- Alors que le lexique est assez équilibré en matière de genre, 85 % de ces emprunts sont masculins (Roché, 1992).

## Les emprunts contemporains à des langues “sans genre”

- Alors que le lexique est assez équilibré en matière de genre, 85 % de ces emprunts sont masculins (Roché, 1992).
- Cette disparité contribue à la perception du masculin comme le genre “par défaut” ou “non marqué” en français.

## Les emprunts contemporains à des langues “sans genre”

- Alors que le lexique est assez équilibré en matière de genre, 85 % de ces emprunts sont masculins (Roché, 1992).
- Cette disparité contribue à la perception du masculin comme le genre “par défaut” ou “non marqué” en français.
- Les mots anglais reçoivent souvent le genre de leurs équivalents français (Haden et Joliat, 1940; Nymansson, 1995; Lupu, 2005), par ex. *une love affair* < *une affaire*.

## Les emprunts contemporains à des langues “sans genre”

- Alors que le lexique est assez équilibré en matière de genre, 85 % de ces emprunts sont masculins (Roché, 1992).
- Cette disparité contribue à la perception du masculin comme le genre “par défaut” ou “non marqué” en français.
- Les mots anglais reçoivent souvent le genre de leurs équivalents français (Haden et Joliat, 1940; Nymansson, 1995; Lupu, 2005), par ex. *une love affair* < *une affaire*.
- Un autre facteur important est l'ellipse avec un nom français non exprimé, par ex. *une (voiture) Ford*.

## Les emprunts contemporains à des langues “sans genre”

- Alors que le lexique est assez équilibré en matière de genre, 85 % de ces emprunts sont masculins (Roché, 1992).
- Cette disparité contribue à la perception du masculin comme le genre “par défaut” ou “non marqué” en français.
- Les mots anglais reçoivent souvent le genre de leurs équivalents français (Haden et Joliat, 1940; Nymansson, 1995; Lupu, 2005), par ex. *une love affair* < *une affaire*.
- Un autre facteur important est l'ellipse avec un nom français non exprimé, par ex. *une (voiture) Ford*.
- L'analogie phonétique joue un rôle réduit (Belleau, 2016), et peut entrer en conflit avec les facteurs ci-dessus, donnant lieu à la variation, par ex. *le/la new beat* (fém. par ellipse, masc. en raison de [it] final) (Nymansson, 1995).

## Différences régionales : Le genre & les emprunts (Europe & Amérique)

- Ces variétés ne présentent pas de différences significatives en ce qui concerne le genre des noms partagés ou l'attribution du genre aux emprunts anglais (Haden et Joliat, 1940; Nymansson, 1995; Belleau, 2016), avec quelques exceptions importantes.



## Différences régionales : Le genre & les emprunts (Europe & Amérique)

- Ces variétés ne présentent pas de différences significatives en ce qui concerne le genre des noms partagés ou l'attribution du genre aux emprunts anglais (Haden et Joliat, 1940; Nymansson, 1995; Belleau, 2016), avec quelques exceptions importantes.
- Le genre de certains mots (par ex. *party*) et morphèmes (par ex. *-ing*) à part, deux facteurs principaux distinguent ces variétés.

## Différences régionales : Le genre & les emprunts (Europe & Amérique)

- Ces variétés ne présentent pas de différences significatives en ce qui concerne le genre des noms partagés ou l'attribution du genre aux emprunts anglais (Haden et Joliat, 1940; Nymansson, 1995; Belleau, 2016), avec quelques exceptions importantes.
- Le genre de certains mots (par ex. *party*) et morphèmes (par ex. *-ing*) à part, deux facteurs principaux distinguent ces variétés.
  - D'abord, les mots anglais à voyelle finale ont tendance à être masculins et ceux à consonne finale féminins en français canadien, contrairement au français européen (Léard, 1995).

## Différences régionales : Le genre & les emprunts (Europe & Amérique)

- Ces variétés ne présentent pas de différences significatives en ce qui concerne le genre des noms partagés ou l'attribution du genre aux emprunts anglais (Haden et Joliat, 1940; Nymansson, 1995; Belleau, 2016), avec quelques exceptions importantes.
- Le genre de certains mots (par ex. *party*) et morphèmes (par ex. *-ing*) à part, deux facteurs principaux distinguent ces variétés.
  - D'abord, les mots anglais à voyelle finale ont tendance à être masculins et ceux à consonne finale féminins en français canadien, contrairement au français européen (Léard, 1995).
  - Enfin, les mots monosyllabiques (par ex., *job*) ont tendance à être masculins en français européen mais féminins en français canadien (Belleau, 2016).

## Différences régionales : Le genre & les emprunts (Afrique)

- On remarque une difficulté à appliquer systématiquement les distinctions de genre chez les apprenants de certaines variétés africaines (Calvet et Dumont, 1969; Biloa, 2003; Holtzer, 2004).

## Différences régionales : Le genre & les emprunts (Afrique)

- On remarque une difficulté à appliquer systématiquement les distinctions de genre chez les apprenants de certaines variétés africaines (Calvet et Dumont, 1969; Biloa, 2003; Holtzer, 2004).
- L'omission des déterminants signalant le genre est une autre caractéristique récurrente dans certaines variétés africaines, entraînant une confusion sur l'utilisation du masculin et du féminin (Herault et Vonrospach, 1967; Ayewa, 2009).

## Différences régionales : Le genre & les emprunts (Afrique)

- On remarque une difficulté à appliquer systématiquement les distinctions de genre chez les apprenants de certaines variétés africaines (Calvet et Dumont, 1969; Biloa, 2003; Holtzer, 2004).
- L'omission des déterminants signalant le genre est une autre caractéristique récurrente dans certaines variétés africaines, entraînant une confusion sur l'utilisation du masculin et du féminin (Herault et Vonrospach, 1967; Ayewa, 2009).
- Cette omission se produit dans le français de Côte d'Ivoire (Jabet, 2006; Boutin, 2007) et dans le *camfrançais*, une langue mixte du Cameroun (de Féral, 2006; Telep, 2014).

## Différences régionales : Le genre & les emprunts (Afrique)

- On remarque une difficulté à appliquer systématiquement les distinctions de genre chez les apprenants de certaines variétés africaines (Calvet et Dumont, 1969 ; Biloa, 2003 ; Holtzer, 2004).
- L'omission des déterminants signalant le genre est une autre caractéristique récurrente dans certaines variétés africaines, entraînant une confusion sur l'utilisation du masculin et du féminin (Herault et Vonrospach, 1967 ; Ayewa, 2009).
- Cette omission se produit dans le français de Côte d'Ivoire (Jabet, 2006 ; Boutin, 2007) et dans le *camfrançais*, une langue mixte du Cameroun (de Féral, 2006 ; Telep, 2014).
- Nous n'avons pas pu trouver de synthèse sur le genre d'emprunts anglais, mais nous notons quelques emprunts féminins dans certaines variétés, sans pouvoir en déduire des tendances significatives.

## Une brève histoire du genre d'une maladie

- Le 11 février 2020, l'Organisation mondiale de la santé (OMS) a donné à la maladie causée par le SRAS-CoV-2 le nom abrégé de "COVID-19" (World Health Organization, 2020), et le terme était généralement masculin jusqu'au 6 mars, quand l'OMS a opté pour le féminin (Avanzi, 2020).



## Une brève histoire du genre d'une maladie

- Le 11 février 2020, l'Organisation mondiale de la santé (OMS) a donné à la maladie causée par le SRAS-CoV-2 le nom abrégé de "COVID-19" (World Health Organization, 2020), et le terme était généralement masculin jusqu'au 6 mars, quand l'OMS a opté pour le féminin (Avanzi, 2020).
- Le même jour, Radio-Canada (Bonsaint, courriel) et l'OQLF (Darras, courriel) ont mis à jour leurs fiches terminologiques dans le même but (Radio-Canada, 2020 ; Office québécois de la langue française, 2020).

## Une brève histoire du genre d'une maladie

- Le 11 février 2020, l'Organisation mondiale de la santé (OMS) a donné à la maladie causée par le SRAS-CoV-2 le nom abrégé de "COVID-19" (World Health Organization, 2020), et le terme était généralement masculin jusqu'au 6 mars, quand l'OMS a opté pour le féminin (Avanzi, 2020).
- Le même jour, Radio-Canada (Bonsaint, courriel) et l'OQLF (Darras, courriel) ont mis à jour leurs fiches terminologiques dans le même but (Radio-Canada, 2020 ; Office québécois de la langue française, 2020).
- Ce n'est que le 7 mai que l'Académie française a émis une recommandation du féminin (Académie française, 2020).

## Une brève histoire du genre d'une maladie

- Le 11 février 2020, l'Organisation mondiale de la santé (OMS) a donné à la maladie causée par le SRAS-CoV-2 le nom abrégé de "COVID-19" (World Health Organization, 2020), et le terme était généralement masculin jusqu'au 6 mars, quand l'OMS a opté pour le féminin (Avanzi, 2020).
- Le même jour, Radio-Canada (Bonsaint, courriel) et l'OQLF (Darras, courriel) ont mis à jour leurs fiches terminologiques dans le même but (Radio-Canada, 2020 ; Office québécois de la langue française, 2020).
- Ce n'est que le 7 mai que l'Académie française a émis une recommandation du féminin (Académie française, 2020).
- Dans le trois cas, le raisonnement pour l'emploi du féminin est le même : le référent de base est féminin (*maladie*), qu'il soit exprimé ou non.

# Méthodologie

---

## Remarques générales

- Par souci de concision, nous employons les dénominations *Amérique*, *Afrique* et *Europe*.

## Remarques générales

- Par souci de concision, nous employons les dénominations *Amérique*, *Afrique* et *Europe*.
- Chaque région est hétérogène et implique une variation, mais des échantillons de données ont été vérifiés manuellement.

## Remarques générales

- Par souci de concision, nous employons les dénominations *Amérique*, *Afrique* et *Europe*.
- Chaque région est hétérogène et implique une variation, mais des échantillons de données ont été vérifiés manuellement.
- La mise en commun des données de divers pays est inévitable pour des raisons pratiques et méthodologiques : rareté des données, généralisation, etc.

## Remarques générales

- Les données ont récupérées à partir de la compilation COVID-19-TweetIDs (Chen et al., 2020).



## Remarques générales

- Les données ont récupérées à partir de la compilation COVID-19-TweetIDs (Chen et al., 2020).
- Notre jeu de données couvre du 21 janvier 2020 à la fin juin 2020.

## Remarques générales

- Les données ont récupérées à partir de la compilation COVID-19-TweetIDs (Chen et al., 2020).
- Notre jeu de données couvre du 21 janvier 2020 à la fin juin 2020.
- La part du français était (23 juin) estimée à 3%, soit 5,5M de tweets par les auteurs du jeu de données.

## Préparation des données

- Récupération des données (*hydrating*) avec Python.

## Préparation des données

- Récupération des données (*hydrating*) avec Python.
- **Délestage** des tweets dans les langues autres que le français.

## Préparation des données

- Récupération des données (*hydrating*) avec Python.
- **Délestage** des tweets dans les langues autres que le français.
- Élimination des doublons et des reprises (RT).

## Préparation des données

- Récupération des données (*hydrating*) avec Python.
- **Délestage** des tweets dans les langues autres que le français.
- Élimination des doublons et des reprises (RT).
- Nettoyage des tweets (URL, caractères Unicode, etc.)

## Préparation des données

- Récupération des données (*hydrating*) avec Python.
- **Délestage** des tweets dans les langues autres que le français.
- Élimination des doublons et des reprises (RT).
- Nettoyage des tweets (URL, caractères Unicode, etc.)
- Identification et normalisation de l'**origine géographique**.

## Préparation des données

- Récupération des données (*hydrating*) avec Python.
- **Délestage** des tweets dans les langues autres que le français.
- Élimination des doublons et des reprises (RT).
- Nettoyage des tweets (URL, caractères Unicode, etc.)
- Identification et normalisation de l'**origine géographique**.
- Sélection des tweets où une marque de genre était observable (*le, au, du, ce, la* et *cette*).



## Préparation des données

- Récupération des données (*hydrating*) avec Python.
- **Délestage** des tweets dans les langues autres que le français.
- Élimination des doublons et des reprises (RT).
- Nettoyage des tweets (URL, caractères Unicode, etc.)
- Identification et normalisation de l'**origine géographique**.
- Sélection des tweets où une marque de genre était observable (*le, au, du, ce, la* et *cette*).
- Récupération des dates de publication.

## Préparation des données

- Récupération des données (*hydrating*) avec Python.
- **Délestage** des tweets dans les langues autres que le français.
- Élimination des doublons et des reprises (RT).
- Nettoyage des tweets (URL, caractères Unicode, etc.)
- Identification et normalisation de l'**origine géographique**.
- Sélection des tweets où une marque de genre était observable (*le, au, du, ce, la et cette*).
- Récupération des dates de publication.
- Regroupement des *locuteurs* et fonction de la taille de leur public en **3 groupes** : petit, moyen, grand.

## Préparation des données

- Récupération des données (*hydrating*) avec Python.
- **Délestage** des tweets dans les langues autres que le français.
- Élimination des doublons et des reprises (RT).
- Nettoyage des tweets (URL, caractères Unicode, etc.)
- Identification et normalisation de l'**origine géographique**.
- Sélection des tweets où une marque de genre était observable (*le, au, du, ce, la et cette*).
- Récupération des dates de publication.
- Regroupement des *locuteurs* et fonction de la taille de leur public en **3 groupes** : petit, moyen, grand.
- Chaque groupe d'un même continent contient environ  $1/3$  des observations (tweets).

## Préparation des données : données géographiques

- Seulement environ 1% des tweets comportaient des métadonnées sur leur origine géographique (`place.country`).

## Préparation des données : données géographiques

- Seulement environ 1% des tweets comportaient des métadonnées sur leur origine géographique (**place.country**).
- Afin de compléter les données, nous avons aussi analysé le champ **user.location**.

## Préparation des données : données géographiques

- Seulement environ 1% des tweets comportaient des métadonnées sur leur origine géographique (**place.country**).
- Afin de compléter les données, nous avons aussi analysé le champ **user.location**.
- Le format de ce dernier champ étant libre, on observe une grande variation et beaucoup de créativité.

## Préparation des données : données géographiques

- Seulement environ 1% des tweets comportaient des métadonnées sur leur origine géographique (**place.country**).
- Afin de compléter les données, nous avons aussi analysé le champ **user.location**.
- Le format de ce dernier champ étant libre, on observe une grande variation et beaucoup de créativité.
- En conséquence, à des fins de normalisation, nous avons utilisé des expressions régulières afin de relier le contenu de ce champ à celui d'une base de données géographiques personnalisée pour enfin identifier le continent d'origine.

## Préparation des données : données géographiques

- Seulement environ 1% des tweets comportaient des métadonnées sur leur origine géographique (**place.country**).
- Afin de compléter les données, nous avons aussi analysé le champ **user.location**.
- Le format de ce dernier champ étant libre, on observe une grande variation et beaucoup de créativité.
- En conséquence, à des fins de normalisation, nous avons utilisé des expressions régulières afin de relier le contenu de ce champ à celui d'une base de données géographiques personnalisée pour enfin identifier le continent d'origine.
- Une validation manuelle indique que les résultats sont bons (Afrique à 87.3%, Amérique à 95.3% et Europe à 98.7%).



## Préparation des données : données géographiques

- Les données analysées proviennent de la base de données Eureka.cc, un agrégateur de journaux.

## Préparation des données : données géographiques

- Les données analysées proviennent de la base de données Eureka.cc, un agrégateur de journaux.
- Les mêmes indicateurs de genre ont été utilisés pour compiler des données hebdomadaires pour une période identique à celle couverte par le corpus Twitter.

## Préparation des données : données géographiques

- Les données analysées proviennent de la base de données Eureka.cc, un agrégateur de journaux.
- Les mêmes indicateurs de genre ont été utilisés pour compiler des données hebdomadaires pour une période identique à celle couverte par le corpus Twitter.
- Les continents ont été délimités à l'aide des filtres de recherche d'Eureka.cc.

## Préparation des données : données géographiques

- Les données analysées proviennent de la base de données Eureka.cc, un agrégateur de journaux.
- Les mêmes indicateurs de genre ont été utilisés pour compiler des données hebdomadaires pour une période identique à celle couverte par le corpus Twitter.
- Les continents ont été délimités à l'aide des filtres de recherche d'Eureka.cc.
- Le nombre de sources disponibles au moment de l'interrogation était le suivant : 78 pour l'Afrique, 653 pour l'Amérique et 825 pour l'Europe.

## Préparation des données : données géographiques

- Les données analysées proviennent de la base de données Eureka.cc, un agrégateur de journaux.
- Les mêmes indicateurs de genre ont été utilisés pour compiler des données hebdomadaires pour une période identique à celle couverte par le corpus Twitter.
- Les continents ont été délimités à l'aide des filtres de recherche d'Eureka.cc.
- Le nombre de sources disponibles au moment de l'interrogation était le suivant : 78 pour l'Afrique, 653 pour l'Amérique et 825 pour l'Europe.
- Le nombre d'articles pour chaque requête a été retenu et non le nombre d'occurrences.

## Préparation des données : données géographiques

- Les données analysées proviennent de la base de données Eureka.cc, un agrégateur de journaux.
- Les mêmes indicateurs de genre ont été utilisés pour compiler des données hebdomadaires pour une période identique à celle couverte par le corpus Twitter.
- Les continents ont été délimités à l'aide des filtres de recherche d'Eureka.cc.
- Le nombre de sources disponibles au moment de l'interrogation était le suivant : 78 pour l'Afrique, 653 pour l'Amérique et 825 pour l'Europe.
- Le nombre d'articles pour chaque requête a été retenu et non le nombre d'occurrences.
- En raison des limites du moteur de recherche, nous n'avons pas pu éliminer les fils de presse du corpus.

# Résultats

---

	Afrique	Amérique	Europe	Total
Février	47	64	417	528
Mars	904	728	4 636	6 268
Avril	2 270	1 460	9 367	13 097
Mai	2 385	1 817	10 169	14 371
Juin	6 349	5 197	30 244	41 790
<b>Total</b>	<b>11 955</b>	<b>9 266</b>	<b>54 833</b>	<b>76 054</b>

**TABLE 1 :** Nombre de tweets par continent et par mois



Afrique	Amérique	Europe	Total
6 649	4 712	32 767	44 128

**TABLE 2 :** Nombre de contributeurs par continent

	Afrique	Amérique	Europe
Petit	0-213	0-285	0-196
Moyen	214-1 558	286-1 595	197-1 017
Grand	1 559+	1 596+	1 018+

**TABLE 3 :** Taille des regroupements (selon le # d'abonnés) par continent

# Résultats - Twitter

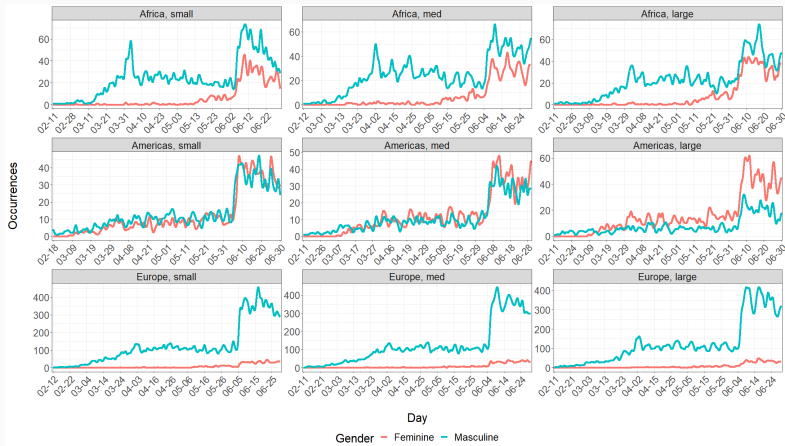


FIGURE 1 : Occurrences de “COVID” masculin/féminin

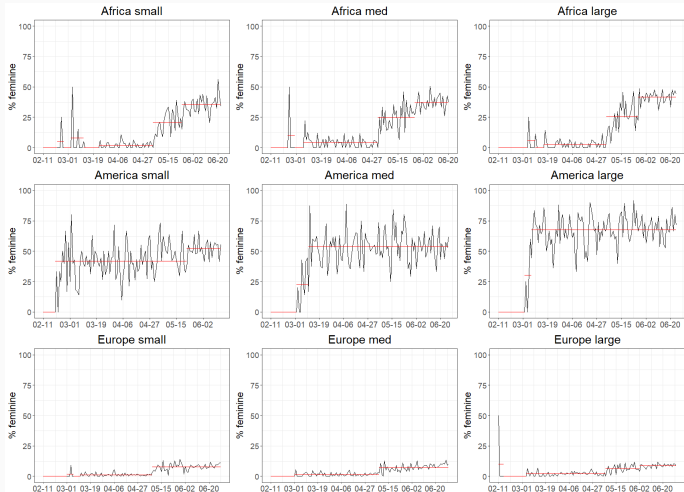


FIGURE 2 : Pourcentage d'occurrences au fém. avec points de rupture

# Résultats - Journaux

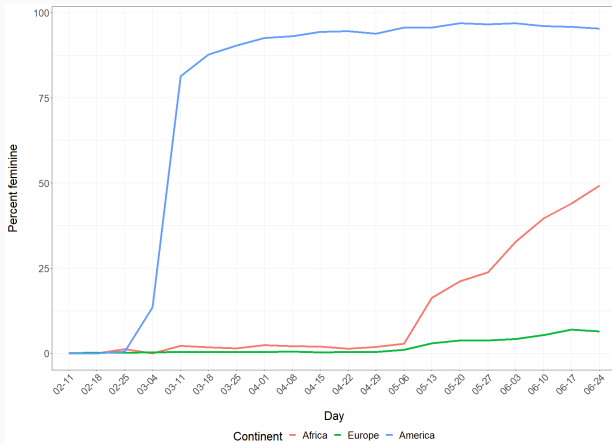


FIGURE 3 : Pourcentage d'occurrences au fém. par semaine, par continent

## Conclusion

---

## Observations

- Les utilisateurs Twitter et les médias américains ont adopté très soudainement le féminin à la suite des recommandations de l'OMS, de Radio-Canada et de l'Office québécois de la langue Française dès mars 2020.

## Observations

- Les utilisateurs Twitter et les médias américains ont adopté très soudainement le féminin à la suite des recommandations de l'OMS, de Radio-Canada et de l'Office québécois de la langue Française dès mars 2020.
- Cet usage est désormais stable.

## Observations

- Les utilisateurs Twitter et les médias américains ont adopté très soudainement le féminin à la suite des recommandations de l'OMS, de Radio-Canada et de l'Office québécois de la langue Française dès mars 2020.
- Cet usage est désormais stable.
- Du côté de l'Afrique, l'usage féminin a connu une augmentation importante à la suite de la recommandation de l'Académie française en mai 2020.



## Observations

- Les utilisateurs Twitter et les médias américains ont adopté très soudainement le féminin à la suite des recommandations de l'OMS, de Radio-Canada et de l'Office québécois de la langue Française dès mars 2020.
- Cet usage est désormais stable.
- Du côté de l'Afrique, l'usage féminin a connu une augmentation importante à la suite de la recommandation de l'Académie française en mai 2020.
- Enfin, l'usage féminin est toujours négligeable dans l'ensemble du corpus européen où le masculin semble bien implanté.

## Récapitulatif du travail effectué

- Nous avons récupéré et reconstruit le corpus COVID-19-TweetIDs (Chen et al., 2020).
- Notre jeu de données couvre du 21 janvier 2020 à la fin juin 2020.
- Nous avons mis en place un corpus français de 5,5M de tweets pour étudier l'évolution du genre grammatical de COVID-19 quasi en temps réel.
- Après filtrage nous avons obtenu un corpus de 76 054 tweets *géolocalisés* non ambigus pour le genre de COVID-19.
- Ce corpus nous a permis de comparer l'usage de trois grandes régions : l'Afrique, l'Amérique et l'Europe.
- Les observations faites sur le corpus Twitter ont été mis en correspondance avec des observations faites sur des journaux pour la même période.

## Explications potentielles

- Un variété de facteurs (linguistiques et extralinguistiques) peut venir influencer l'usage.

## Explications potentielles

- Un variété de facteurs (linguistiques et extralinguistiques) peut venir influencer l'usage.
- L'attribution du genre grammatical n'est pas la même dans tous les pays francophones et les variétés de français, plus particulièrement en ce qui a trait aux emprunts.

## Explications potentielles

- Un variété de facteurs (linguistiques et extralinguistiques) peut venir influencer l'usage.
- L'attribution du genre grammatical n'est pas la même dans tous les pays francophones et les variétés de français, plus particulièrement en ce qui a trait aux emprunts.
- Le rôle joué par les organismes exerçant une influence sur la langue n'est pas le même d'une région à une autre.

## Explications potentielles

- Un variété de facteurs (linguistiques et extralinguistiques) peut venir influencer l'usage.
- L'attribution du genre grammatical n'est pas la même dans tous les pays francophones et les variétés de français, plus particulièrement en ce qui a trait aux emprunts.
- Le rôle joué par les organismes exerçant une influence sur la langue n'est pas le même d'une région à une autre.
- L'attitude face aux recommandation des ces organismes varie elle aussi selon la région, tel qu'il a été observé par Kim (2017).

## Explications potentielles

- Un variété de facteurs (linguistiques et extralinguistiques) peut venir influencer l'usage.
- L'attribution du genre grammatical n'est pas la même dans tous les pays francophones et les variétés de français, plus particulièrement en ce qui a trait aux emprunts.
- Le rôle joué par les organismes exerçant une influence sur la langue n'est pas le même d'une région à une autre.
- L'attitude face aux recommandation des ces organismes varie elle aussi selon la région, tel qu'il a été observé par Kim (2017).
- Le long délai entre la recommandation de l'OMS et celle de *l'Académie* d'utiliser le féminin a donné à l'usage masculin le temps de s'implanter en Europe.

# Remerciements

---



## Remerciements

- **Nathalie Bonsaint**, consultante linguistique, Service linguistique, Radio-Canada.
- **Xavier Darras**, coordonnateur de la production linguistique de Direction de l'expertise linguistique de l'Office québécois de la langue française.

# Références

---

Académie française. 2020. Le covid 19 ou la covid 19. Fiche terminologique.

<http://www.academie-francaise.fr/le-covid-19-ou-la-covid-19>.

Avanzi, Mathieu. 2020. Le/la covid? Réouvrir ou rouvrir? Les leçons de grammaire du coronavirus.

*The Conversation* URL [https://theconversation.com/](https://theconversation.com/le-la-covid-reouvrir-ou-rouvrir-les-lecons-de-grammaire-du-coronavi)

[le-la-covid-reouvrir-ou-rouvrir-les-lecons-de-grammaire-du-coronavi](https://theconversation.com/le-la-covid-reouvrir-ou-rouvrir-les-lecons-de-grammaire-du-coronavi)

Ayewa, Kouassi Noël. 2009. Une enquête linguistique : le français, une langue ivoirienne. *Le français en Afrique* (25) : 117–134.

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*. Mémoire de maîtrise, Université Laval.

Bilola, Edmond. 2003. *La langue française au Cameroun : analyse linguistique et didactique*. Peter Lang.

Boutin, Akissi. 2007. Déterminant zéro ou omission du déterminant en français de Côte d'Ivoire. *Le français en Afrique* 22 : 161–182.

Calvet, Maurice et Pierre Dumont. 1969. Le français au Sénégal : interférences du wolof dans le français des élèves sénégalais. *Collection IDERIC* 7(1) : 71–90.

- Chen, Emily, Kristina Lerman, et 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.
- Corbett, Greville G. 1991. *Gender*. Cambridge University Press.
- de Féral, Carole. 2006. Étudier le camfranglais : recueil des données et transcription. *Le français en Afrique* 21 : 211–218.
- Haden, Ernest F et 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.
- Herault, Georges et Jean-Paul Vonrosnach. 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.
- Holmes, Virginia M et 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.
- Jabet, Marita. 2006. Noms sans déterminant en français abidjanais : trait sociolinguistique, sémantique et/ou pragmatique ? *Le français en Afrique* 21 : 325–337.

- Karmiloff-Smith, Anette. 1979. *A Functional Approach to Language Acquisition*. Cambridge : Cambridge University Press.
- Kim, Minchai. 2017. *Variation terminologique en francophonie : Élaboration d'un modèle d'analyse des facteurs d'implantation terminologique*. Thèse de doctorat, Université Paris Sorbonne.
- Léard, Jean-Marcel. 1995. *Grammaire québécoise d'aujourd'hui : comprendre les québécismes*. Guérin universitaire.
- Lupu, Mihaela. 2005. La masculinisation du lexique français : le rôle catalyseur des anglicismes. *Analele Universității „Alexandru Ioan Cuza” din Iași. Secțiunea IIIe. Lingvistică* .
- Meteyer, Madeleine. 2020. Les gens qui disent LA covid sont-ils seulement des snobs? *Le Figaro* .
- Nymansson, Karin. 1995. Le genre grammatical des anglicismes contemporains en français. *Cahiers de lexicologie* 66(1) : 95–113.
- Office québécois de la langue française. 2020. Covid-19. Fiche terminologique. [http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id\\_Fiche=26557671](http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=26557671).
- 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>.

- Roché, Michel. 1992. Le masculin est-il plus productif que le féminin? *Langue française* (96) : 113–124.
- Telep, Suzie. 2014. Le camfranglais sur internet : pratiques et représentations. *Le français en Afrique* (28) : pp–27.
- Tucker, G. Richard, Wallace E. Lambert, et 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.
- World Health Organization. 2020. Prévention et contrôle des infections (PCI) pour le virus de la COVID-19. <https://openwho.org/courses/COVID-19-IPC-FR>.