

Science virale: comment la COVID-19 a changé la représentation médiatique de la science

Annexe technique

François Claveau, Jean-Hugues Roy et Olivier Santerre

Décembre 2022

Contents

Introduction	1
Tâches préliminaires	1
Dictionnaires	2
Corpus Facebook et mesure des interactions	4
Information sur les corpus	4
Corpus médiatique de l'OCI	4
Corpus Facebook	9
Information sur la COVID-19 en général	13
Progression de la pandémie au Québec	13
Mentions de la COVID dans notre 1er corpus	14
Plus de science?	15
Part du contenu médiatique	15
Intensité de l'intérêt du public pour la science	25
Quelle science?	70
Verbes spécifiques à la période COVID	70
Moteur de recherche sémantique	72
Test de la prévalence des champs sémantiques liés au politique	79
La présence de l'incertitude	93
Session R utilisée	99

Introduction

Ce document est l'annexe technique d'un article pour le numéro spécial « Médiatiser la pandémie » de la Revue canadienne de communication (*Canadian Journal of Communication*), à paraître en 2023.

Vous lisez ce document en format PDF. **Notez que la version HTML de l'annexe est plus facile et agréable à consulter, voir [ici](#).**

La fonction principale de ce document, dans l'esprit de la science ouverte, est de montrer comment nos résultats sont produits, bien que nos données brutes ne puissent pas être mises dans un dépôt public pour des

raisons de propriété intellectuelle. Si vous avez besoin de données pour reproduire certains résultats, merci de contacter François Claveau.

La fonction secondaire de cette annexe est de fournir des explications et des résultats additionnels qui n'entraient pas dans l'article.

Tâches préliminaires

Chargement des paquets. Pour l'information complète sur la session R utilisée, voir à la toute fin de ce document.

```
package_list <- c("ggplot2", "dplyr", "mongolite", "data.table", "tidyr", "stringr",
                  "ggrepel", "fuzzyjoin", "anytime", "topicmodels", "tm", "knitr",
                  "tidytext", "doParallel", "zoo", "lubridate", "udpipe",
                  "splitstackshape", "Matrix", "text2vec", "stargazer",
                  "sandwich", "mgcv", "viridis", "directlabels", "kableExtra",
                  "quantda.textplots")
for(p in package_list){
  if (p %in% installed.packages()==FALSE){install.packages(p,dependencies = TRUE)}
  library(p,character.only=TRUE)
}
```

Création de variables globales et chargement de la liste des médias.

```
accessible_data_folder <- "data/"
data_folder <- "/data/covid/"

liste_media <- read.csv(paste0(accessible_data_folder, "liste_finale_medias.csv")) %>%
  data.table()
liste_media <- liste_media[dans_oci == TRUE]

size_axis_text <- 11
pos_vline <- as.Date("2020-03-14")

fig_dir <- "figs/"
dir.create(fig_dir, showWarnings = FALSE)

Sys.setlocale("LC_TIME", "fr_CA.UTF-8")
```

Chargement de fonctions.

```
source("scripts/FCT_util.R")
```

Prétraitement des corpus.

```
# OCI données qui peuvent être partagées
source("scripts/01-OCI_data_create_sharable_object.R")

# OCI données sur le texte brut (ne peut être partagé)
source("scripts/02-Importing_OCI_data.R")
source("scripts/03-Cleaning.R")
source("scripts/04-Morphosyntactic_analysis.R")

# Production d'un objet qui peut être partagé à partir du corpus Facebook
source("scripts/05-Facebook_corpus_preparation.R")
```

Pour ce prétraitement, notons qu'une analyse morphosyntaxique est effectuée sur les paragraphes détectés comme étant en lien avec la science. L'analyse est faite avec la bibliothèque R `udpipe`. Elle nous permet de faire de l'étiquetage syntaxique (p. ex., étiqueter les verbes) et de lemmatiser chaque mot (c'est-à-dire, ramener tous les verbes à l'infinitif ainsi que tous les noms et adjectifs à leur forme neutre (masculin singulier)).

Dictionnaires

Dans le prétraitement, nous utilisons des dictionnaires pour identifier les textes liés à la science et ceux explicitement liés à la COVID-19.

Pour la catégorie « science », l'expression régulière utilisée est `(\s)scien|chercheu` avec une recherche insensible à la casse. Dans une première itération de notre analyse, nous avons inclus le terme « étude* ». Comme il produisait trop de faux positifs, nous l'avons supprimé. Quelques faux positifs demeurent avec nos deux termes, c'est-à-dire des expressions qui correspondent à notre expression régulière, mais qui n'indiquent pas un véritable lien avec la science. Une inspection détaillée du corpus nous a permis de détecter trois cas principaux. Les deux premiers cas sont les expressions « science-fiction » et « chercheur d'emploi » (et ses variantes). De plus, l'expression « science politique » est détectée, bien qu'une proportion significative de ces mentions ne serait pas généralement considérée comme liée à la science. Comme toute méthode de détection doit accepter une précision imparfaite, la question est surtout de voir si les faux positifs sont particulièrement nombreux. Pour répondre à cette question, nous avons calculé la proportion de paragraphes et d'articles qui ne seraient plus associés à la science si nous raffinions notre méthode pour exclure les trois cas principaux.

```
# Chargement des paragraphes identifiés comme scientifiques:
load(paste0(data_folder, "dt_science.RData"))

# Regex pour couvrir les faux positifs les plus usités:
string_to_remove1 <- "science-fiction|chercheu(r|s)[e]?[s]?\\sd['|']emploi"
string_to_remove2 <- "science[s]?\\spolitique[s]?"

# Regex pour détecter le contenu scientifique
initial_string <- "(\\s)scien|chercheu"

# Création de paragraphes sans le 1er faux positif
dt_science[, para_wo_false_pos := str_remove_all(tolower(paragraphe), string_to_remove1)]

# Test si encore regex de détection quand faux positifs sont enlevés
dt_science[, with_sci_after_fp_removed := grepl(initial_string, para_wo_false_pos)]

# Proportion de faux positifs
prop_para_fp1 <- dt_science[with_sci_after_fp_removed == FALSE, .N]/dt_science[, .N]
prop_art_fp1 <- dt_science[with_sci_after_fp_removed == FALSE,
                           unique(art_ID) %>% length]/dt_science[, unique(art_ID) %>%
                                                                    length]

# On applique la même chose avec le 2e faux positif
dt_science[, para_wo_false_pos := str_remove_all(para_wo_false_pos, string_to_remove2)]

# Test si encore regex de détection quand faux positifs sont enlevés
dt_science[, with_sci_after_fp_removed := grepl(initial_string, para_wo_false_pos)]

# Proportion de faux positifs
prop_para_fp2 <- dt_science[with_sci_after_fp_removed == FALSE, .N]/dt_science[, .N]
prop_art_fp2 <- dt_science[with_sci_after_fp_removed == FALSE,
                           unique(art_ID) %>% length]/dt_science[, unique(art_ID) %>%
```

length]

rm(dt_science)

Nous trouvons que les expressions « science-fiction » et « chercheur d'emploi » génèrent ensemble une identification de 2.9% des paragraphes pris étant liés à la science (et 4.2% pour l'ensemble des articles). Lorsque l'on prend en compte « science politique », la part des faux positifs monte à 4.8% pour les paragraphes (et 7.4% pour les articles), bien qu'une proportion non déterminée de ces articles parlent vraiment de recherche en science politique, et sont donc de vrais positifs. Ceci pris en compte, nous sommes très confortablement en haut d'une précision de 90%. Nous avons donc maintenu notre dictionnaire simple sans ajouter des critères d'exclusion sur la base d'expressions régulières.

Notre dictionnaire est similaire à celui utilisé par Chinn et ses collègues (2020) pour identifier les mentions scientifiques dans leur étude sur la politisation et la polarisation de la couverture du changement climatique dans les médias américains entre 1985 et 2017. Ils n'ont utilisé que trois chaînes de caractères : *scientist*, *research* et *professor*. Nous ne pouvions pas inclure ce dernier terme dans un corpus francophone, car le mot « professeur » peut aussi référer à des textes sur l'enseignement primaire ou secondaire.

Il convient également de préciser que notre dictionnaire n'identifie pas tous les textes à contenu scientifique. Par exemple, un article pourrait utiliser des termes spécifiques à un domaine, comme « microbiologie » ou « physicien », et ne pas être pris en compte par notre dictionnaire. Le problème de l'ajout de termes spécifiques à un domaine est que, d'une part, il faut en ajouter beaucoup pour éviter les biais qui seraient alors dus à la sous-représentation (ou la sur-représentation) de certains domaines et, d'autre part, la multiplication des termes génère un grand nombre de faux positifs. Nous nous en sommes donc tenus à deux termes.

Pour la catégorie « COVID-19 », la recherche insensible à la casse utilise l'expression:

`virus|covid|sras-cov|sars-cov`

Cette expression semble produire très peu de faux positifs lorsque l'expression est utilisée sur le corpus 2020.

Corpus Facebook et mesure des interactions

CrowdTangle ne donne accès qu'au contenu qui a été publié par les organisations médiatiques elles-mêmes sur leurs propres pages. Il n'englobe pas les articles ou vidéos de médias que des abonnés Facebook peuvent publier directement sur leur fil d'actualité, sur d'autres pages ou sur des groupes.

CrowdTangle renvoie, pour chaque publication, trois variables principales :

1. le nombre de fois où la publication a été partagée;
2. le nombre de réactions à cette publication (« j'aime », ainsi que les émoticônes « love », « wow », « haha », « tristesse », « colère » et « solidarité »);
3. le nombre de commentaires laissés par d'autres abonnés Facebook (les réactions à ces commentaires ne sont pas comptées).

La somme de ces mesures constitue ce que Facebook appelle *engagement* et que nous traduisons par « interactions ». Il faut noter que les données d'interactions sont différentes de la « portée » (*reach*) ou des « impressions » et ne permettent pas de les prédire. Elles donnent une idée du nombre d'utilisateurs de Facebook qui ont interagi avec un contenu donné (en le partageant, en le commentant ou en cliquant sur une icône de réaction), et non du nombre de personnes qui ont vu ou lu ce contenu.

Table 1: Les articles du corpus de l'OCI divisés selon nos deux variables dichotomiques

Statut	Nombre
Sans science avant COVID	665520
Science avant COVID	32996
Sans science pendant COVID	247001
Science pendant COVID	17310

Information sur les corpus

Corpus médiatique de l'OCI

Notre premier corpus est tiré de l'Observatoire de la circulation de l'information. Voici quelques données générales sur ce corpus.

```
# Chargement des données créées par le script 01
dt_full_text <- fread(paste0(accessible_data_folder, "dt_full_text_for_count_sci.csv"))

dt_full_text <- dt_full_text[month(date) %in% 1:6 &
                             year(date) %in% 2017:2020] # première moitié des années
dt_full_text[is_science == TRUE & year(date) ==2020, Status := "Science pendant COVID"]
dt_full_text[is_science == TRUE & year(date) <2020, Status := "Science avant COVID"]
dt_full_text[is_science == FALSE & year(date) <2020, Status := "Sans science avant COVID"]
dt_full_text[is_science == FALSE & year(date) ==2020,
              Status := "Sans science pendant COVID"]

count <- dt_full_text[,list(doc_ID, Status)] %>% unique()

caption_txt <-
  "Les articles du corpus de l'OCI divisés selon nos deux variables dichotomiques"

kable(count[,.N, by = Status][,list(Statut= Status, Nombre = N)],
       caption = caption_txt) %>% kable_styling("hover")

caption_txt <- "Les articles de l'OCI par média francophone généraliste
(ordre inversé du nombre total d'articles)"

if(outputFormat == "html"){
  dt_full_text[,list(`Nombre total` = .N, `Lié à la science` = sum(is_science),
                    `Pendant COVID` = sum(is_covid)),by = root_url][order(-`Nombre total`)] %>%
    kable(caption = caption_txt) %>% kable_styling() %>%
    scroll_box(width = "100%", height = "200px")
} else{
  dt_full_text[,list(`Nombre total` = .N, `Lié à la science` = sum(is_science),
                    `Pendant COVID` = sum(is_covid)),by = root_url][order(-`Nombre total`)] %>%
    kable(caption = caption_txt, booktabs = TRUE, longtable = TRUE) %>% kable_styling()
}
```

Table 2: Les articles de l'OCI par média francophone généraliste
(ordre inversé du nombre total d'articles)

root_url	Nombre total	Lié à la science	Pendant COVID
www.lapresse.ca	112937	7233	9755

www.journaldemontreal.com	103204	5020	10179
www.journaldequebec.com	101236	4900	9715
ici.radio-canada.ca	84064	4877	12249
journalmetro.com	50518	2624	2589
www.ledevoir.com	40893	4681	2967
www.tvnouvelles.ca	40377	2189	6395
www.lesoleil.com	29470	2088	4853
www.latribune.ca	22984	1387	3856
quebec.huffingtonpost.ca	20752	2084	2881
www.lenouvelliste.ca	20473	1149	3720
www.lequotidien.com	19196	1181	3246
www.lavoixdelest.ca	18397	1342	3606
www.lanouvelle.net	13997	343	1366
lactualite.com	12368	1504	2939
www.fm93.com	12089	109	457
www.fm1069.ca	11167	109	528
www.lecourrierdusud.ca	8488	281	547
www.lavantage.qc.ca	7833	161	53
www.lereflet.qc.ca	7614	156	569
www.journalexpress.ca	7280	226	703
www.cybersoleil.com	6819	146	535
www.infodimanche.com	6449	196	518
www.cihofm.com	5824	65	442
www.lhebdojournal.com	5337	206	366
www.enbeauce.com	4807	133	666
www.chga.fm	4754	33	176
www.fm1033.ca	4699	75	46
chipfm.com	4644	37	229
www.lecourriersud.com	4629	142	567
www.courrierfrontenac.qc.ca	4536	134	372
www.lhebdo dustmaurice.com	4394	145	348
www.laction.com	4186	95	401
www.journallenord.com	3936	136	237
www.lerefletdulac.com	3762	82	233
www.lavoixdusud.com	3702	90	279
www.leplacoteux.com	3632	118	312
lenord-cotier.com	3614	87	691
courrierlaval.com	3568	109	12
www.lechodemaskinonge.com	3559	116	343
www.granbyexpress.com	3486	99	367
www.lemanic.ca	3409	115	700
www.les2rives.com	3390	119	24
www.lechodelatuque.com	3387	75	337
www.leclaireurprogres.ca	3362	88	412
www.fm1047.ca	3317	41	443
www.larevue.qc.ca	3254	85	27
www.hebdorivenord.com	3248	68	302
www.lecharlevoisien.com	3246	78	80
www.journaldechambly.com	3150	103	274
www.nordinfo.com	3001	87	48

www.levail.com	2938	93	36
www.letincelle.qc.ca	2872	79	169
www.journalleguide.com	2843	95	327
www.leprogres.net	2830	70	154
www.nouvelleshebdo.com	2780	80	191
www.letoledulac.com	2776	84	197
www.laveniretdesrivieres.com	2741	88	328
www.lelacstjean.com	2678	70	213
www.coupdoeil.info	2672	85	306
www.lactiondautray.com	2493	62	263
www.canadafrancais.com	2396	71	114
www.radiogaspesie.ca	2334	68	348
www.lareleve.qc.ca	2330	125	54
linformationdunordmonttremlant.ca	2314	65	134
www.courrierdeportneuf.com	2309	54	186
ckmn.fm	2173	58	115
journaldesvoisins.com	2160	167	243
beauce.tv	2155	11	10
www.quebechebdo.com	2139	83	66
www.rcinet.ca	2127	368	131
www.lapetitenation.com	2040	47	144
www.estrieplus.com	2022	75	0
www.journalsaint-francois.ca	2022	49	0
www.oelregional.com	1992	60	125
www.beaucemedia.ca	1936	47	215
linformationdunordvalleedelarouge.ca	1768	56	103
www.chamblyexpress.ca	1475	55	311
www.radio-acton.com	1460	10	16
www.lejournaldejoliette.ca	1426	52	55
lautjournal.info	1422	224	184
www.tvrs.ca	1415	32	185
www.lechodelaval.ca	1336	95	52
www.lechodelarivenord.ca	1112	24	55
www.carrefourdequebec.com	1103	66	11
www.lechodetroisrivieres.ca	1091	60	62
echodefrontenac.com	1074	22	33
www.lepeuplelotbiniere.ca	967	44	130
radiox.com	848	8	56
www.sorel-tracyexpress.ca	793	24	59
www.lecantonnier.com	747	19	68
www.1031fm.ca	727	3	5
www.lattisee.com	695	26	27
mediat.ca	685	7	111
www.journallereffet.com	684	30	103
www.lavantagegaspesien.com	673	9	11
www.journalhorizon.com	653	21	212
www.leradar.qc.ca	595	10	25
autourdelile.com	580	15	64
lametropole.com	571	53	26
www.indicebohemien.org	524	33	13

monlimoilou.com	501	23	7
www.lecourrier.qc.ca	481	2	27
www.lavantposte.ca	473	12	5
www.journal-le-sentier.ca	445	26	30
radio-centreville.com	436	30	95
tremblantexpress.com	425	18	9
www.journal-ensemble.org	419	14	24
journaloutremont.com	418	23	27
lapensee.qc.ca	401	9	72
journalmobiles.com	397	46	9
www.lasentinelles.ca	328	5	5
laviedici.com	317	5	17
www.lelavallois.com	310	14	9
www.ckrl.qc.ca	305	3	3
monsaintroch.com	279	14	11
csur.tv	275	1	16
929cfut.com	274	4	17
echodenhaut.org	274	8	4
www.lebulletindeschenaux.com	266	7	5
www.entree libre.info	255	20	17
www.droitdeparole.org	235	16	24
www.lexpressmontcalm.com	230	1	46
www.moutonnoir.com	223	32	8
www.leclairon.qc.ca	222	6	49
graffici.ca	205	15	8
monsaintsauveur.com	202	8	7
www.tvr9.org	197	3	25
ski-se-dit.info	184	10	10
tvc.qc.ca	181	1	27
echosmontreal.com	146	11	4
www.lecitoyenrouynlasarre.com	135	5	59
cfmf.rocks	117	0	48
cfm.ca	102	9	14
www.versants.com	89	0	39
ctbtv.ca	57	3	35
www.via905.fm	46	0	8
tcfrontenac.org	33	0	1
www.gaboteur.ca	10	3	4

```
n_media_OCI <- dt_full_text$root_url %>% unique() %>% length()

rm(dt_full_text, count)
```

Le nombre total de sources médiatiques francophones généralistes dans ce corpus est 139. L'OCI a été construite pour couvrir l'ensemble des médias canadiens majeurs (ici, francophones) et s'étendre jusqu'à des médias régionaux. Bien entendu, il est difficile de couvrir de façon exhaustive les plus petits médias. On voit toutefois dans la distribution des articles du tableau ci-dessus que le poids d'un média local (voir à la fin du tableau) est infime par rapport au nombre de publications des médias majeurs comme *La Presse* et le *Journal de Montréal*. L'impact des médias locaux sur nos résultats est donc très faible.

Nous avons donné ci-dessus le nombre d'articles liés à la science selon notre procédure de détection. Plus bas,

Table 3: Les publications du corpus Facebook divisées selon nos deux variables dichotomiques

Statut	Nombre
Sans science avant COVID	102166
Science avant COVID	1046
Sans science pendant COVID	60100
Science pendant COVID	562

nous allons parfois nous concentrer sur une unité plus petite : le paragraphe lié à la science. Une question peut donc se poser : quelle proportion de l'espace d'un article lié à la science est-elle composée de paragraphes qui ne seraient pas détectés comme étant en lien avec la science selon notre méthode?

```
# Chargement des données
load(paste0(data_folder, "dt_full_text.RData")) # textes complets
load(paste0(data_folder, "dt_science.RData")) # paragraphes liés à la science

# extraction de tous les paragraphes dans les articles
# détectés comme étant liés à la science:
full_sci_text <- dt_full_text[art_ID %in% unique(dt_science$art_ID),full_article_text]
full_sci_text <- full_sci_text %>% str_split("\n\n") %>% unlist()
full_sci_text <- full_sci_text[nchar(full_sci_text)> 0] # retranche paragraphes vides

nb_para_in_art <- length(full_sci_text) # nombre de paragraphes dans tous ces articles
nb_char_in_art <- full_sci_text %>% nchar() %>% sum() # nombre caractères dans articles

# Même chose pour les paragraphes liés à la science:
nb_para_w_sci <- dt_science %>% nrow()
nb_char_w_sci <- dt_science$paragraphe %>% as.character() %>% nchar() %>% sum()

rm(dt_full_text, dt_science, full_sci_text)
```

Il y a au moins deux façons de calculer cette proportion. Premièrement, la proportion des paragraphes contenant notre expression régulière `(\s)scien|chercheu` parmi les articles considérés comme liés à la science est de 8%. Deuxièmement, la proportion des caractères de ces paragraphes sur le nombre total de caractères dans les articles étant liés à la science est de 12%. Bien entendu, il est fort probable qu'il soit question de science dans beaucoup d'autres paragraphes de ces articles sans que l'expression régulière utilisée soit présente.

Corpus Facebook

Notre deuxième corpus est tiré de Facebook grâce à l'outil CrowdTangle. Voici quelques données générales sur ce corpus.

```
fb_posts <- fread(file = paste0(accessible_data_folder, "FB_data.csv"))

caption_txt <- "Les publications du corpus Facebook divisées
selon nos deux variables dichotomiques"

kable(fb_posts[,.N, by = Status][,list(Statut= Status, Nombre = N)],
      caption = caption_txt) %>% kable_styling("hover")
```

```
caption_txt <- "Les publications du corpus Facebook par média
(ordre inversé du nombre total de publications)"

if(outputFormat == "html"){
  fb_posts[,list(`Nombre total` = .N, `Lié à la science` = sum(is_science),
                `Pendant COVID` = sum(since_COVID)),by = media][
    order(-`Nombre total`)] %>%
  kable(caption = caption_txt) %>% kable_styling() %>%
  scroll_box(width = "100%", height = "200px")
} else{
  fb_posts[,list(`Nombre total` = .N, `Lié à la science` = sum(is_science),
                `Pendant COVID` = sum(since_COVID)),by = media][
    order(-`Nombre total`)] %>%
  kable(caption = caption_txt, booktabs = TRUE, longtable = TRUE) %>%
  kable_styling()
}
```

Table 4: Les publications du corpus Facebook par média (ordre inversé du nombre total de publications)

media	Nombre total	Lié à la science	Pendant COVID
Le Journal de Québec	13346	80	3533
Le Journal de Montréal	10885	101	3334
La Presse	8877	176	2726
TVA Nouvelles	8727	72	2819
HuffPost Québec	8690	88	1653
Le Soleil	8027	108	2970
Le Devoir	6304	174	2084
Radio-Canada Information	5604	130	1937
Le Nouvelliste	5492	55	2248
Le Quotidien	4646	66	2008
La Tribune	4564	63	2319
Journal Métro	3993	64	1020
La Voix de l'Est	3123	37	1597
ICI Québec	2586	27	1033
www.lanouvelle.net	2289	2	880
ICI Mauricie Centre-du-Québec	1985	18	804
TVA Gatineau-Ottawa	1842	4	870
98,5 FM	1764	9	599
ICI Ottawa Gatineau	1750	6	527
ICI Saguenay-Lac-Saint-Jean	1546	4	672
ICI Estrie	1542	9	663
ICI Grand Montréal	1458	27	465
FM Charlevoix - CIHO 96,3	1267	1	655
Le Charlevoisien	1267	1	796
L'actualité	1232	63	379
Le Courrier du Sud	1215	1	502
ICI Côte-Nord	1195	17	457
104,7 Outaouais	1179	5	812
Le Citoyen Rouyn - La Sarre	1168	2	436
ICI Abitibi-Témiscamingue	1145	7	562

L'Avantage	1089	5	218
L'Hebdo du St-Maurice	1075	4	315
ICI Bas-Saint-Laurent	1075	22	493
EnBeauce.com	1059	5	468
Les 2 Rives	1043	2	322
Journal Mobiles	1034	7	464
Le Courrier Sud	1032	4	249
ICI Gaspésie-Les-Îles	1016	6	445
Le Nord-Côtier	999	2	485
Journal L'Express de Drummondville	971	2	562
CFIM, la radio des Îles de la Madeleine	954	16	484
Info Dimanche	948	3	493
TVA Québec	890	3	437
Journal Le Reffet	884	0	609
FM 103,3 - La radio allumée	884	33	329
L'Éclaireur Progrès	824	3	245
Journal Le Manic	798	2	446
TVA Nouvelles Abitibi-Témiscamingue	780	1	335
Le Soleil de Châteauguay	778	0	572
TVA Est-du-Québec	712	2	298
TVA Saguenay-Lac-St-Jean	705	2	382
FM93 Québec	672	1	269
MédiAT	662	0	314
Écho de La Tuque	613	0	268
Journal Saint-François	588	1	437
L'Écho de Maskinongé	568	0	219
TVA Sherbrooke	568	2	309
Journal L'Action	557	0	272
Journal Le Lac-St-Jean	542	0	229
Courrier Laval	540	3	336
L'Hebdo Journal	530	2	200
Courrier Frontenac	528	1	175
L'info du Nord Mont-Tremblant	517	0	226
L'info de la Basse-Lièvre	516	2	193
Journal de Chambly	498	0	148
Le Reffet du Lac	487	0	162
CHGA FM, la radio de la Vallée-de-la-Gatineau	480	0	270
L'Avantage Gaspésien	461	6	123
Beauce Média	458	0	122
Hebdomadaire La Revue de Terrebonne	434	0	236
Journal Nouvelles Hebdo	427	2	174
Journal La Voix du Sud	420	0	188
Le Placoteux	417	3	221
Radio X - CHOI 98,1	407	0	256
Hebdo Rive Nord	405	2	195
Journal Coup d'oeil	391	2	157
BeauceTV	389	0	185
Radio Gaspésie	389	3	264
GranbyExpress.com	380	0	119
L'info Petite Nation	378	0	119

Québec Hebdo	365	1	117
Journal L'Étoile du Lac	363	0	148
VIA 90.5 FM	358	2	229
Journal Le Canada Français	350	1	131
EstriePlus.com Le Journal Internet	349	4	103
Journal La Pensée de Bagot	338	0	117
TVR9	309	1	79
L'OEil Régional	304	0	121
Radio Acton	297	0	123
Monlimoilou	291	2	86
Journal Le Nord	275	1	144
106.9 FM Mauricie	273	2	169
Le Courrier de Saint-Hyacinthe	269	0	110
Le Journal de Joliette	266	1	153
L'Avant-Poste	266	1	94
Country Pop 103,1	261	0	83
La Sentinelle	251	0	37
Journal L'Express Montcalm	251	0	104
L'Écho de Laval	249	0	86
Monsaintroch	248	1	65
L'info du Nord Vallée de la Rouge	243	0	124
CKRL 89,1	242	1	39
Courrier de Portneuf	240	0	124
Monsaintsauveur	215	0	57
Le Progrès de Coaticook	209	1	72
JournalLeGuide.com	189	0	38
L'Indice bohémien	186	4	81
L'Écho de Frontenac	185	0	31
Le Peuple Lotbinière	163	0	49
Actualités - L'Étincelle du Val SF	157	0	56
lAvenirEtDesRivieres.com	157	1	61
Journal Ski-se-Dit	157	1	31
L'aut'journal	156	1	59
Journal L'Action D'Autray	154	0	44
Bulletin des Chenaux	152	1	55
Radio Canada International	148	4	106
Le Radar, l'hebdomadaire des Îles	146	0	33
Journaldesvoisins.com	138	1	53
Journal L'Éveil	136	1	44
92,9 CFUT	131	0	26
CKMN.FM	125	0	40
Journal Autour de l'Île	123	0	94
Journal l'Écho d'en Haut Inc.	98	0	13
TVC Basse-Lièvre	98	0	45
Journal L'Horizon, presse coopérative des Basques	97	1	33
Csur la télé	93	0	13
Journal Le Carrefour de Québec	92	0	29
Le Nord Info	85	0	38
L'Écho de Trois-Rivières	84	0	35
Tremblant Express	75	1	36

Journal Le Sentier	71	0	15
JOURNAL GRAFFICI	67	1	14
Journal Le Mouton Noir	64	0	25
L'Écho de la Rive-Nord	60	0	23
Sorel-Tracy Express.ca	59	0	19
Journal La Relève	56	0	38
CTBTv	55	0	18
Journal Entrée Libre	50	0	27
CFMF 103,1 FM	50	0	24
Le Gaboteur	47	0	13
La Métropole	47	0	6
Journal Le Lavallois	46	0	13
CHIP 101.9	45	0	45
Chambly Express.ca	42	0	10
Journal Communautaire Le Cantonnier	41	0	22
Versants du Mont-Bruno	35	0	27
Journal Droit de Parole	34	0	8
Journal Outremont	34	1	17
Journal communautaire Ensemble pour bâtir	24	0	12
Television Rive-Sud-TVRS	22	0	17
La Vie d'Ici	19	0	3
TCFTV Montréal	11	0	9
Radio Centre-ville	8	0	2

```
n_media_FB <- fb_posts$media %>% unique() %>% length()

rm(fb_posts)
```

Il y a 153 pages Facebook de médias dans ce corpus.

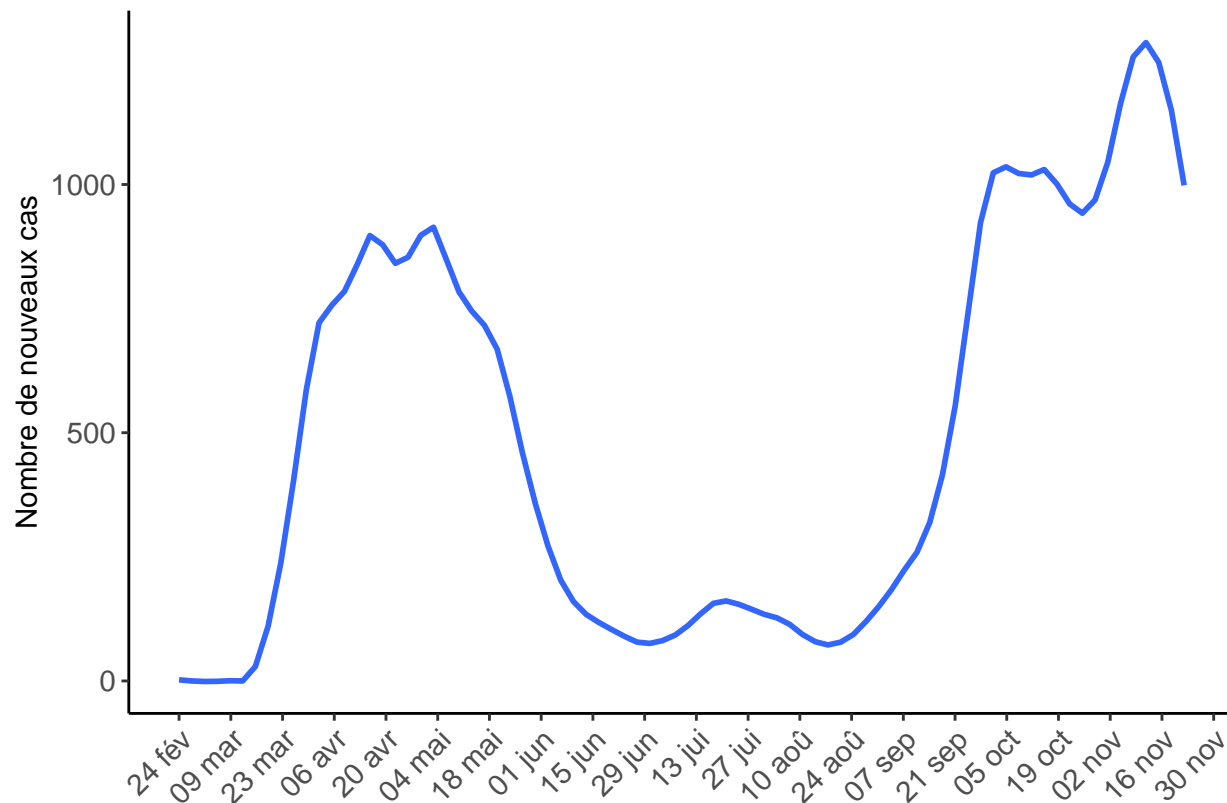
Information sur la COVID-19 en général

Progression de la pandémie au Québec

Avec les données de l'*Institut national de santé publique du Québec* (ici), nous pouvons produire un graphique du nombre de cas confirmés de COVID-19 au Québec en 2020. Voir `figure1.jpg`.

```
dt_covid_cases <- fread(paste0(accessible_data_folder, "nb_cas_COVID_qc.csv"))
dt_covid_cases <- dt_covid_cases[, list(date = as.Date(`Date de déclaration du cas`),
  nb_cases = `Par lien épidémiologique` + `En laboratoire`)]

ggplot(dt_covid_cases, aes(x= date, y = nb_cases)) + geom_smooth(span = .1, se= FALSE) +
  theme_classic() + labs(x= "", y = "Nombre de nouveaux cas") +
  scale_x_date(date_breaks = "2 week", date_labels = "%d %b") +
  theme(axis.text.x = element_text(angle = 45,
    vjust = 1, hjust=1, size = size_axis_text),
    axis.text.y = element_text(size = size_axis_text))
```



```
ggsave(file=paste0(fig_dir,"figure1.jpg"), height=4)
```

Mentions de la COVID dans notre 1er corpus

Quelle proportion des textes dans le corpus de l'OCI mentionne la COVID entre janvier et juin 2020? Voir figure2.jpg.

```
dt_full_text <- fread(paste0(accessible_data_folder, "dt_full_text_for_count_sci.csv"))

dt_articles_thru_time <- dt_full_text[,total := .N, by = date] %>%
  .[is_covid == TRUE, prop_covid := .N / total, by = date] %>%
  .[!is.na(prop_covid),list(date, prop_covid)] %>%
  unique()

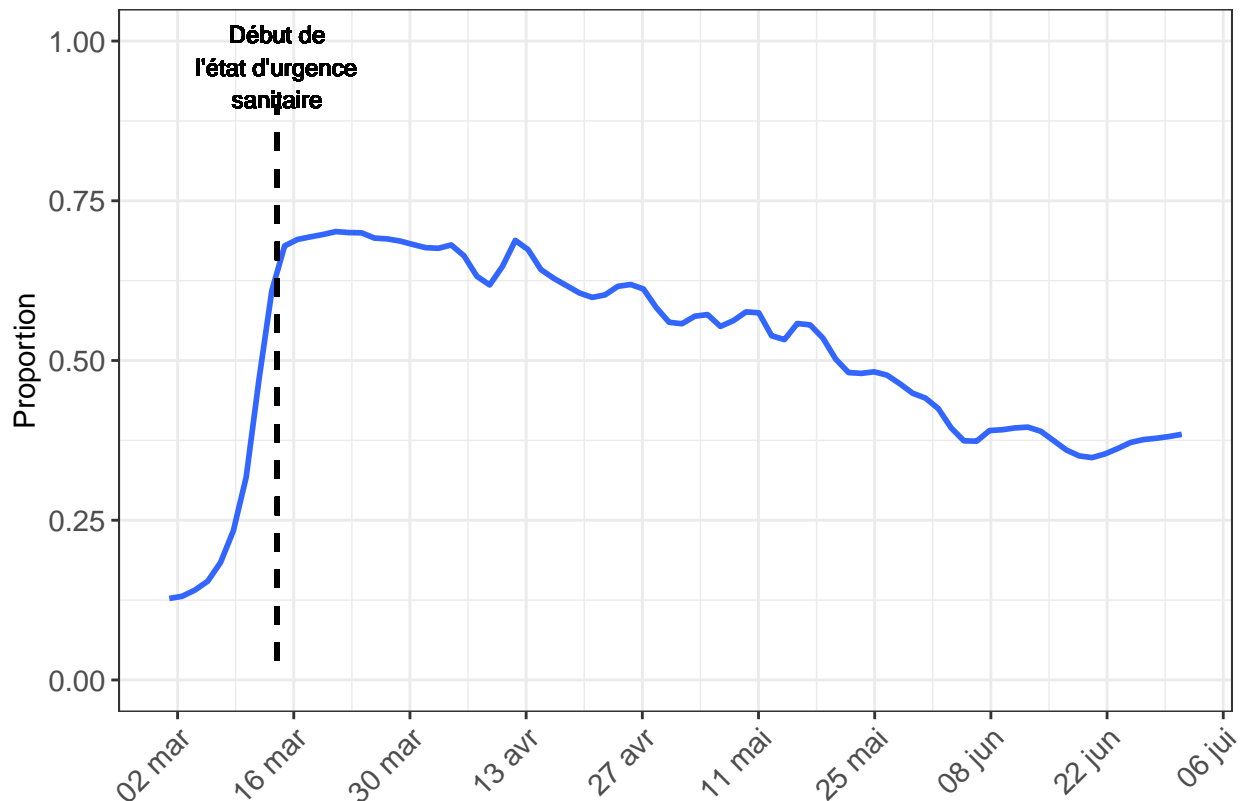
rm(dt_full_text)

ggplot(dt_articles_thru_time[date >= "2020-03-01" ], aes(x=date, y=prop_covid)) +
  geom_smooth(se=F,method = 'loess', show.legend = FALSE, span = 0.1) +
  theme_bw() + ylab("Proportion") + scale_x_date(date_breaks = "2 week",
                                                  date_labels = "%d %b") +
  theme(axis.text.x = element_text(angle = 45,
                                    vjust = 1, hjust=1, size = size_axis_text),
        axis.text.y = element_text(size = size_axis_text)) + xlab("") +
  ylim(0,1) + geom_segment(aes(x = pos_vline,xend=pos_vline,y=.03,yend=.9),
```

```

    linetype="dashed", colour = "black",
    size=.8, alpha = .6) +
geom_text(aes(x=pos_vline, y= Inf,
    label = "Début de\nl'état d'urgence\nsanitaire"),
    colour = "black", vjust = 1.2, size = 3)

```



```

ggsave(file=paste0(fig_dir,"figure2.jpg"),height= 5, width = 8)

```

Plus de science?

Part du contenu médiatique

Notre première grande question concerne la part des textes médiatiques liés à la science. En premier lieu, nous utilisons une moyenne mobile sur une semaine pour suivre cette proportion (voir [figure_A1.jpg](#)). Comme il y a beaucoup de volatilité, nous utilisons ensuite un lissage par régression polynomiale locale (voir [figure3.jpg](#)).

```

dt_full_text <- fread(paste0(accessible_data_folder, "dt_full_text_for_count_sci.csv"))

#On compte d'abord le nombre d'articles total par jour,
#et le nombre d'articles qui mentionnent la science chaque jour
dt_plot <- dt_full_text[, list(total_art = .N, total_sci = sum(is_science)), by = date]

#Pour que les lignes soient superposées, on leur met tous la même année

```

```

dt_plot <- dt_plot[month(date) < 7]
dt_plot[, year := substr(date, 1, 4)]
dt_plot[year == 2017, date := as.Date(date) %m+% years(3)]
dt_plot[year == 2018, date := as.Date(date) %m+% years(2)]
dt_plot[year == 2019, date := as.Date(date) %m+% years(1)]

dt_plot[year %in% 2017:2019, corpus := 1][is.na(corpus), corpus := 2]

dt_plot[, proportion := sum(total_sci)/sum(total_art), by = .(year, date)]
#On utilise la moyenne sur 7 jours (3 jours avant et 3 jours après)
#Pour les 3 premiers jours de la période et les 3 derniers,
# on prend respectivement les 6 jours suivants et les 6 jours précédents.
window_size = 7
dt_plot[, `:=`(week_mean = rollmean(proportion,
                                     window_size, align = "center", na.pad = TRUE),
               first_nas = rollmean(proportion, window_size, align = "left", na.pad = TRUE),
               last_nas = rollmean(proportion, window_size, align = "right", na.pad = TRUE)),
         by = year]
dt_plot[is.na(week_mean) & is.na(first_nas), week_mean := last_nas]
dt_plot[is.na(week_mean) & is.na(last_nas), week_mean := first_nas]

dt_plot[corpus == 1, corpus_years := "2017-2019"]
dt_plot[corpus == 2, corpus_years := "2020"]
dt_plot[, corpus_years := as.factor(corpus_years)]

title_g <- "Proportion d'articles à caractère scientifique\n
dans les médias franco-canadiens"

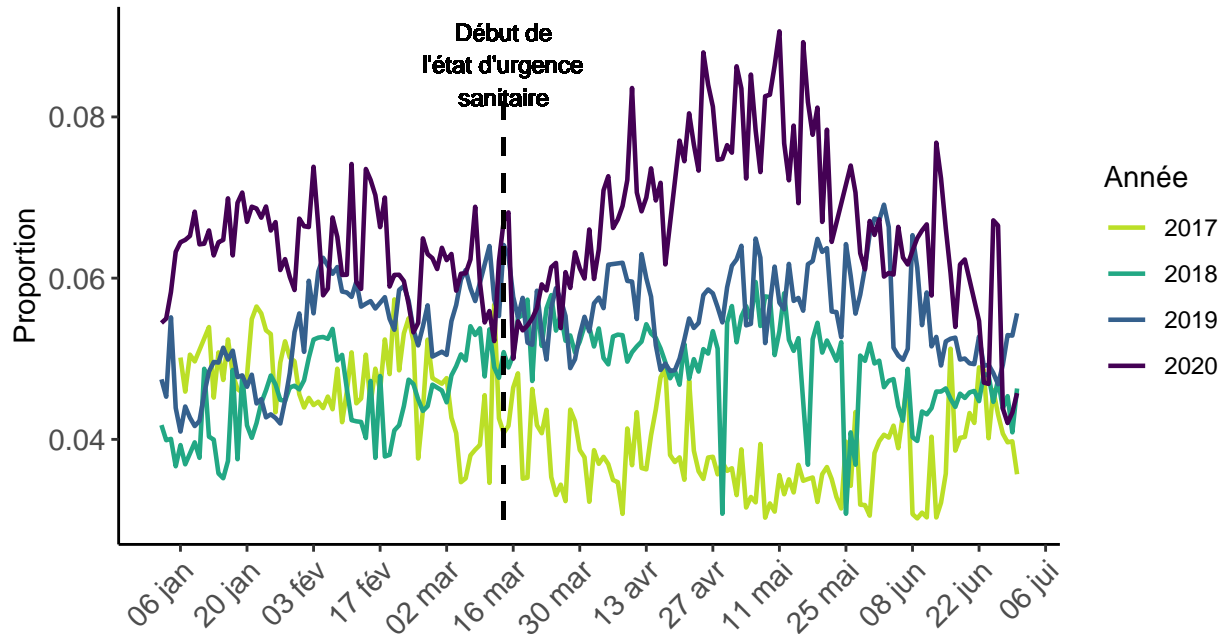
ggplot(dt_plot[date > as.Date("2020-01-01")], aes(x=date, y=week_mean, col = year)) +
  geom_line(size = .8) +
  theme_classic() +
  scale_x_date(date_breaks = "2 week", date_labels = "%d %b") +
  ggtitle(title_g) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1, size = size_axis_text),
        axis.text.y = element_text(size = size_axis_text)) +
  ylab("Proportion") +
  labs(color = "Année", subtitle =
        "(Les lignes représentent des moyennes mobiles sur 7 jours)") + xlab("") +
  scale_color_viridis(discrete=TRUE, end = .9, direction = -1) +
  geom_segment(aes(x = pos_vline, xend=pos_vline, y=.03, yend=.083),
               linetype="dashed", colour = "black", size=.8, alpha = .6) +
  geom_text(aes(x=pos_vline, y= Inf,
                label = "Début de\nl'état d'urgence\nsanitaire"),
            colour = "black", vjust = 1.2, size = 3)

```


Proportion d...articles à caractère scientifique

dans les médias franco-canadiens

(Les lignes représentent des moyennes mobiles sur 7 jours)

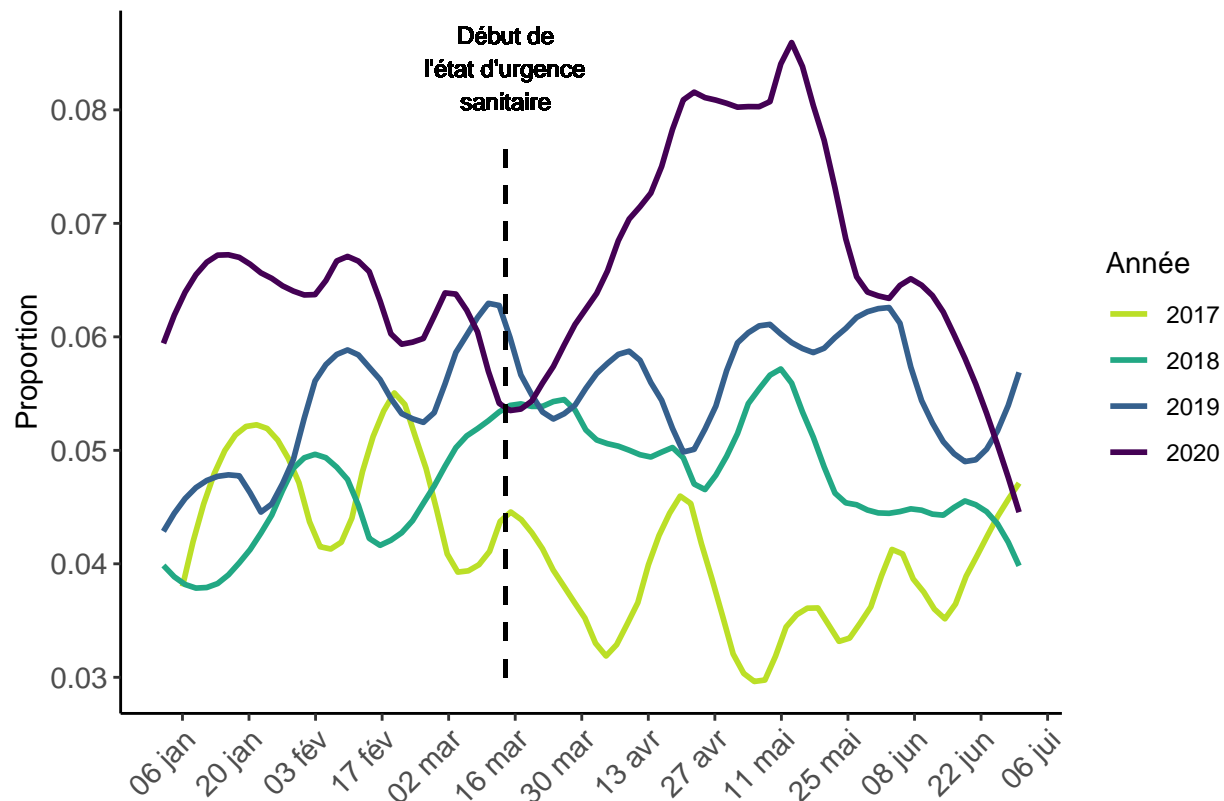


```
ggsave(file=paste0(fig_dir,"figure_A1.jpg"), width = 10)
```

```
p_more_sci <- ggplot(dt_plot[date > as.Date("2020-01-01")],
  aes(x=date, y=proportion, col = year)) +
  geom_smooth(span = .2, se = FALSE) +
  theme_classic() +
  scale_x_date(date_breaks = "2 week", date_labels = "%d %b") +
  # ggtitle(title_g) + #ylim(0, NA) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1, size = size_axis_text),
    axis.text.y = element_text(size = size_axis_text)) +
  ylab("Proportion") +
  labs(color = "Année#", subtitle = "(Curves smoothed by local polynomial regression)"
) + xlab("") + scale_color_viridis(discrete=TRUE, end = .9, direction = -1) +
  geom_segment(aes(x = pos_vline,xend=pos_vline,y=.03,yend=.078),
    linetype="dashed", colour = "black", size=.8, alpha = .6) +
  geom_text(aes(x=pos_vline, y= Inf,
    label = "Début de\nl'état d'urgence\nsanitaire"),
    colour = "black", vjust = 1.2, size = 3)

ggsave(file=paste0(fig_dir,"figure3.jpg"), plot = p_more_sci, width = 7, height =4)

p_more_sci
```



```
prop_sci_first_half_y <- dt_plot[, list(annual_proportion = sum(total_sci)/sum(total_art),
                                     total_art = sum(total_art),
                                     total_sci = sum(total_sci)), by = year]

prop_sci_pandemic <- dt_plot[date >= as.Date("2020-03-14"),
                             list(annual_proportion = sum(total_sci)/sum(total_art),
                                  total_art = sum(total_art),
                                  total_sci = sum(total_sci)), by = year]

prop_sci_before_pandemic <- dt_plot[date < as.Date("2020-03-14"),
                                    list(annual_proportion = sum(total_sci)/sum(total_art),
                                         total_art = sum(total_art),
                                         total_sci = sum(total_sci)), by = year]

# Préparation des données pour utilisation ci-dessous.
n_dt_covid_cases <- "Nouveaux cas de COVID"
n_dt_science_prop <- "Proportion d'articles\nliés à la science"

dt_p <- rbindlist(list(dt_covid_cases[,list(date, y = nb_cases,
                                             type_dt = n_dt_covid_cases)],
                     data.table(date = as.Date(as.Date("2020-01-01")):(min(dt_covid_cases$date)-1)),
                     y = 0, type_dt = n_dt_covid_cases),
                  dt_plot[year == 2020, list(date = as.Date(date), y = proportion,
                                             type_dt = n_dt_science_prop)]
),fill = TRUE)
```

Table 5: Proportion d'articles liés à la science dans les premiers 6 mois de chaque année

Année	Proportion	Taux de croissance
2017	0.041	NA
2018	0.048	0.179
2019	0.055	0.137
2020	0.065	0.200

Table 6: Proportion d'articles liés à la science du 14 mars (lendemain de l'annonce de l'état d'urgence sanitaire au Québec en 2020) jusqu'à la fin juin

Année	Proportion	Taux de croissance
2017	0.0377655	NA
2018	0.0497979	0.319
2019	0.0555046	0.115
2020	0.0673171	0.213

```
rm(dt_plot, dt_full_text)
```

On voit que la proportion de textes liés à la science suit une tendance ascendante sur les quatre ans. Nous calculons les moyennes sur les 6 mois, pour le premier tableau, ou pendant les dates de la période d'urgence sanitaire (du 14 mars au 30 juin) dans le second :

```
nb_round = 3
prop_sci_first_half_y[, `Growth_rate` :=
  round(c(NA,
    diff(annual_proportion)/annual_proportion[1:(.N-1)]),
    nb_round)]

prop_sci_first_half_y[,list(Année = year,
  `Proportion` = round(annual_proportion,nb_round),
  `Taux de croissance` = Growth_rate)] %>%
  kable(caption =
    "Proportion d'articles liés à la science
    dans les premiers 6 mois de chaque année"
  ) %>% kable_styling("hover")

prop_sci_pandemic[, `Growth_rate` :=
  round(c(NA, diff(annual_proportion)/annual_proportion[1:(.N-1)])
    ,3)]

prop_sci_pandemic[,list(Année = year, `Proportion` = annual_proportion,
  `Taux de croissance` = Growth_rate)] %>%
  kable(caption = "Proportion d'articles liés à la science du 14 mars
    (lendemain de l'annonce de l'état d'urgence sanitaire au Québec en 2020)
    jusqu'à la fin juin") %>% kable_styling("hover")
```

Cette tendance à la hausse est surprenante. Il faut s'assurer que ce ne soit pas un artefact de changements dans notre base de données. Comme test de robustesse, nous calculons:

```
dt_full_text <- fread(paste0(accessible_data_folder, "dt_full_text_for_count_sci.csv"))

dt_full_text[root_url %in% c("www.lapresse.ca", "www.journaldemontreal.com"),
```

Table 7: 1. La proportion annuelle dans les médias principaux

Année	Proportion
2017	0.0518921
2018	0.0617191
2019	0.0641443
2020	0.0759209

Table 8: 2. La proportion annuelle d'articles avec plus de 1000 caractères (puisque la moyenne du nombre de caractères suit aussi une pente ascendante)

Année	Proportion
2017	0.0557242
2018	0.0592089
2019	0.0651659
2020	0.0789753

```

      "ici.radio-canada.ca", "www.ledevoir.com",
      "www.tvanouvelles.ca"),
    list(Proportion = sum(is_science)/.N ), by = year(date)) [
, list(Année = year, Proportion)] %>%
kable(caption = "1. La proportion annuelle dans les médias principaux") %>%
  kable_styling("hover")

dt_full_text[n_char_text > 1000, list(Proportion = sum(is_science)/.N ), by = year(date)] [
, list(Année = year, Proportion)] %>%
kable(caption = "2. La proportion annuelle d'articles avec plus de 1000 caractères
(puisque la moyenne du nombre de caractères suit aussi une pente ascendante)") %>%
  kable_styling("hover")

dt_full_text[n_char_text > 1000 &
root_url %in% c("www.lapresse.ca", "www.journaldemontreal.com", "ici.radio-canada.ca",
"www.ledevoir.com", "www.tvanouvelles.ca"),
list(Proportion = sum(is_science)/.N ), by = year(date)] [
, list(Année = year, Proportion)] %>%
kable(caption = "3. La combinaison des deux restrictions précédentes") %>%
  kable_styling("hover")

rm(dt_full_text)

```

La tendance vers le haut est présente sous toutes ces restrictions. Elle ne semble donc pas être un artefact de nos données.

De retour à la distribution des proportions, nous vérifions que les différences sont statistiquement significatives.

Table 9: 3. La combinaison des deux restrictions précédentes

Année	Proportion
2017	0.0638760
2018	0.0723304
2019	0.0729239
2020	0.0868425

Nous définissons l'hypothèse alternative comme étant que la période prépandémique avait une proportion plus faible.

Premier scénario : 2020 contre 2019 pour l'ensemble des premiers 6 mois.

```
y_select = 2019:2020
prop.test(prop_sci_first_half_y[year %in% y_select, total_sci],
          prop_sci_first_half_y[year %in% y_select, total_art],
          alternative = "l")

##
## 2-sample test for equality of proportions with continuity correction
##
## data: prop_sci_first_half_y[year %in% y_select, total_sci] out of prop_sci_first_half_y[year %in% y_select, total_sci]
## X-squared = 250.75, df = 1, p-value < 2.2e-16
## alternative hypothesis: less
## 95 percent confidence interval:
## -1.000000000 -0.009793704
## sample estimates:
## prop 1 prop 2
## 0.05456841 0.06549103
```

L'hypothèse nulle est fortement rejetée.

2e scénario : du 15 mars à la fin juin, en comparant 2020 à 2019:

```
y_select = 2019:2020
prop.test(prop_sci_pandemic[year %in% y_select, total_sci],
          prop_sci_pandemic[year %in% y_select, total_art],
          alternative = "l")

##
## 2-sample test for equality of proportions with continuity correction
##
## data: prop_sci_pandemic[year %in% y_select, total_sci] out of prop_sci_pandemic[year %in% y_select, total_sci]
## X-squared = 178.85, df = 1, p-value < 2.2e-16
## alternative hypothesis: less
## 95 percent confidence interval:
## -1.000000000 -0.01036644
## sample estimates:
## prop 1 prop 2
## 0.05550462 0.06731709
```

L'hypothèse nulle est encore fortement rejetée.

3e scénario : en se concentrant sur 2020, avant et après le début de la pandémie au Québec:

```
y_select = 2020
dt_2020 <- rbind(prop_sci_before_pandemic[year == y_select,
                                           list(pandemic_in_QC = FALSE,
                                                total_sci, total_art)],
                 prop_sci_pandemic[year == y_select, list(pandemic_in_QC = TRUE,
                                                           total_sci, total_art)])

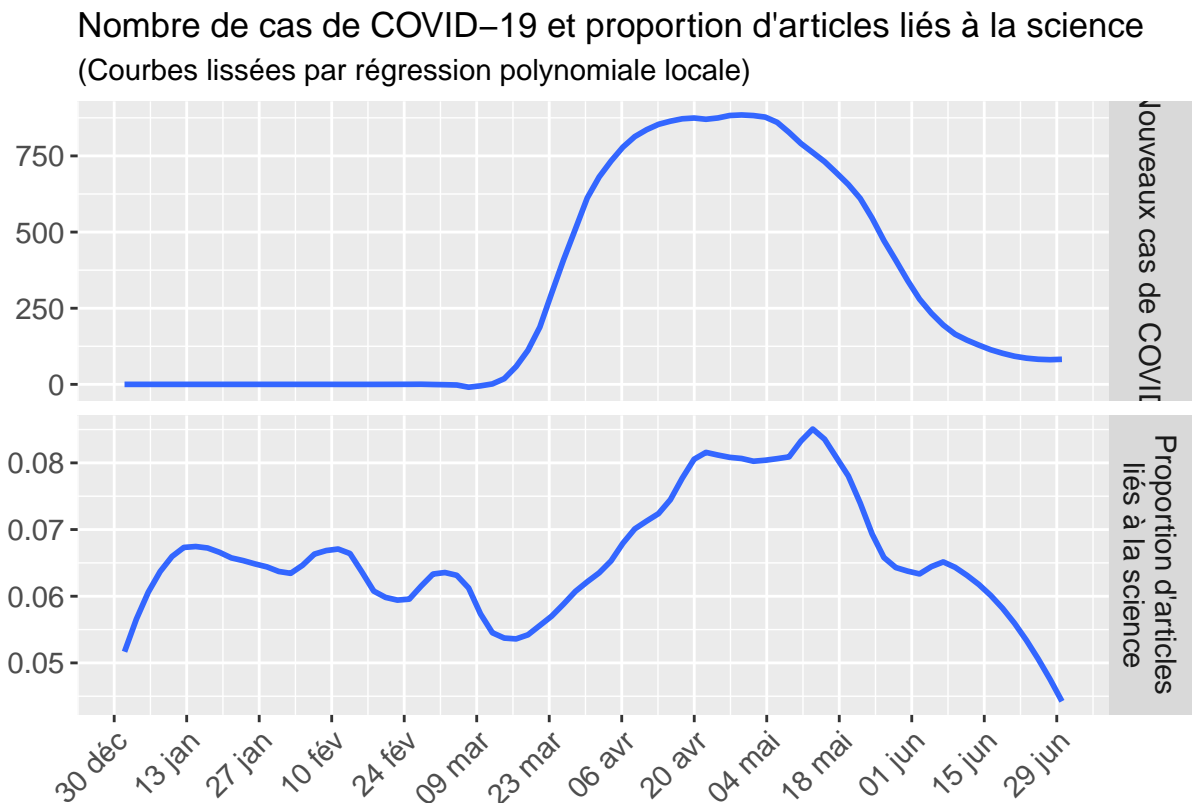
prop.test(dt_2020[, total_sci], dt_2020[, total_art], alternative = "l")

##
## 2-sample test for equality of proportions with continuity correction
##
```

```
## data: dt_2020[, total_sci] out of dt_2020[, total_art]
## X-squared = 23.736, df = 1, p-value = 5.524e-07
## alternative hypothesis: less
## 95 percent confidence interval:
## -1.000000000 -0.003218551
## sample estimates:
##      prop 1      prop 2
## 0.06247113 0.06731709
```

Ici, la différence est moins dramatique, mais l'hypothèse est tout de même fortement rejetée.

La similitude est frappante entre les tendances temporelles du nombre quotidien de cas de COVID et de la proportion d'articles liés à la science (voir [figure_A2.jpg](#)):



Même si c'est peu probable, cette synchronicité pourrait être une coïncidence. Vérification : Est-ce que les articles à propos de la COVID sont plus fréquemment liés à la science? (voir [figure_A3.jpg](#))

```
dt_full_text <- fread(paste0(accessible_data_folder,
                              "dt_full_text_for_count_sci.csv"))

dt_full_text[is_covid == TRUE, est_covid := "Vrai"][
  is.na(est_covid), est_covid := "Faux"]

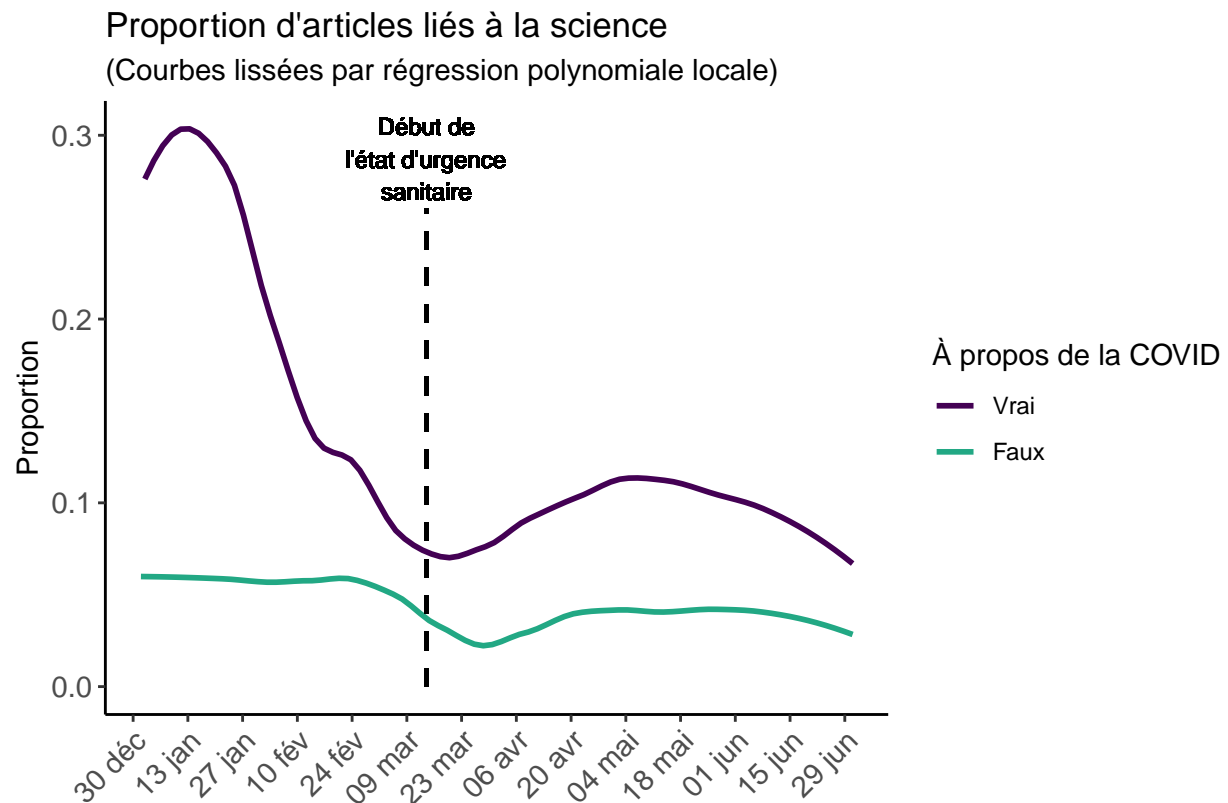
dt_full_text$is_covid <- factor(dt_full_text$is_covid ,
                              levels = c("TRUE", "FALSE"))

dt_full_text$est_covid <- factor(dt_full_text$est_covid ,
                              levels = c("Vrai", "Faux"))
```

```

ggplot(
  dt_full_text[year(date) == 2020, sum(is_science)/.N,
    by = .(date, est_covid)],
  aes(x=date, y = V1, color = est_covid)) +
  geom_smooth(span = .4, se = FALSE) +
  theme_classic() + ylim(0,NA) +
  scale_x_date(date_breaks = "2 week", date_labels = "%d %b") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1,
    hjust=1, size = size_axis_text),
    axis.text.y = element_text(size = size_axis_text)) +
  labs(y = "Proportion", x = "",
    title = "Proportion d'articles liés à la science",
    subtitle = "(Courbes lissées par régression polynomiale locale)",
    color = "À propos de la COVID") +
  geom_segment(aes(x = pos_vline, xend=pos_vline, y=0, yend=.26),
    linetype="dashed", colour = "black", size=.8, alpha = .6) +
  geom_text(aes(x=pos_vline, y= Inf,
    label = "Début de\nl'état d'urgence\nsanitaire"),
    colour = "black", vjust = 1.2, size = 3) +
  scale_color_viridis( discrete=TRUE, end = .6, direction = 1) # +

```



```

ggsave(file=paste0(fig_dir,"figure_A3.jpg"),width = 7)

```

```

#print("Proportion of science-related articles in the first half of 2020")

```

Table 10: Première moitié de 2020

À propos de la COVID	Proportion de science
Faux	0.047272
Vrai	0.095466

Table 11: Seulement janvier et février 2020

À propos de la COVID	Proportion de science
Faux	0.0576441
Vrai	0.1693911

```
dt_full_text[year(date) == 2020,
  list(proportion_of_science =
    sum(is_science)/.N), by = est_covid][
  ,list(`À propos de la COVID` = est_covid,
    `Proportion de science` =
      proportion_of_science)] %>%
kable(caption = "Première moitié de 2020") %>%
kable_styling("hover", full_width = F)

dt_full_text[between(date, as.Date("2020-01-01"), as.Date("2020-02-29")),
  list(proportion_of_science = sum(is_science)/.N), by = est_covid][
  ,list(`À propos de la COVID` = est_covid,
    `Proportion de science` = proportion_of_science)] %>%
kable(caption = "Seulement janvier et février 2020") %>%
kable_styling("hover", full_width = F)

#print("Proportion of science-related articles in March to June of 2020")
dt_full_text[date > as.Date("2020-02-29"),
  list(proportion_of_science = sum(is_science)/.N), by = est_covid][
  ,list(`À propos de la COVID` = est_covid,
    `Proportion de science` = proportion_of_science)] %>%
kable(caption = "Mars à juin 2020") %>%
kable_styling("hover", full_width = F)

rm(dt_full_text)
```

La réponse est clairement affirmative : le lien à la science est plus fort pour les articles qui traitent de la COVID. Ce lien est particulièrement fort au tout début de 2020, mais il demeure significatif tout au long de la première vague.

Table 12: Mars à juin 2020

À propos de la COVID	Proportion de science
Faux	0.0383117
Vrai	0.0922036

Intensité de l'intérêt du public pour la science

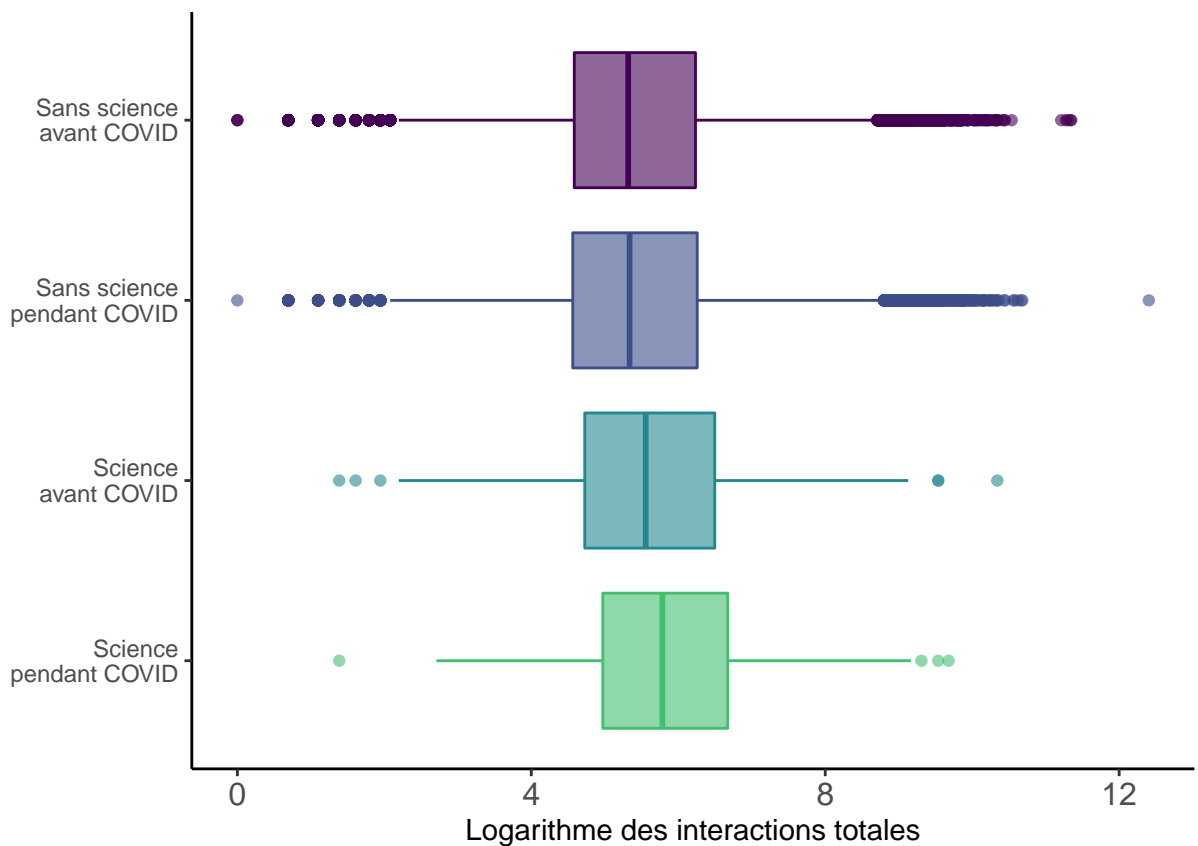
Pour répondre à notre deuxième question dans ce volet sur la quantité de science dans les médias – la science a-t-elle davantage intéressé le public des médias – nous nous tournons vers le corpus Facebook et utilisons les interactions (ou *engagement* en anglais) comme mesure. À noter que nous prenons le logarithme des interactions puisque celui-ci suit une loi de puissance.

Nous produisons d'abord une représentation par boîtes à moustaches de la distribution des interactions pour nos quatre catégories de publications (voir [figure4.jpg](#)).

```
fb_posts <- fread(paste0(accessible_data_folder, "FB_data.csv"))

fb_posts$Status <- factor(fb_posts$Status ,
                          levels = fb_posts[order(is_science, since_COVID),
                                                Status %>% unique() %>% rev()] )

### Représentation avec diagramme de boîtes à moustaches:
ggplot(fb_posts[semaine >= 10], aes(y= Status, x=log(interaction),
                                     fill= Status, col= Status )) +
  geom_boxplot(alpha= .6) +
  theme_classic() + guides(fill="none", col="none") +
  labs(x = "Logarithme des interactions totales", y="") +
  theme(axis.text.x = element_text(size = 12)) +
  scale_fill_viridis(discrete=TRUE, end = .7, direction = -1) +
  scale_color_viridis(discrete=TRUE, end = .7, direction = -1)
```



```

ggsave(file=paste0(fig_dir,"figure4.jpg"),width = 7,height = 4 )

## Préparation des données pour les tests statistiques
dt_to_fit <- fb_posts[semaine >= 10,list(log_interaction = log(interaction),
  is_science, since_COVID, semaine, Status2)]

rm(fb_posts)

```

Nous testons ensuite avec un modèle ANOVA pour l'égalité des moyennes des distributions:

```

## test ANOVA sur les quatre sous populations
model_w_anova <- aov(log_interaction ~ Status2, dt_to_fit)
summary(model_w_anova)

```

```

##              Df Sum Sq Mean Sq F value    Pr(>F)
## Status2        3     104    34.80   21.38 7.75e-14 ***
## Residuals    112781 183561     1.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

L'hypothèse nulle de l'égalité des distributions est fortement rejetée. Nous proposons ensuite un modèle de régression avec un terme d'interaction pour séparer les différentes composantes. Le modèle est:

$$\log(E) = \alpha + \beta_1 S + \beta_2 C + \delta SC + \varepsilon \quad \varepsilon \sim NID(\theta, \sigma^2)$$

Où E est la somme des interactions (l'engagement selon Facebook), S est une variable binaire pour la référence à la science ($S = 1$ si oui; $S = 0$ si non) et C est une variable binaire pour la période de la pandémie ($C = 1$ si mars à juin 2020; $C = 0$ si mars à juin, de 2017 à 2019). L'estimation du modèle par la méthode des moindres carrés produit ces résultats :

```

## Modèle de régression avec un terme d'interaction entre nos deux variables binaires:
model_w_interaction <- lm(log_interaction ~ is_science*since_COVID, data = dt_to_fit)

```

```

## Estimation du modèle avec des erreurs standardisées robustes
stargazer(model_w_interaction,
  se=list(sqrt(diag(vcovHC(model_w_interaction, type = "HC1")))),
  title="Regression des interactions sur Facebook", type="text",
  #align = TRUE,#type = "latex", float = TRUE, #
  df=FALSE, digits=3)

```

```

##
## Regression des interactions sur Facebook
## =====
##                               Dependent variable:
##                               -----
##                               log_interaction
## -----
## is_science                   0.207***
##                               (0.050)
##
## since_COVID                   0.021***
##                               (0.008)
##
## is_scienceTRUE:since_COVID   0.173**
##                               (0.079)

```

```
##
## Constant                5.462***
##                        (0.005)
## -----
## Observations            112,785
## R2                      0.001
## Adjusted R2             0.001
## Residual Std. Error     1.276
## F Statistic              21.378***
## =====
## Note:                    *p<0.1; **p<0.05; ***p<0.01
```

L'ensemble des paramètres sont significativement différents de zéro. Le paramètre principal pour nous est δ puisqu'il extrait la hausse spécifique liée à la science pendant la pandémie.

Comme notre variable dépendante (les interactions) est en logarithme, les paramètres estimés donnent des relations entre moyennes géométriques (et non arithmétiques):

1. Sans science avant COVID: $e^\alpha = 236$
2. Sans science pendant COVID: $e^{\alpha+\beta_2} = 241$
3. Science avant COVID: $e^{\alpha+\beta_1} = 290$
4. Science pendant COVID: $e^{\alpha+\beta_1+\beta_2+\delta} = 352$

Ces valeurs correspondent à une augmentation de 49.3% de la moyenne géométrique des interactions entre les publications prépandémiques pas explicitement sur la science et celles pendant la pandémie liées à la science. Le contraste le plus important est entre les publications liées à la science avant et pendant la pandémie : une augmentation de 21.3% de la moyenne géométrique des interactions.

Pour interpréter plus finement cette augmentation des interactions, nous allons :

1. Étudier les publications liées à la science qui génèrent plus d'interactions
2. Étudier l'évolution temporelle des interactions sur la période

Types de publications liées à la science générant beaucoup d'interactions durant la pandémie

Pour étudier les publications liées à la science qui génèrent beaucoup d'interactions, nous pouvons d'abord diviser notre corpus par page :

```
fb_posts <- fread(paste0(accessible_data_folder, "FB_data.csv"))
# Focussing on science posts during covid
fb_sci_cov <- fb_posts[is_science == TRUE & since_COVID == TRUE]

fb_sci_cov[, list(number_posts = .N, mean_engagement = mean(log(interaction))),
            by = media][ order(-mean_engagement),
                        list(Média = media, `Publications` = number_posts,
                          `Interactions moyennes` = mean_engagement)]
```

##		Média Publications	Interactions moyennes
## 1:	TVA Nouvelles	32	8.089974
## 2:	Radio-Canada Information	37	7.119920
## 3:	Le Journal de Montréal	48	6.748127
## 4:	La Presse	52	6.555075
## 5:	ICI Estrie	7	6.390691
## 6:	FM Charlevoix - CIHO 96,3	1	6.383507
## 7:	Le Devoir	51	6.227487
## 8:	TVA Est-du-Québec	1	6.226537

## 9:	ICI Gaspésie-Les-Îles	2	6.200721
## 10:	Le Nord-Côtier	1	6.200509
## 11:	Les 2 Rives	1	6.169611
## 12:	L'Éclaireur Progrès	1	6.137727
## 13:	Le Journal de Québec	32	5.967271
## 14:	TVA Gatineau-Ottawa	3	5.911250
## 15:	FM 103,3 - La radio allumée	4	5.777029
## 16:	Le Charlevoisien	1	5.733341
## 17:	ICI Bas-Saint-Laurent	6	5.731439
## 18:	ICI Québec	10	5.670293
## 19:	TVA Sherbrooke	1	5.616771
## 20:	ICI Côte-Nord	3	5.616287
## 21:	L'Avantage	2	5.496513
## 22:	Journal L'Express de Drummondville	1	5.484797
## 23:	Journal Le Manic	2	5.407232
## 24:	ICI Ottawa Gatineau	1	5.384495
## 25:	Journal Saint-François	1	5.351858
## 26:	TVA Saguenay-Lac-St-Jean	1	5.347108
## 27:	98,5 FM	7	5.341966
## 28:	HuffPost Québec	21	5.325771
## 29:	104,7 Outaouais	5	5.277032
## 30:	Journal Mobiles	6	5.157548
## 31:	ICI Mauricie Centre-du-Québec	8	5.147413
## 32:	Le Placoteux	2	5.063595
## 33:	Le Soleil	32	5.008053
## 34:	L'actualité	22	4.978473
## 35:	ICI Abitibi-Témiscamingue	3	4.918454
## 36:	Courrier Laval	3	4.905311
## 37:	Le Quotidien	30	4.877176
## 38:	Journal Métro	17	4.870356
## 39:	Info Dimanche	2	4.826647
## 40:	La Tribune	23	4.808452
## 41:	ICI Grand Montréal	6	4.756319
## 42:	Le Nouvelliste	24	4.752249
## 43:	L'Avant-Poste	1	4.718499
## 44:	www.lanouvelle.net	1	4.663439
## 45:	106.9 FM Mauricie	2	4.657215
## 46:	La Voix de l'Est	15	4.600764
## 47:	TVA Québec	1	4.488636
## 48:	L'Hebdo du St-Maurice	2	4.353741
## 49:	Radio Gaspésie	2	4.353741
## 50:	Journal Le Nord	1	4.304065
## 51:	Le Courrier Sud	1	4.204693
## 52:	Monlimoilou	1	4.077537
## 53:	Hebdo Rive Nord	1	3.988984
## 54:	L'Hebdo Journal	2	3.977888
## 55:	VIA 90.5 FM	2	3.702748
## 56:	L'aut'journal	1	3.610918
## 57:	Radio Canada International	4	3.511980
## 58:	EstriePlus.com Le Journal Internet	2	3.465247
## 59:	Journal Outremont	1	3.433987
## 60:	CFIM, la radio des Îles de la Madeleine	9	2.441499
##	Média Publications Interactions moyennes		

```
rm(fb_posts)
```

On constate que ce sont surtout des pages de médias « nationaux » qui produisent le plus de publications liées à la science et qui génèrent le plus d'interactions moyennes.

Une analyse lexicale nous permet de caractériser davantage le contenu des publications générant beaucoup d'interactions. Nous avons d'abord effectué un étiquetage morphosyntaxique du sous-corpus des publications liées à la science pendant la COVID (`scripts/05-Facebook_corpus_preparation.R`). Nous divisons ensuite les publications en deux groupes : celles dont le log des interactions est sous la moyenne (interactions faibles) et celles dont le log est au-dessus de la moyenne (interactions fortes). Nous pouvons ainsi extraire les termes caractéristiques de ces deux groupes avec une analyse des résidus standardisés.

```
# Chargement des publications étiquetées
load(paste0(data_folder, "dt_tagged_tokens_FB_science_in_COVID.RData"))

# Calcul d'un score par publication par rapport à la moyenne du log des interactions
fb_sci_cov[, z_score := (log(interaction) -
                        mean(log(interaction)))/mean(log(interaction))]
dt_tagged_tokens_science <- merge(dt_tagged_tokens_science,
                                   fb_sci_cov[,list(postID,z_score)],
                                   by.x = "doc_id", by.y = "postID",
                                   all.x = TRUE)

save(dt_tagged_tokens_science, file =
     paste0(data_folder,
            "dt_tagged_tokens_FB_science_in_COVID_extended.RData"))

# On se limite à des segments de phrases pertinents
dt_lemmas <- dt_tagged_tokens_science[upos %in% c("ADJ","NOUN", "PROPN","VERB")]

# La distinction entre les deux corpus: faibles/fortes interactions
dt_lemmas[z_score < 0 , corpus:=1][is.na(corpus), corpus:=2]

dt_chi_sqr <- dt_lemmas[lemma != "<unknown>",
                       list(count = .N, weighted_count = sum(abs(z_score))),
                       by = .(lemma, corpus)]
dt_chi_sqr <- merge(dt_chi_sqr[corpus==1], dt_chi_sqr[corpus==2],
                   by = "lemma", all = TRUE)
dt_chi_sqr[is.na(corpus.x), `:=`(corpus.x = 1, count.x = 0, weighted_count.x = 0)]
dt_chi_sqr[is.na(corpus.y), `:=`(corpus.y = 2, count.y = 0, weighted_count.y = 0)]
setnames(dt_chi_sqr, c("corpus.x", "corpus.y", "count.x", "count.y",
                      "weighted_count.x", "weighted_count.y"),
         c("corpus_1", "corpus_2", "Lemma_count_1", "Lemma_count_2",
           "Lemma_weighted_count_1", "Lemma_weighted_count_2"))
#Éviter les divisions par zéro
dt_chi_sqr[, `:=`(Lemma_count_1 = Lemma_count_1 + 1, Lemma_count_2 = Lemma_count_2 + 1)]

res_chi <- chisq.test(dt_chi_sqr[,list(Lemma_weighted_count_1, Lemma_weighted_count_2)])

rm(dt_tagged_tokens_science)
```

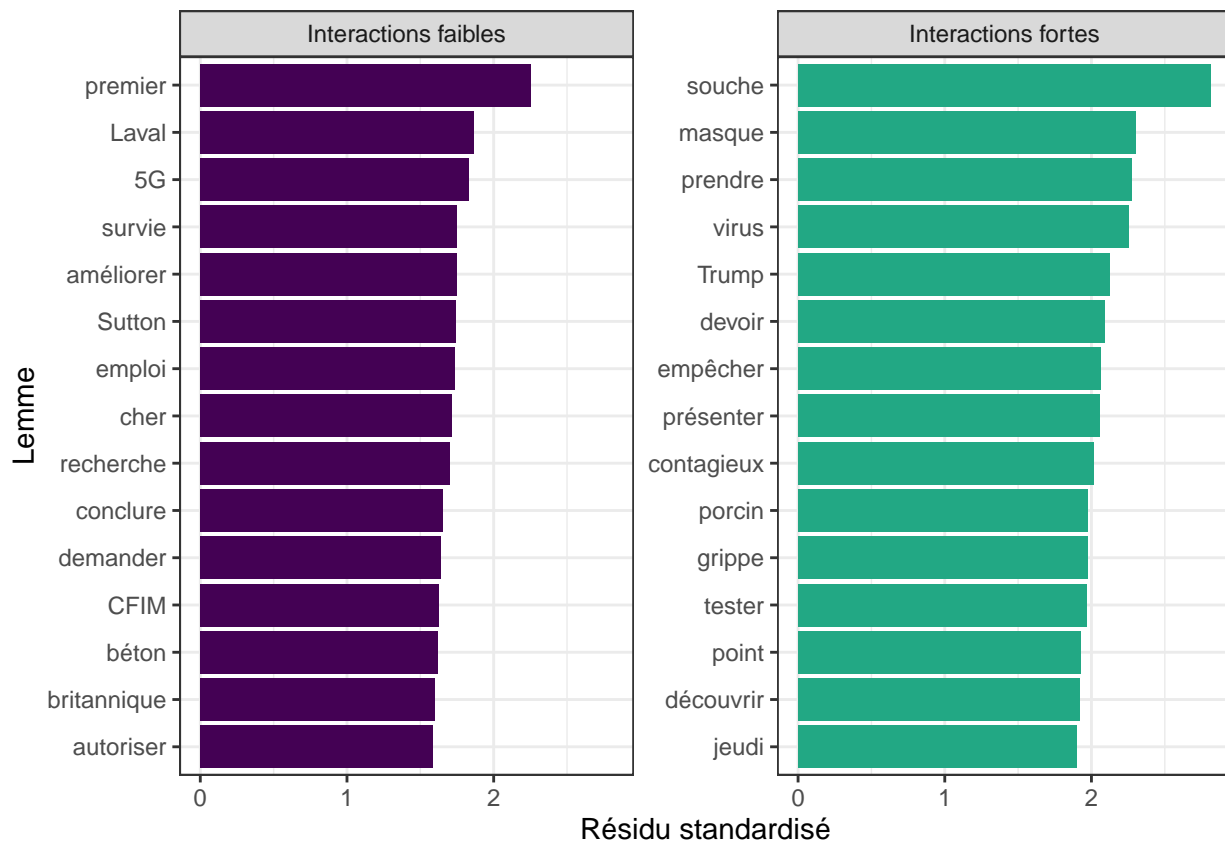
Création du graphique sur la base des résidus standardisés. Voir `figure_A4.jpg`.

```

dt_chi_sqr[, residual := res_chi$stdres[,2]]
setorder(dt_chi_sqr, residual)
lemmas_that_are_not_verbs <- c("quir", "unrir", "hydroxychloroquiner", "infecté",
                              "influenzer", "éternir", "immunitier", "épidémie",
                              "ne", "professeeur", "âgé", "professeur")
dt_chi_sqr <- dt_chi_sqr[!lemma %in% lemmas_that_are_not_verbs]
nb_inc <- 15
dt_plot <- rbind(dt_chi_sqr[, list(lemma, residual)] %>% head(nb_inc),
                dt_chi_sqr[, list(lemma, residual)] %>% tail(nb_inc))

dt_plot[residual > 0, corpus_title := "Interactions fortes"] [
  is.na(corpus_title), corpus_title := "Interactions faibles"]
dt_plot[corpus_title == "Interactions faibles", residual := -residual]
ggplot(dt_plot, aes(x = reorder(lemma, residual), y = residual, fill = corpus_title),
      col = corpus_title) +
  geom_bar(stat = "identity") + ylab("Résidu standardisé") + xlab("Lemme") +
  facet_wrap(~corpus_title, scales = "free_y",) +
  scale_x_reordered() + guides(fill="none") +
  # scale_y_continuous(expand=c(0,1)) +
  theme_bw() + coord_flip() + scale_fill_viridis( discrete=TRUE, end = .6)

```



```

ggsave(file=paste0(fig_dir, "figure_A4.jpg"), width = 6, height = 4)

```

Quelques termes dans ce graphique sont un peu surprenants. Nous extrayons le contexte de leur utilisation pour mieux comprendre :

Contexte du terme 'souche'

. Une nouvelle souche de le virus
identifié une nouvelle souche de le coronavirus
ont découvert une souche de virus de
monde . Cette souche serait devenue dominante
COVID-19 Une nouvelle souche de le virus

. La deuxième souche aurait frappé à
pandémie . La souche « possède tous

```
load(paste0(data_folder, "dt_tagged_tokens_FB_science_in_COVID_extended.RData"))
dt_tagged_tokens_science[, termid := 1:.N]

fct_for_context_extraction <- function(lemma_to_search, n_before_after = 3,
                                       high_engagement = TRUE){
  if(high_engagement == TRUE){
    ids <- dt_tagged_tokens_science[lemma == lemma_to_search & z_score > 0, termid]
  } else if(high_engagement == FALSE){
    ids <- dt_tagged_tokens_science[lemma == lemma_to_search & z_score < 0, termid]
  }
  ids_neighbor <- lapply(ids, function(x) (x - n_before_after):(x + n_before_after))

  strings <- lapply(ids_neighbor,
                    function(x) dt_tagged_tokens_science[termid %in% x,
                    paste(token, collapse = " ")] ) %>%
  unlist %>% unique()

  df <- data.frame(strings)
  colnames(df) <- paste0("Contexte du terme '", lemma_to_search, "'")

  return( df %>% kable(booktabs = T) %>% kable_styling( ) %>%
    scroll_box(width = "100%", height = "200px")
  )
}

fct_for_context_extraction(lemma_to_search = "souche")
fct_for_context_extraction(lemma_to_search = "prendre")
fct_for_context_extraction(lemma_to_search = "Trump")
fct_for_context_extraction(lemma_to_search = "devoir")
fct_for_context_extraction(lemma_to_search = "jeudi")
fct_for_context_extraction(lemma_to_search = "premier", high_engagement = FALSE)
fct_for_context_extraction(lemma_to_search = "Laval", high_engagement = FALSE)
fct_for_context_extraction(lemma_to_search = "5G", high_engagement = FALSE)
fct_for_context_extraction(lemma_to_search = "survie", high_engagement = FALSE)
```

Contexte du terme 'prendre'

un pari que prend le gouvernement provincial
là les décideurs prennent de les décisions
quelques-unes décortiquées . Prise en charge de
, j' en prends depuis une semaine
, qu' il prenait un comprimé d'
pour les parents pris à la maison
COVID-19 qui avaient pris en automédication ...
COVID-19 qui avaient pris en automédication de
Arruda s' en prend à les «
, j' en prends depuis une semaine
le Canada avait pris de les mesures
«Si le gouvernement prend à le sérieux
différentes communautés et prend en considération la
décision , également prise par Starbucks et
et les connaissances prennent de le temps
.» Donald Trump prend de l' hydroxychloroquine
ces deux branches prend source dans un
de les mesures prises pour protéger les

Contexte du terme 'Trump'

endossé par Donald Trump pour soigner les
Même si Donald Trump et un scientifique
désinfectant » ? Trump accusé d' être
déclarations de Donald Trump jeudi ont consterné
a déclaré Donald Trump lors de son
presse quotidien . Trump , coronavirus et
la COVID-19 . Trump , mépris de
York demande à Trump d' écouter la
bit.ly/39TQCpO Donald J. Trump a partagé un
de les patchs Trump , coronavirus et
par 2500 » Trump , mépris de
coup , Donald Trump a annoncé lundi
les propos de Trump sur le désinfectant
les propos de Trump soulèvent la controverse
le gouvernement , Trump recommande les masques
président américain Donald Trump a annoncé vendredi
les propos de Trump déclenchent un tollé
américaines . Donald Trump : mépris de
Le virus , Trump et la prise
de le gouvernement Trump pour affronter la
politique .» Donald Trump prend de l'
a déclaré M. Trump . Tous les

Contexte du terme 'devoir'

contraire qu' ils doivent être maintenus .
système d' éducation doit passer avant un
: les régions devront rester fermées ,
pourrait -il être dû à les recherches
. «Arruda ne doit pas perdre sa

Laliberté «Arruda ne doit pas perdre sa
vedettariat qu' il doit à le premier
. Le gouvernement doit imposer un moratoire
. Les régions devront rester fermées ,
coronavirus , elle devrait être l' une

chercheuse Les régions devront rester fermées ,
pandémie de coronavirus devrait durer à le
pandémie de coronavirus devrait durer de 18
de la population doit porter le masque
Ce qui ne devait être qu' une

le sapin baumier devrait laisser sa place
brevetée , mais devra maintenant passer l'
même nombreux à devoir être hospitalisés .
les enseignants ne devraient pas hésiter à
de l' Éducation devrait leur envoyer une

à la ferme devrait être illégale ,
à la ferme devrait être illégal ,
climatique , il doit inscrire dans la
— via Le Devoir - Opinion Un
2030 . Legault doit respecter la science

— via Le Devoir - Opinion Katherine
— via Le Devoir - Opinion Il
— via Le Devoir - Opinion La
— via Le Devoir - Opinion Surveillés
— via Le Devoir - Opinion Découverte

autres ? Le Devoir a demandé à
— via Le Devoir - Opinion Se
— via Le Devoir - Opinion Sur

Contexte du terme 'jeudi'

de Donald Trump jeudi ont consterné la
Machouf a imploré jeudi le gouvernement de
rapport universitaire publié jeudi . Selon les
, a annoncé jeudi un scientifique à
Une note publiée jeudi par l' Institut
étudiants ont réclamé jeudi l' inclusion ,

Contexte du terme 'premier'

. Pour la première fois de l'
laboratoire . Ma première réaction en a
à lancer les premiers tests d' un
a autorisé les premiers essais cliniques pour
annoncé samedi le premier ministre Justin Trudeau
: pour la première fois , un
montré pour la première fois qu' il
sont à les premières lignes dans ce
, monsieur le Premier Ministre ? «
] Comme le premier ministre de le
alimenter . Le premier ministre François Legault
atteindre son but premier : devenir médecin
: Sutton , première ville à le
Ville est la première à adopter une
. Dans une première entrevue accordée à
... Après une première vague de faible
autorités ... Le premier ministre François Legault
été la toute première personne de la
C' est la première enquête sur les
chef de le premier ministre de le
a autorisé un premier test de dépistage
Alerte : La 1ère transfusion de plasma
à Montréal La première transfusion de plasma

Contexte du terme 'Laval'

de l' Université Laval , la professeure
de l' Université Laval Luca Sorelli que
de l' Université Laval . La consommation
de l' Université Laval . COVID-19 :
de l' Université Laval est ouverte et
CHU de Québec-Université Laval , le Dr
de l' Université Laval . Les conséquences
CHU de Québec-Université Laval et professeur à
de l' Université Laval . Le Dr
de l' Université Laval . Les fameux
de l' Université Laval OPINION / L'
de l' Université Laval GNL Québec: la
de l' Université Laval . Trois médicaments
à l' Université Laval Une centaine de
de l' Université Laval pourraient être euthanasiés
de l' Université Laval . Coronavirus :
COVID-19 | Courrier Laval Les chercheurs et
Canada | Courrier Laval Santé Canada a
machine fabriquée à Laval et développé ...
machine fabriquée à Laval et développée à

Contexte du terme '5G'

général . Technologie 5G : de la
envers la technologie 5G et les ondes
. Téléphonie cellulaire 5G : Sutton ,
la technologie cellulaire 5G jusqu' à ce
regroupement citoyen Stop 5G Sutton . Le
regroupement citoyen Stop 5G Sutton . Mégadoses
Montréal . Technologie 5G : de la
regroupement citoyen Stop 5G Sutton . Les
... Téléphonie cellulaire 5G : Sutton ,
regroupement citoyen Stop 5G Sutton . Rendre
. Technologie cellulaire 5G : Valleyfield se
la technologie cellulaire 5G . Celle-ci provoque

Contexte du terme 'survie'

semble améliorer la survie PARIS — Un
il améliorerait la survie chez de les

```
rm(dt_tagged_tokens_science)
```

Avec ces contextes, nous voyons que les publications générant plus d'interactions sont davantage liées à la pandémie. Une surprise est la présence de "5G" dans les termes pour les interactions faibles. Nous regardons les publications liées à ce terme :

```
load(paste0(data_folder, "dt_tagged_tokens_FB_science_in_COVID_extended.RData"))

dt_tagged_tokens_science[lemma == "5G" & z_score < 0,
                           list( `Publications mentionnant 5G` =
                                sentence %>% unique) ] %>%
  kable( booktabs = T) %>% kable_styling(full_width = F) %>%
  column_spec(1, width = "45em")
```

Publications mentionnant 5G

Technologie 5G: de la vitesse au détriment de la santé ?

L'inquiétude d'une partie de la population envers la technologie 5G et les ondes électromagnétiques en général est partagée par le physicien, professeur à la Faculté de médecine de l'Université McGill et directeur du programme de santé au travail, Paul Héroux.

Téléphonie cellulaire 5G: Sutton, première ville au Canada à demander un moratoire

Les élus de Sutton demandent au gouvernement fédéral de déclarer un moratoire sur l'implantation de la technologie cellulaire 5G jusqu'à ce qu'un consensus scientifique prouve qu'elle n'est pas nuisible ni à la santé ni à l'environnement.

La Ville est la première à adopter une telle résolution au Canada, selon le regroupement citoyen Stop 5G Sutton.

Technologie cellulaire 5G: Valleyfield se laissera guider par la science - Journal Saint-François Aucune demande n'a été formulée de la part d'un citoyen afin que Valleyfield se positionne sur la technologie cellulaire 5G.

```
rm(dt_tagged_tokens_science)
```

Les extraits sur la 5G ne sont effectivement pas liés à la pandémie.

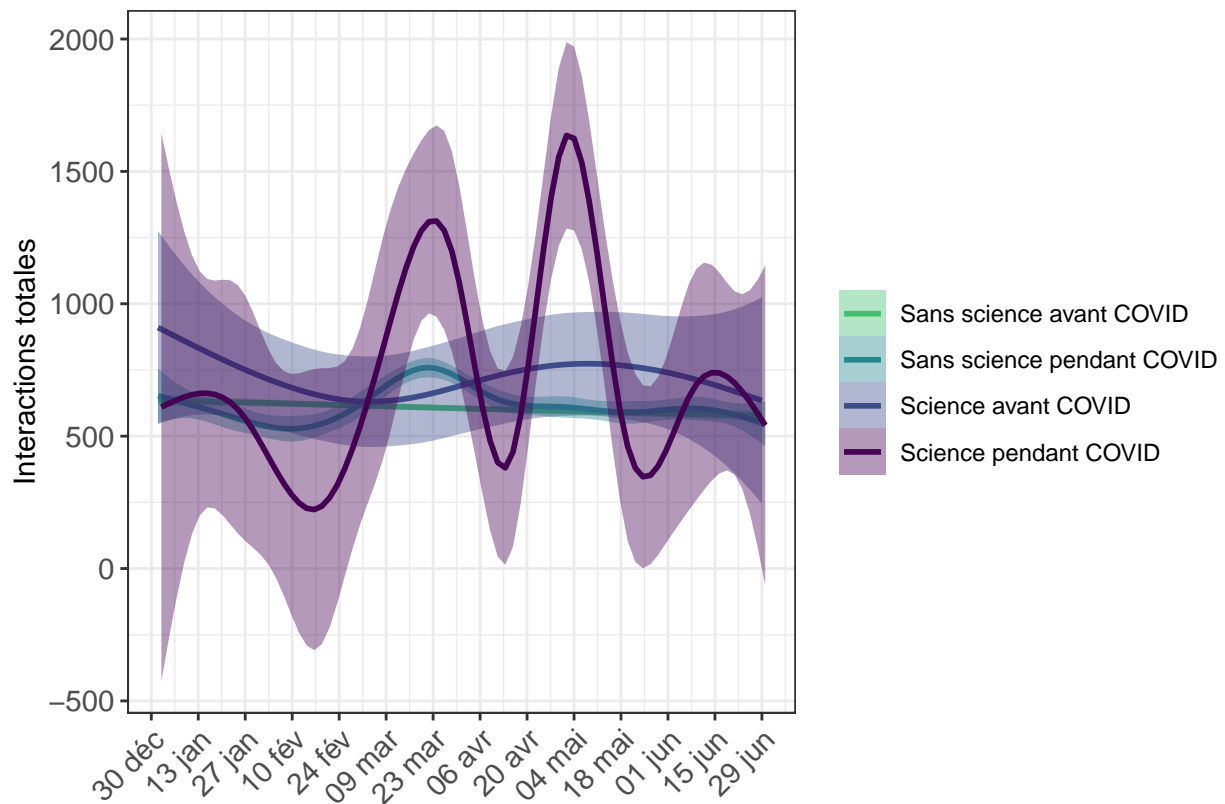
Évolution temporelle des interactions

Maintenant, nous étudions la distribution des interactions totales sur les 6 premiers mois de chaque année (avant et pendant la première vague de COVID). Voir la [figure5.jpg](#).

```
fb_posts <- fread(paste0(accessible_data_folder, "FB_data.csv"))

p <- ggplot(fb_posts#[semaine >= 10]#[is_science == FALSE]
, aes(x=as.Date(day_of_year-1,format = "%j", origin = as.Date("2020-01-01")),
      y= interaction, fill= Status2, col= Status2 )) +
geom_smooth(alpha=.4, method = "gam", method.args = list(method="GACV.Cp")) + labs(
y = "Interactions totales", x = "", fill = "", col = "") +
scale_fill_viridis(discrete=TRUE, end = .7, direction = -1) +
scale_color_viridis(discrete=TRUE, end = .7, direction = -1) +
scale_x_date(date_breaks = "2 week", date_labels = "%d %b") + theme_bw() +
theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1, size = size_axis_text),
axis.text.y = element_text(size = size_axis_text))
```

p



```
ggsave(file=paste0(fig_dir,"figure5.jpg"),width = 10, height=5)
```

```
rm(fb_posts, p)
```

On voit deux périodes où la distribution des interactions pour les publications liées à la science pendant la pandémie surpassent significativement les autres distributions:

1. De la mi-mars au début avril
2. De la mi-avril à la mi-mai

Alors, quelles publications ont généré beaucoup d'interactions durant ces périodes?

La première période :

```
load(paste0(data_folder, "dt_fb_posts_all_after_preprocessing.RData"))

dt <- fb_posts[day_of_year %in% c(75:95) & is_science == TRUE & since_COVID == TRUE &
  high_engagement == TRUE] [
  order(-interaction), list(interaction, day = jour %>% strftime(format = "%B %d"),
    titre, message, description, media,lien)
]

if(outputFormat == "html"){
  dt %>% kable() %>% kable_styling() %>%
    scroll_box(width = "100%", height = "200px")
} else{
  dt %>% kable( booktabs = T, longtable = TRUE) %>%
    kable_styling(full_width = F) %>%
    column_spec(3:7, width = "7em")
}
```

interaction	day	titre	message	description	media	lien
15986	mars 23	Un traitement potentiel du coronavirus testé au Canada dès lundi Coronavirus	6000 Canadiens sont appelés à participer à l'étude clinique.	Une équipe de chercheurs montréalais veut recruter 6000 Canadiens d'ici huit semaines.	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/168-coronavirus-covid19-tardif-montreal-canada-quebec-colcorona
13881	mars 16	Un nouveau-né testé positif à la COVID-19 peu après sa naissance	Les chercheurs tentent de savoir à quel point le virus est dangereux pour les bébé naissants.	Un bébé né à Londres a été testé positif au nouveau coronavirus quelques minutes seulement après sa naissance.	TVA Nouvelles	https://www.tvanouvelles.ca/13881-nouveau-ne-teste-positif-a-la-covid-19-peu-apres-sa-naissance?sd

8300	mars 24	Des traces du virus 17 jours plus tard	Des chercheurs ont découvert des traces du virus dans des cabines du Diamond Princess, 17 jours après le retrait des passagers et membres d'équipage.	Un rapport des Centres pour le contrôle et la prévention des maladies américains permet d'en apprendre plus sur la durée de vie du coronavirus.	TVA Nouvelles	https://www.tvanouvelles.ca/2020/03/24/des-traces-du-virus-17-jours-plus-tard
4944	mars 18	Le virus peut survivre plusieurs heures à l'air libre	Pour mener leur étude, les scientifiques ont cherché à imiter sa propagation lors d'éternuements	Le nouveau coronavirus provoquant la COVID-19 peut survivre en suspension dans l'air pendant environ trois heures.	TVA Nouvelles	https://www.tvanouvelles.ca/2020/03/18/le-virus-peut-survivre-plusieurs-heures-a-lair-libre
4743	mars 23	Des chercheurs québécois testeront un médicament contre les complications graves	Il se trouve déjà sur les tablettes des pharmacies de partout sur la planète, coûte moins de 1 \$ par jour et pourrait empêcher les complications graves de la COVID-19. Des chercheurs de l'Institut de cardiologie de Montréal démarrent une étude à grande échelle visant à tester un médicament appelé colchicine contre la détresse respiratoire causée par la COVID-19.	Il se trouve déjà sur les tablettes des pharmacies de partout sur la planète, coûte moins de 1 \$ par jour et pourrait empêcher les complications graves de ...	La Presse	https://www.lapresse.ca/2020/03/23/01-5265931-des-chercheurs-quebecois-testeront-un-medicament-contre-les-complications-graves.php?utm_source=google

3588	mars 25	Créer un respirateur artificiel en deux semaines pour sauver des vies	Ingénieurs et scientifiques du monde entier répondent au défi de concevoir un respirateur artificiel qui serait simple et peu coûteux à produire.	Des ingénieurs et scientifiques du monde entier répondent au défi lancé au Québec pour concevoir un respirateur artificiel qui serait simple et peu coûteux à produire. Car l'appareil médical sauvera la vie de milliers de personnes gravement atteintes de la COVID-19 et les hôpitaux craignent...	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/1688-artificiel-code-vie-hopital-general-cusm-coronavirus
3149	avril 02	Des chercheurs pensent avoir trouvé une autre façon de décontaminer les masques N95 Coronavirus	Des chercheurs manitobains ont obtenu des résultats encourageants pour décontaminer et réutiliser jusqu'à 10 fois certains masques.	Alors que le personnel médical peine à se procurer des masques N95 pour se protéger du nouveau coronavirus, des chercheurs manitobains ont obtenu des résultats encourageants pour décontaminer et réutiliser jusqu'à 10 fois certains masques, à partir d'une technique d'autoclavage, qui pourrait...	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/1690-respirateurs-coronavirus-medecin-hopitaux-penurie-decontaminer

3142	mars 20	Les symptômes gastro-intestinaux sous-estimés ?	Une nouvelle étude chinoise indique que les symptômes gastro-intestinaux comme la diarrhée peuvent précéder l'apparition de la toux et des problèmes respiratoires chez les patients atteints de la COVID-19. Pendant ce temps, des chercheurs américains confirment que si les jeunes meurent beaucoup moins que les vieux de la maladie, ils sont quand même nombreux à devoir être hospitalisés.	Une nouvelle étude chinoise indique que les symptômes gastro-intestinaux comme la diarrhée peuvent précéder l'apparition de la toux et des problèmes respir...	La Presse	https://www.lapresse.ca/19/202003/19/01-5265568-les-symptomes-gastro-intestinaux-sous-estimes-.php?utm_source=faceb
2867	mars 25	Il faut demander aux profs d'appeler leurs élèves, suggère une chercheure	Non seulement les enseignants ne devraient pas hésiter à entrer en contact avec les élèves et leurs familles, mais en plus, le ministre de l'Éducation devrait leur envoyer une directive claire pour leur dire de le faire, recommande une chercheure de l'Université McGill.	Non seulement les enseignants ne devraient pas hésiter à entrer en contact avec les élèves et leurs familles, mais en plus, le ministre de l'Éducation devr...	La Presse	https://www.lapresse.ca/19/202003/25/01-5266437-il-faut-demander-aux-profs-dappeler-leurs-eleves-suggere-une-chercheure.php?utm_so

2404	mars 18	Les mesures de confinement pourraient durer des mois	Les scientifiques ont une bonne et une mauvaise nouvelle pour les parents pris à la maison avec de jeunes enfants, les aînés isolés et tous les Québécois qui souffrent du confinement.	Les scientifiques ont une bonne et une mauvaise nouvelle pour les parents pris à la maison avec de jeunes enfants, les aînés isolés et tous les Québécois q...	La Presse	https://www.lapresse.ca/19/202003/17/01-5265172-les-mesures-de-confinement-pourraient-durer-des-mois.php?utm_source=
2179	mars 23		Des chercheurs vont recruter 6000 Canadiens dans l'espoir de trouver un traitement efficace contre la COVID-19. rc.ca/RSQm36		Radio-Canada Information	https://www.facebook.c
2038	avril 03	Trois médicaments contre la COVID-19 testés à Sherbrooke	Des chercheurs sherbrookoïses testeront trois médicaments auprès de patients volontaires déclarés positifs à la COVID-19 qui sont hospitalisés au CIUSSS de l'Estrie-CHUS.	Des chercheurs sherbrookoïses testeront trois médicaments auprès de patients volontaires déclarés positifs à la COVID-19 qui sont hospitalisés au CIUSSS de l'Estrie-CHUS.	La Tribune	https://www.latribune.c/medicaments-contre-la-covid-19-testes-a-sherbrooke-bde8f0335a212d5a063c9
1881	mars 30	Pandémie de coronavirus: une aide respiratoire conçue avec des ingénieurs de Formule 1	Des chercheurs, des médecins et des ingénieurs l'équipe Mercedes de Formule 1 ont développé en moins d'une semaine une assistance respiratoire permettant de soulager les poumons des malades.	Des chercheurs, des médecins et des ingénieurs l'équipe Mercedes de Formule 1 ont développé en moins d'une semaine une assistance respiratoire.	Le Journal de Montréal	https://www.journalden/de-coronavirus-une-aide-respiratoire-concue-avec-des-ingenieurs-de-formule-1

1655	avril 01	Les jeunes, un facteur de propagation déterminant de la COVID-19? Coronavirus	Au Canada, une personne sur trois infectée par le coronavirus a moins de 40 ans. Ce chiffre important fait dire aux experts qu'ici et ailleurs les jeunes risquent, sans le savoir, d'accélérer la propagation de la COVID-19.	Des chercheurs affirment que les jeunes risquent, sans le savoir, d'accélérer la propagation de la COVID-19.	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/169enfants-covid-19-coronavirus-propagation
1607	mars 19	Coronavirus et alimentation : aucune pénurie en vue, assure un chercheur Coronavirus	Le Canada sera le dernier pays à manquer de nourriture, selon Sylvain Charlebois, analyste du secteur agroalimentaire canadien.	La capacité logistique du système d'approvisionnement serait particulièrement forte, estime le spécialiste.	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/167coronavirus-covid-19-canada-penurie
1537	mars 28	Le gouverneur de New York demande à Trump d'écouter la science et d'aider son État Coronavirus	Pendant que les politiciens peinent à s'entendre, les besoins explosent dans l'État de New York.	Le refus du président de fournir plus de matériel médical à l'État de New York enrage son gouverneur.	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/168hopitaux-andrew-cuomo-new-york-trump-ecouter-science

1421	mars 29	Des «cas de toxicité cardiaque» après la prise d'hydroxychloroquine	Des « cas de toxicité cardiaque » ont été signalés sur des personnes présentant des symptômes de la COVID-19 qui avaient pris en automédication de l'hydroxychloroquine, promu par certains scientifiques comme un remède possible contre le virus, mettent en garde dimanche les autorités sanitaires de Nouvelle-Aquitaine, une région de France.	(Bordeaux) Des « cas de toxicité cardiaque » ont été signalés sur des personnes présentant des symptômes de la COVID-19 qui avaient pris en automédication ...	La Presse	https://www.lapresse.ca/5267000-des-cas-de-toxicite-cardiaque-apres-la-prise-dhydroxychloroquine.ph
1246	mars 15	Coronavirus : vous nous avez posé des questions, nos experts vous répondent	«Un récent article scientifique nous apprend que le COVID-19 peut survivre 72 heures sur du plastique ou de l'acier inoxydable et 24 heures sur une surface en carton.» - Dr Caroline Quach-Thanh, pédiatre, microbiologiste et infectiologue au CHU Sainte-Justine	8 questions ont été retenues.	Le Journal de Montréal	http://www.journaldemontreal.com/vous-nous-avez-pose-des-questions-nos-experts-vous-repondent

1241	avril 01		[Carbone] Dans les cas les plus graves de COVID-19, la principale menace à la vie du patient n'est pas le virus lui-même, mais bien une forte réaction de son système immunitaire. Voici deux scientifiques qui cherchent à calmer cette tempête inflammatoire.		Radio-Canada Information	https://www.facebook.com/radio-canada
943	mars 16	La course au vaccin	Les scientifiques estiment qu'il faudra attendre au moins jusqu'à l'automne 2020 avant qu'un vaccin contre la COVID-19 ne voie le jour.	Les scientifiques estiment qu'il faudra attendre au moins jusqu'à l'automne 2020 avant qu'un vaccin contre la COVID-19 ne voie le jour.	Le Journal de Montréal	https://www.journaldelequipe.com/covid-19/la-course-au-vaccin
902	mars 24	Les fameux examens du ministère!	OPINION / Ce texte a été rédigé par Nicole Monney et Stéphane Allaire, professeurs au Département des sciences de l'éducation à l'Université du Québec à Chicoutimi (UQAC).	OPINION / Ce texte a été rédigé par Nicole Monney et Stéphane Allaire, professeurs au Département des sciences de l'éducation à l'Université du Québec à Chicoutimi (UQAC).	Le Quotidien	https://www.lequotidien.com/actualites/les-fameux-examens-du-ministere-34ee7a2689101f92074804

784	mars 24	Les données cellulaires pour cibler les lieux de rassemblements ?	Le maire de Toronto ainsi qu'un groupe de chercheurs de prestigieuses universités américaines suggèrent d'utiliser des données de positionnement de millions de cellulaires pour surveiller les efforts de distanciation sociale et créer une carte des points chauds de rassemblements. Bell dit être prête à « examiner » cette possibilité en gardant en tête les lois sur la protection de la vie privée.	Le maire de Toronto ainsi qu'un groupe de chercheurs de prestigieuses universités américaines suggèrent d'utiliser des données de positionnement de million...	La Presse	https://www.lapresse.ca/19/202003/24/01-5266284-les-donnees-cellulaires-pour-cibler-les-lieux-de-rassemblements-.php?utm_source=faceb
734	mars 27		Découvert au milieu du 19e siècle par l'obstétricien hongrois Ignace Semmelweis, le lavage des mains est aujourd'hui une pratique banale. À l'époque, la communauté scientifique rejette cependant la découverte du médecin, dont la vie se termine tragiquement.		Radio-Canada Information	https://www.facebook.c

720	mars 25	Un superordinateur à la rescousse	L'ordinateur le plus puissant du monde s'est creusé les méninges pour trouver 77 composés médicamenteux qui ont le potentiel d'empêcher le coronavirus d'infecter le corps humain. (via En 5 minutes)	Deux chercheurs ont analysé les structures de plus de 8000 médicaments, produits chimiques, drogues et autres produits naturels.	Le Journal de Montréal	https://www.journaldenouvelles.com/superordinateur-a-la-rescousse
687	mars 30	Non, il n'y a pas de liens démontrés entre les groupes sanguins et la COVID-19	Est-ce que notre groupe sanguin peut influencer sur notre niveau de vulnérabilité face au nouveau coronavirus (SRAS-CoV-2) ? Cette hypothèse circule abondamment sur le web et des dizaines de personnes ont écrit aux Décodeurs pour en savoir plus à ce sujet.	Groupes sanguins et nouveau coronavirus : pas de preuves scientifiques validées.	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/1688000-groupes-sanguins-coronavirus-pas-preuve
669	mars 28	COVID-19 : y a-t-il un lien entre la pollution et la propagation du coronavirus?	Il existerait un lien entre la pollution atmosphérique et la propagation rapide du coronavirus, selon une étude récente de chercheurs italiens. Mais cette étude ne fait pas l'unanimité.	Selon des chercheurs italiens, la pollution atmosphérique serait un vecteur de propagation du coronavirus.	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/1688000-contamination-infection-epidemie

649	mars 29	Sophie Grégoire Trudeau est guérie de la COVID-19	Elle a ajouté qu'elle croit «fermement que la science ET la compassion nous aideront à traverser cette crise».	Dans un message Facebook publié samedi soir, la conjointe du premier ministre, Sophie Grégoire Trudeau, a annoncé être guérie du coronavirus.	Le Journal de Québec	https://www.journaldequbec.com/2020/03/29/sophie-gregoire-trudeau-est-gueree-de-la-covid-19/
580	mars 24	Médecine et sciences infirmières: les finissants pourraient être appelés en renfort	Ils pourraient participer à l'effort de guerre, même s'ils n'ont pas encore obtenu leur diplôme.	Les finissants en médecine et en sciences infirmières pourraient intégrer le système de santé rapidement pour participer à l'effort de guerre.	Le Journal de Québec	https://www.journaldequbec.com/2020/03/24/medecine-et-sciences-infirmieres-les-finissants-pourraient-etre-appelles-en-renfort/
558	avril 03		Avancer plus rapidement que le virus pour enfin l'arrêter, c'est le défi des chercheurs en intelligence artificielle. Aperçu de différents modèles développés actuellement dans le monde.		Radio-Canada Information	https://www.facebook.com/radiocanada/posts/10158444444444444
538	mars 31	La poudre de perlimpinpin	La chronique de Richard Martineau.	S'il y a quelque chose que cette crise aura prouvé, c'est que pour se sortir du pétrin, il n'y a rien comme la science.	Le Journal de Montréal	https://www.journaldequbec.com/2020/03/31/la-poudre-de-perlimpinpin/

534	mars 25	Il pourrait y avoir un lien entre la COVID-19 et la pollution atmosphérique	L'hypothèse reste encore à vérifier, mais le virus pourrait être en suspension dans l'air ambiant là où il y a le plus de particules, et donc être plus facile à contracter.	Les particules dans l'air ambiant pourraient être un véhicule de transport du virus, selon des chercheurs italiens.	Le Devoir	https://www.ledevoir.com/sant%C3%A9/5266360-des-chercheurs-de-ludem-sattaquent-au-profil-genetique-du-virus.php?utm_source=
526	mars 25	Des chercheurs de l'UdeM s'attaquent au profil génétique du virus	Des chercheurs de l'Université de Montréal s'attaquent à la caractérisation génétique du coronavirus responsable de la COVID-19.	(Montréal) Des chercheurs de l'Université de Montréal s'attaquent à la caractérisation génétique du coronavirus responsable de la COVID-19.	La Presse	https://www.lapresse.ca/monreal/202003/25/01-5266360-des-chercheurs-de-ludem-sattaquent-au-profil-genetique-du-virus.php?utm_source=
516	mars 18	COVID-19: Pascal Bérubé interpelle les conspirationnistes	«Tant qu'à s'isoler du consensus scientifique, j'invite les expéditeurs de ces messages à s'isoler de leur clavier et à cesser de nuire aux efforts collectifs».	Le chef intérimaire du Parti québécois, Pascal Bérubé, en a eu assez des messages évoquant des théories du complot qu'il recevait sur la COVID-19.	Le Journal de Montréal	https://www.journaldeledevoir.com/monreal/516-berube-interpelle-les-conspirationnistes
462	avril 03	Un chercheur et un médecin s'excusent après des propos «méprisants» sur l'Afrique	Les deux experts s'interrogeaient sur l'opportunité de tester un vaccin en Afrique contre le coronavirus, «un peu comme c'est fait d'ailleurs sur certaines études avec le sida, ou chez les prostituées».	Les deux experts français s'interrogeaient sur l'opportunité de tester un vaccin en Afrique contre le coronavirus.	Le Devoir	https://www.ledevoir.com/sant%C3%A9/462-chercheur-et-un-medecin-s-excuse-apres-des-propos-tres-critiques-sur-l-afrique?utm_medium=

441	avril 03	Le coronavirus transmet en parlant, selon des scientifiques du gouvernement, Trump recommande les masques	Le président américain Donald Trump a annoncé vendredi que les autorités sanitaires conseillaient désormais aux Américains de se couvrir le visage lorsqu'ils sortent de chez eux.	(Washington) Les scientifiques du gouvernement américain estiment que le nouveau coronavirus est sans doute transmis par les gens lorsqu'ils parlent et res...	La Presse	https://www.lapresse.ca/unis/202004/03/01-5267924-les-americains-appelles-a-se-couvrir-le-visage-a-lexterieur.php?utm_sou
412	mars 25	Créer un respirateur artificiel en deux semaines pour sauver des vies	Depuis une semaine, Louis-Pierre Fortin et son équipe tentent de mettre au point un ventilateur composé de pièces d'équipement automobile et d'autres composantes faciles à trouver.	Des ingénieurs et scientifiques du monde entier répondent au défi lancé au Québec pour concevoir un respirateur artificiel qui serait simple et peu coûteux à produire. Car l'appareil médical sauvera la vie de milliers de personnes gravement atteintes de la COVID-19 et les hôpitaux craignent...	ICI Québec	http://ici.radio-canada.ca/nouvelle/168-artificiel-code-vie-hopital-general-cusm-coronavirus
370	mars 30	On pourrait désinfecter et réutiliser les masques N95, selon des chercheurs	Il serait possible de désinfecter et de réutiliser les précieux masques N95, sans que le nettoyage nuise à leur efficacité.	(Montréal) Il serait possible de désinfecter et de réutiliser les précieux masques N95, sans que le nettoyage nuise à leur efficacité.	La Presse	https://www.lapresse.ca/19/202003/30/01-5267066-on-pourrait-desinfecter-et-reutiliser-les-masques-n95-selon-des-chercheurs.php?utm_so
369	mars 26		Les scientifiques du monde entier participent à la course contre la montre pour trouver un remède et un vaccin à la COVID-19.		Le Journal de Montréal	https://www.facebook.c

353	mars 25	COVID-19 et pollution atmosphérique pourraient être liés	Il pourrait exister un lien entre la pollution atmosphérique et la propagation de la COVID-19, croient des chercheurs italiens.	(Montréal) Il pourrait exister un lien entre la pollution atmosphérique et la propagation de la COVID-19, croient des chercheurs italiens.	La Presse	https://www.lapresse.ca/19/202003/25/01-5266345-covid-19-et-pollution-atmospherique-pourraient-etre-lies.php?utm_source=fa
-----	---------	--	---	--	-----------	---

347	mars 23	Les conséquences dramatiques du déni de la science	ANALYSE / Dans les milieux ultraconservateurs en Occident, mais plus particulièrement aux États-Unis, il existe une forte tendance à rejeter les études scientifiques et les analyses des experts sur une variété de tragédies appréhendées. Dans cette guerre à la science, les conclusions scientifiques deviennent tout simplement de fausses nouvelles et des canulars inventés par des partisans démocrates ou des penseurs libéraux. Cela apparaît très clairement dans le déni du réchauffement climatique. La présente pandémie est en partie aussi le résultat de cette attitude de l'autruche.	ANALYSE / Dans les milieux ultraconservateurs en Occident, mais plus particulièrement aux États-Unis, il existe une forte tendance à rejeter les études scientifiques et les analyses des experts sur une variété de tragédies appréhendées. Dans cette guerre à la science, les conclusions scient...	Le Soleil	https://www.lesoleil.com/consequences-dramatiques-du-deni-de-la-science-1c7f47799b60056c3009c9
340	mars 20	La vague de l'épidémie de coronavirus continue à monter en Afrique subsaharienne	«Les scientifiques estiment que jusqu'à 60 % de la population [sud-africaine] pourrait être affectée par le virus».	En Afrique du Sud, «jusqu'à 60 % de la population pourrait être affectée par le virus», selon les scientifiques.	Le Devoir	https://www.ledevoir.com/vague-de-l-epidemie-de-coronavirus-continue-a-monter-en-afrique-subsaharienne?utm_campaign=

340	mars 23	Un traitement potentiel du coronavirus testé au Canada dès lundi Coronavirus	Le médicament – un anti-inflammatoire très puissant – pourrait réduire le risque de complications pulmonaires liées à la COVID-19.	Une équipe de chercheurs montréalais veut recruter 6000 Canadiens d'ici huit semaines.	ICI Estrie	https://ici.radio-canada.ca/nouvelle/168-coronavirus-covid19-tardif-montreal-canada-quebec-colcorona
324	mars 24	Gestion de crise, tartelettes et hybridité médiatique	Horacio Arruda s'est inscrit dans la culture populaire ; cela traduit l'efficacité de son entreprise de communication. Un texte d'opinion de Philippe Dubois, doctorant en science politique, Groupe de recherche en communication politique, Université Laval – via Le Devoir - Opinion		Le Devoir	https://www.ledevoir.com/opinion/575607/gestion-de-crise-tartelettes-et-hybridite-mediatique?utm_medium

```
rm(fb_posts)
```

La deuxième période :

```
load(paste0(data_folder, "dt_fb_posts_all_after_preprocessing.RData"))

dt <- fb_posts[day_of_year %in% c(115:135) & is_science == TRUE & since_COVID == TRUE &
  high_engagement == TRUE][
  order(-interaction), list(interaction, day = jour %>% strptime(format = "%B %d"),
    titre, message, description, media, lien)
]

if(outputFormat == "html"){
  dt %>% kable() %>% kable_styling() %>%
    scroll_box(width = "100%", height = "200px")
} else{
  dt %>% kable( booktabs = T, longtable = TRUE) %>%
    kable_styling(full_width = F) %>%
    column_spec(3:7, width = "7em")
}
```

}

interaction	day	titre	message	description	media	lien
9577	mai 03	Des «frelons meurtriers» trouvés en Amérique du Nord inquiètent des chercheurs	Ils sont capables d'exterminer des colonies d'abeilles entières en quelques heures et de transpercer l'équipement de protection d'apiculteurs.	Ces frelons tuent jusqu'à 50 personnes annuellement au Japon.	Le Journal de Montréal	http://www.journaldemontreal.com/actualites/monde/2020/05/03/frelons-meurtriers-trouves-en-amerique-du-nord-inquietent-des-chercheurs-1
7197	mai 07	Le coronavirus est-il aussi une MTS? Des traces dans du sperme	Des traces du coronavirus ont été retrouvées dans du sperme	Une équipe de chercheurs chinois a constaté que le sperme de plusieurs personnes ayant contracté la COVID-19 était positif au virus.	TVA Nouvelles	https://www.tvanouvelles.ca/2020/05/07/le-coronavirus-est-il-aussi-une-mst-des-traces-dans-du-sperme
6063	mai 05	Une nouvelle souche du virus plus contagieuse: des chercheurs sonnent l'alerte	La deuxième souche aurait frappé à partir de la mi-mars dans l'État de New York, elle serait beaucoup plus contagieuse.	Des scientifiques ont identifié une nouvelle souche du coronavirus qui serait devenue dominante dans le monde.	TVA Nouvelles	https://www.tvanouvelles.ca/2020/05/05/une-nouvelle-souche-du-virus-plus-contagieuse
6035	mai 05	Une nouvelle souche du virus plus contagieuse: des chercheurs sonnent l'alerte	Des scientifiques ont identifié une nouvelle souche du coronavirus qui serait devenue dominante dans le monde et qui semble être plus contagieuse que les versions qui se sont propagées aux premiers jours de la pandémie de COVID-19.	Des scientifiques ont identifié une nouvelle souche du coronavirus qui serait devenue dominante dans le monde.	Le Journal de Montréal	https://www.journaldemontreal.com/actualites/monde/2020/05/05/une-nouvelle-souche-du-virus-plus-contagieuse-des-chercheurs-sonnent-l-alerte

5890	avril 30	Une réouverture des écoles ferait «immédiatement» repartir la pandémie	Selon un chercheur qui s'exprimait lors de la conférence hebdomadaire de l'Institut supérieur de la santé	Une réouverture des écoles ferait «immédiatement» repartir à la hausse la pandémie de coronavirus.	TVA Nouvelles	https://www.tvanouvelles.ca/2020/04/30/une-reouverture-des-ecoles-ferait-immEDIATEMENT-repartir-la-pandemie-1
5053	mai 02	Des «frelons meurtriers» trouvés en Amérique du Nord inquiètent des chercheurs	Ils mesurent jusqu'à deux pouces de longueur.	Ces frelons tuent jusqu'à 50 personnes annuellement au Japon. «Nous avons une fenêtre pour les empêcher de s'installer.»	TVA Nouvelles	http://www.tvanouvelles.ca/2020/05/02/frelons-meurtriers-trouves-en-amerique-du-nord-inquietent-des-chercheurs
5038	avril 27	Incertitude et gérants d'estrade	«Ce qui frappe le plus dans ce bruit ambiant de solutions contradictoires — le retour du "y a qu'à faire ceci, y a qu'à faire cela" — est l'absence totale de prise en compte de la temporalité des prises de décision en situation d'incertitude et même d'ignorance.» Un texte d'opinion d'Yves Gingras, directeur scientifique de l'Observatoire des sciences et des technologies de l'UQAM. — via Le Devoir - Opinion	La situation est totalement inusitée et les connaissances prennent du temps à se construire.	Le Devoir	https://www.ledevoir.com/opinion/577763/incertitude-et-gerants-d-estrade?utm_medium=Social

4985	mai 06	Assassinat d'un chercheur sur le point de faire «des découvertes importantes»	Le meurtre du chercheur pourrait-il être dû aux recherches qu'il menait sur la COVID-19?	Un professeur de l'Université de Pittsburgh qui menait des recherches prometteuses sur la COVID-19 a été assassiné ce week-end.	TVA Nouvelles	https://www.tvanouvelles.ca/2020/04/24/01-5270714-des-propos-de-trump-declenchent-un-tolle-chez-les-scientifiques
4941	avril 24	Des propos de Trump déclenchent un tollé chez les scientifiques	Une « injection » de « désinfectant » pour combattre le coronavirus ? Des déclarations de Donald Trump jeudi ont consterné la communauté scientifique et de nombreux spécialistes ont accusé le président américain d'être irresponsable en faisant cette suggestion « dangereuse ».	(Paris) Une « injection » de « désinfectant » pour combattre le coronavirus ? Des déclarations de Donald Trump jeudi ont consterné la communauté scientifique...	La Presse	https://www.lapresse.ca/unis/202004/24/01-5270714-des-propos-de-trump-declenchent-un-tolle-chez-les-scientifiques.php?utm_source=twitter&utm_medium=social
3917	mai 06	COVID-19: assassinat d'un chercheur sur le point de faire «des découvertes importantes»	Le meurtre du chercheur pourrait-il être dû aux recherches qu'il menait sur la COVID-19?	Un professeur de l'Université de Pittsburgh qui menait des recherches prometteuses sur la COVID-19 a été assassiné ce week-end.	Le Journal de Montréal	https://www.journaldelemonde.com/2020/05/06/covid-19-assassinat-dun-chercheur-sur-le-point-de-faire-des-decouvertes-importantes
3889	mai 02	La pandémie de coronavirus devrait durer au moins 18 mois	Selon les prédictions de chercheurs universitaires dans un nouveau rapport	La pandémie de coronavirus devrait durer de 18 à 24 mois, ont prédit des chercheurs dans un rapport universitaire publié jeudi.	TVA Nouvelles	http://www.tvanouvelles.ca/2020/05/02/01-5270714-pandemie-de-coronavirus-devrait-durer-au-moins-18-mois

3871	mai 08	La COVID-19, aussi une maladie du cœur	«On apprend que la COVID-19 peut causer des arythmies cardiaques, elle peut causer des infarctus. elle peut même attaquer le muscle cardiaque»	Les médecins et chercheurs se rendent compte de corollaires entre les maladies cardio-vasculaires et les patients les plus gravement atteints.	TVA Nouvelles	https://www.tvanouvelles.ca/3871/covid-19-aussi-une-maladie-du-cœur
3050	avril 30	Syndrome inflammatoire chez l'enfant: un spécialiste fait le point	Une équipe de chercheurs de l'Hôpital Sainte-Justine travaille sur une étude	Des enfants qui auraient été en contact avec le coronavirus ont développé une maladie inflammatoire. Des cas ont été signalés en Europe et ici.	TVA Nouvelles	https://www.tvanouvelles.ca/3050/inflammatoire-chef-l'enfant-un-specialiste-fait-le-point
2705	avril 28	Les régions devront rester fermées, croient des chercheurs	Selon un «plan de déconfinement parfait» préparé par une équipe de chercheurs québécois	Alors que la Santé publique évoque la levée des barrages, un «plan de déconfinement parfait» recommande au contraire de les maintenir.	TVA Nouvelles	https://www.tvanouvelles.ca/2705/regions-devront-rester-fermees-croient-des-chercheurs
2657	mai 09	La vitamine D à la rescousse pour mieux combattre la COVID-19	Les chercheurs responsables de l'étude se sont penchés sur le cas de 20 pays européens	Le niveau de vitamine D absorbé par une personne a un effet sur sa capacité à mieux combattre la COVID-19, démontre une étude.	TVA Nouvelles	http://www.tvanouvelles.ca/2657/19-la-vitamine-d-a-la-rescousse-1
2508	mai 04	Identification d'un anticorps monoclonal prometteur	Capable en laboratoire de neutraliser le virus responsable de la COVID-19	Un anticorps monoclonal capable en laboratoire de neutraliser le virus SARS-CoV-2 a été identifié par une équipe de chercheurs néerlandais.	TVA Nouvelles	https://www.tvanouvelles.ca/2508/dun-anticorps-monoclonal-prometteur

2465	avril 24	Combattre le coronavirus avec du «désinfectant»? Trump accusé d'être «irresponsable»	«Je vois que le désinfectant l'assomme (le coronavirus) en une minute», a déclaré Donald Trump lors de son point de presse quotidien.	Des déclarations de Donald Trump jeudi ont consterné la communauté scientifique.	Le Journal de Montréal	https://www.journaldemontreal.com/le-coronavirus-avec-du-desinfectant-trump-accuse-detre-irresponsable-1
2408	mai 05	Découverte d'une concentration record de pollution plastique dans les fonds marins	Selon une étude menée par des chercheurs de l'Université McGill sur le tronçon du Saint-Laurent entre Montréal et Québec, le fleuve serait parmi les pires cours d'eau de la planète en matière de pollution par les microplastiques.	Un équivalent aux «continents de plastique» a été trouvée en débris minuscules dans le fond des mers.	Le Devoir	https://www.ledevoir.com/actualites/monde/2408-dune-concentration-record-de-pollution-plastique-dans-les-fonds-marins?utm_medium=S
2346	avril 27	Le Québec a réussi à éviter «une catastrophe», assure une experte	La courbe s'aplatit «depuis plusieurs semaines», selon cette chercheuse	La courbe s'aplatit «depuis plusieurs semaines», selon cette chercheuse	TVA Nouvelles	https://www.tvanouvelles.ca/2346-le-quebec-a-reussi-a-eviter-une-catastrophe-assure-une-experte
2269	mai 05	Une nouvelle souche du virus plus contagieuse: des chercheurs sonnent l'alerte	Cette souche serait devenue dominante dans le monde et semble être plus contagieuse que les versions qui se sont propagées auparavant...	Des scientifiques ont identifié une nouvelle souche du coronavirus qui serait devenue dominante dans le monde.	Le Journal de Québec	https://www.journaldequebec.com/2269-une-nouvelle-souche-du-virus-plus-contagieuse-des-chercheurs-sonnent-l-alerte
2228	mai 07	Un lama pourrait être la clé pour vaincre la COVID	Les chercheurs étudiaient déjà les anticorps des lamas pour développer un traitement contre le SRAS	Des scientifiques se tournent vers une source à tout le moins obscure pour trouver un remède contre la COVID-19 : Winter le lama.	TVA Nouvelles	https://www.tvanouvelles.ca/2228-un-lama-pourrait-etre-la-cle-pour-vaincre-la-covid

2214	avril 24	Trump, coronavirus et désinfectant: confusion et consternation	Le fabricant du désinfectant Lysol s'est senti obligé de faire une mise au point écrite «en raison des spéculations intenses et de l'activité sur les réseaux sociaux»	Une «injection» de «désinfectant» pour combattre le coronavirus? Des déclarations de Donald Trump jeudi ont consterné la communauté scientifique.	TVA Nouvelles	https://www.tvanouvelles.ca/2020/04/le-coronavirus-avec-du-desinfectant-trump-accuse-detre-irresponsable
1895	mai 06	COVID-19: assassinat d'un chercheur sur le point de faire «des découvertes importantes»	Un professeur de l'Université de Pittsburgh qui menait des recherches prometteuses sur la COVID-19 a été assassiné ce week-end...	Un professeur de l'Université de Pittsburgh qui menait des recherches prometteuses sur la COVID-19 a été assassiné ce week-end.	Le Journal de Québec	https://www.journaldequebec.com/2020/05/19-assassinat-dun-chercheur-sur-le-point-de-faire-des-decouvertes-importantes
1765	mai 07	La stratégie de dépistage du Québec vivement critiquée par une conseillère de Trudeau Coronavirus	« Il y a une différence entre le discours et ce qu'on entend sur le terrain », déplore la conseillère scientifique en chef du Canada.	« Il y a une différence entre le discours et ce qu'on entend sur le terrain », déplore Mona Nemer.	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/1700000/quebec-trudeau-legault-nemer-strategie-coronavirus
1720	avril 28	COVID-19: les régions devront rester fermées, croient des chercheurs	Une équipe de chercheurs québécois recommande de maintenir les barrages empêchant de se rendre vers les régions peu affectées par la COVID-19.	Un «plan de déconfinement parfait» préparé par une équipe de chercheurs recommande de maintenir les barrages empêchant de se rendre vers les régions.	Le Journal de Québec	https://www.journaldequebec.com/2020/04/19-les-regions-devront-rester-fermees-croient-des-chercheurs
1597	mai 06	5G et COVID: les théories du complot échauffent les esprits	Des tours de télécommunication ont successivement été ciblées par des incendies criminels à Laval, Prévost et Piedmont au cours des derniers jours.	«Les auteurs de théories arrivent à créer une histoire qui ne tient pas la route d'un point de vue scientifique.»	TVA Nouvelles	https://www.tvanouvelles.ca/2020/05/et-covid-les-theories-du-complot-echauffent-les-esprits

1515	mai 08	Plandemic , un documentaire truffé de fausses affirmations Coronavirus	[Décrypteurs] La chercheuse mise de l'avant dans le documentaire, Judy Mikovits, fait un grand nombre de fausses affirmations. En voici quelques-unes décortiquées.	La chercheuse mise de l'avant dans le documentaire a un passé controversé.	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/1701plandemic-judy-mikovits-fauci-covid-19-coronavirus-faux
1358	mai 02	On écoute les experts pour la COVID-19, pourquoi pas pour les changements climatiques? Coronavirus	«Quand il y a vraiment la perception d'un danger imminent, alors là les décideurs prennent des décisions qu'on n'aurait jamais pu imaginer.»	Les scientifiques n'ont jamais autant eu l'oreille des décideurs. Pourquoi n'est-ce pas le cas pour le climat?	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/169819-coronavirus-experts-scientifiques-changements-climatiques-climat-crise
1282	avril 30	«Si t'as pas compris ça, t'es pas fort en science!»	Mario Dumont se lance dans un petit calcul mathématique afin d'appuyer son propos		TVA Nouvelles	https://www.tvanouvelles.ca/tas-pas-compris-ca-tes-pas-fort-en-science?fbclid=IwAR0ZvYtLtJIQYQGwN5UgeyYQj_eGVp_TJEs0z8yMbV

1172	mai 08	Arruda s'en prend aux «gérants d'estrade»	Québec se défend d'improviser quant à son plan de déconfinement, au lendemain de vives critiques de la part de la conseillère scientifique de Justin Trudeau. Le Dr Horacio Arruda en a profité pour dénoncer les « gérants d'estrade » et annoncer une volte-face concernant les grands-parents, qui pourront garder leurs petits-enfants sous certaines conditions.	Québec se défend d'improviser quant à son plan de déconfinement, au lendemain de vives critiques de la part de la conseillère scientifique de Justin Trudeau...	La Presse	https://www.lapresse.ca/19/202005/08/01-5272705-arruda-sen-prend-aux-gerants-destrade.php?utm_source=
1169	mai 13	Une fillette de Matane atteinte de la maladie de Kawasaki Au coeur du monde	L'important c'est de reconnaître les symptômes de la maladie de Kawasaki rapidement. Ce fut le cas en ce qui concerne la petite Alyssa.	La jeune matanaise Alyssa, 6 ans, a été diagnostiquée de la maladie de Kawasaki récemment. Elle a commencé à avoir des symptômes à la mi-février. On entend beaucoup plus parler de cette maladie ces jours-ci puisque les chercheurs tentent de voir si les nouveaux cas ont un lien avec la COVID...	ICI Gaspésie- Les-Îles	https://ici.radio-canada.ca/premiere/emission/coeur-du-monde/segments/entrevue/kawasaki-matane

1131	avril 27	Il n'est « pas prudent » d'anticiper l'immunité collective, dit Trudeau Coronavirus	« On ne sait pas encore si quelqu'un qui a eu la COVID-19 est protégé de la rattraper quelques mois plus tard; la science n'est pas [concluante] là-dessus. »	L'immunité collective n'est « pas prouvée », admet peu après le premier ministre du Québec, François Legault.	Radio-Canada Information	https://ici.radio-canada.ca/nouvelle/1697-trudeau-pandemie-coronavirus-mesures-gouvernementales-covid-19
1049	mai 09	Le bouillant Dr Arruda	«Ce scientifique au tempérament vif a conquis les Québécois en réussissant à leur imposer un confinement dont ils ne mesuraient pas à l'évidence les contraintes.» La chronique de Denise Bombardier.	En quelques semaines, le directeur de la santé publique du Québec est passé de l'ombre à la lumière.	Le Journal de Montréal	https://www.journaldemontreal.com/bouillant-dr-arruda
1024	avril 28	COVID-19: les régions devront rester fermées, croient des chercheurs	Alors que la Santé publique évoque la levée des barrages, un «plan de déconfinement parfait» préparé par une équipe de chercheurs québécois recommande au contraire de les maintenir.	Un «plan de déconfinement parfait» préparé par une équipe de chercheurs recommande de maintenir les barrages empêchant de se rendre vers les régions.	Le Journal de Montréal	https://www.journaldemontreal.com/19-les-regions-devront-rester-fermees-croient-des-chercheurs
1013	mai 13	«Arruda ne doit pas perdre sa spontanéité, son naturel»	Pourquoi la population s'est-elle profondément attachée à ce scientifique hors norme et plutôt coloré? Un expert en communications publiques explique :	Le directeur de la Santé publique du Québec, Horacio Arruda, était inconnu de la population il y a 6 mois de cela.	Le Journal de Montréal	https://www.journaldemontreal.com/ne-doit-pas-perdre-sa-spontaneite-son-naturel-1

986	avril 24	Combattre le coronavirus avec du «désinfectant»? Trump accusé d'être «irresponsable»	Les propos du président américain, qui a évoqué une «injection» de «désinfectant» pour combattre le coronavirus, ont déclenché un tollé chez les scientifiques.		Le Devoir	https://www.ledevoir.com/unis/577673/combattre-le-coronavirus-avec-du-desinfectant-trump-accuse-d-etre-irresponsable?utm_medium=
970	mai 04	COVID-19: identification d'un anticorps monoclonal prometteur	Un anticorps monoclonal capable en laboratoire de neutraliser le virus SARS-CoV-2, responsable de la COVID-19, a été identifié par une équipe de chercheurs néerlandais.	Un anticorps monoclonal capable en laboratoire de neutraliser le virus SARS-CoV-2 a été identifié par une équipe de chercheurs néerlandais.	Le Journal de Montréal	https://www.journaldenouvelles.com/19-identification-dun-anticorps-monoclonal-prometteur
962	mai 02	Écoutons aussi les spécialistes de l'éducation. . .	OPINION / L'auteure, Nancy Goyette, est professeure au Département des sciences de l'éducation de l'Université du Québec à Trois-Rivières.	OPINION / L'auteure, Nancy Goyette, est professeure au Département des sciences de l'éducation de l'Université du Québec à Trois-Rivières.	Le Nouvelliste	https://www.lenouvelliste.com/aussi-les-specialistes-de-leducation-80db4f1e1c37f215de7cf4b
909	mai 06	Les masques artisanaux les plus efficaces	Pour bien choisir quels tissus utiliser pour fabriquer un masque « fait maison », des chercheurs de l'Université de Chicago ont analysé la capacité de différentes matières à retenir les postillons et petits aérosols.	Pour bien choisir le tissu à utiliser, des chercheurs ont analysé la capacité de différentes matières à retenir les postillons et petits aérosols.	Le Journal de Montréal	https://www.journaldenouvelles.com/masques-artisanaux-les-plus-efficaces

908	avril 27		Existe-t-il vraiment un remède miracle pour échapper à la COVID-19 ? L'équipe de L'épicerie s'est penchée sur la question avec l'aide de scientifiques.		Radio-Canada Information	https://www.facebook.com/
892	avril 25	Un appel à la prudence sur l'immunité collective	L'immunité collective sur laquelle mise entre autres la santé publique du Québec est en quelque sorte un pari que prend le gouvernement provincial puisque scientifiquement parlant, rien n'est encore confirmé.	En raison des comportements particuliers de la COVID-19, l'OMS a voulu faire un appel à la prudence.	Le Journal de Montréal	http://www.journaldemontreal.com/appel-a-la-prudence-sur-limmunit%C3%A9-collective-1
882	avril 24	Les scientifiques consternés par les propos de Trump sur le désinfectant	Non, ingérer du désinfectant ne combat pas la COVID-19... C'est juste dangereux.		HuffPost Québec	http://huffp.st/XGoiJER
869	avril 25	COVID-19: Ottawa ne mise pas sur l'immunité collective	Justin Trudeau rappelle « qu'il n'y a aucune science encore qui dit que si on a attrapé la COVID-19 une fois, on ne peut l'attraper une deuxième fois ».	Justin Trudeau a averti qu'il ne faut pas, pour l'instant, considérer l'immunité collective comme une solution pour endiguer la pandémie de COVID-19.	Le Journal de Montréal	http://www.journaldemontreal.com/19-ottawa-ne-mise-pas-sur-limmunit%C3%A9-collective

799	avril 25		Des chercheurs de l'Université de Sherbrooke ont créé en quelques jours un des ingrédients indispensables aux tests de COVID-19. Il s'agit d'un amplificateur qui permet de déceler le virus. Son nom : le Sherby-20.		ICI Estrie	https://www.facebook.com/
727	avril 30	Gare à l'infection volontaire	Tentés d'attraper volontairement la COVID-19 pour sortir au plus vite du confinement ? Minute, mettent en garde les scientifiques. L'idée est probablement plus dangereuse qu'elle n'y paraît.	Tentés d'attraper volontairement la COVID-19 pour sortir au plus vite du confinement? Minute, mettent en garde les scientifiques. L'idée est probablement ...	La Presse	https://www.lapresse.ca/19/202004/29/01-5271486-gare-a-linfection-volontaire.php?utm_source=
592	avril 28	Des experts croient qu'il faut maintenir les contrôles routiers - Les nouvelles - CIHO FM 96,3	Une équipe de chercheurs québécois recommande à la Santé publique de maintenir les barrages routiers aux frontières des régions moins affectées par la COVID-19, dont Charlevoix.	Alors que la Santé publique évoque la levée des contrôles routiers aux frontières des régions peu affectées par la COVID-19, dont Charlevoix, un comité de chercheurs québécois croit au contraire qu'ils doivent être maintenus.	FM Charlevoix - CIHO 96,3	http://www.cihofm.com/experts-croient-qu-il-faut-maintenir-les-controles-routiers-2020-04-28-06-27

562	mai 08		Les piscines seront-elles ouvertes au cours de l'été? Le gouvernement du Québec se penche sur la question, tandis que les scientifiques et les gestionnaires de bassins répondent par l'affirmative. rc.ca/RWDHPY		Radio-Canada Information	https://www.facebook.com/
557	avril 24	Combattre le coronavirus avec du «désinfectant»? Trump accusé d'être «irresponsable»	Une «injection» de «désinfectant» pour combattre la COVID-19?	Des déclarations de Donald Trump jeudi ont consterné la communauté scientifique.	Le Journal de Québec	https://www.journaldequebec.com/le-coronavirus-avec-du-desinfectant-trump-accuse-detre-irresponsable-1
546	mai 10	Maladie liée à la COVID-19 chez les enfants: le maire de New York s'inquiète	Le maire de New York a fait part de sa « grave inquiétude » concernant une augmentation de cas d'un syndrome touchant les enfants que les scientifiques pensent être lié au coronavirus.	Le maire de New York Bill de Blasio a fait part de sa «grave inquiétude» concernant une augmentation de cas d'un syndrome touchant les enfants.	Le Journal de Montréal	https://www.journaldequebec.com/liee-a-la-covid-19-chez-les-enfants-le-maire-de-new-york-sinquiete-1

533	mai 09		[Décrypteurs] Près d'un Canadien sur dix croirait aux théories du complot entourant la pandémie de COVID-19, selon les résultats préliminaires d'une recherche menée par une équipe de chercheurs de l'Université de Sherbrooke. rc.ca/RWD2k2		Radio-Canada Information	https://www.facebook.com/
528	avril 24	Des propos de Trump soulèvent la controverse chez les scientifiques	Une «injection» de «désinfectant» pour combattre le coronavirus? Des déclarations de Donald Trump jeudi ont consterné la communauté scientifique.	Une «injection» de «désinfectant» pour combattre le coronavirus? Des déclarations de Donald Trump jeudi ont consterné la communauté scientifique et de nombreux spécialistes ont accusé le président américain d'être irresponsable en faisant cette suggestion «dangereuse» «Je vois que ...	Journal Métro	https://journalmetro.com/propos-de-trump-soulevent-la-controverse-chez-les-scientifiques/
527	mai 13	«Arruda ne doit pas perdre sa spontanéité, son naturel»	Pourquoi la population s'est-elle profondément attachée à ce scientifique hors norme et plutôt coloré?	Le directeur de la Santé publique du Québec, Horacio Arruda, était inconnu de la population il y a 6 mois de cela.	Le Journal de Québec	https://www.journaldequebec.com/ne-doit-pas-perdre-sa-spontaneite-son-naturel-1

515	avril 27	Le Québec a réussi à éviter «une catastrophe», assure une experte	La courbe s'aplatit «depuis plusieurs semaines».	La courbe s'aplatit «depuis plusieurs semaines», selon cette chercheuse	Le Journal de Montréal	https://www.journaldemontreal.com/quebec-a-reussi-a-eviter-une-catastrophe-assure-une-experte
506	avril 28	Les régions devront rester fermées, croient des chercheurs	Même si l'interdiction des déplacements interrégionaux a été l'une des dernières mesures mises en place par le gouvernement provincial pour limiter la propagation du coronavirus, elle devrait être l'une des dernières à être levée, estiment-ils.	Alors que la Santé publique évoque la levée des barrages, un «plan de déconfinement parfait» recommande au contraire de les maintenir.	TVA Est-du-Québec	https://www.tvanouvelles.ca/regions-devront-rester-fermees-croient-des-chercheurs
454	avril 27	Le maintien à domicile des aînés serait bénéfique durant une pandémie	On aurait tout intérêt, en période de pandémie, à favoriser le maintien des aînés à domicile, surtout quand on constate les ravages causés par le coronavirus dans les CHSLD, croit un chercheur du Centre interuniversitaire de recherche en analyse des organisations (CIRANO).	(Montréal) On aurait tout intérêt, en période de pandémie, à favoriser le maintien des aînés à domicile, surtout quand on constate les ravages causés par l...	La Presse	https://www.lapresse.ca/19/202004/27/01-5271084-le-maintien-a-domicile-des-aines-serait-benefique-durant-une-pandemie.php?utm_source=twitter&utm_medium=social

428	avril 26	Un déconfinement graduel reste souhaitable	L'important est de continuer à tester la population pour mesurer la présence de la COVID-19 et prévenir une deuxième vague d'infections.	L'absence de preuve scientifique concernant l'immunité supposée des personnes guéries de la COVID-19 ne remet pas en cause le déconfinement.	Le Journal de Montréal	https://www.journaldemontreal.com/deconfinement-graduel-reste-souhaitable
403	mai 03	Surveillés et dociles?	«Aux mesures classiques de quarantaine succède la mise en place de dispositifs de contrôle inédits, qu'on aurait aisément pu qualifier de dystopiques il y a encore quelques semaines.» Un texte d'opinion d'Alexandre Klein, de l'unité de recherche sur l'histoire du nursing de l'École des sciences infirmières de l'Université d'Ottawa. — via Le Devoir - Opinion	Les décisions que nous prenons actuellement sont loin d'être anecdotiques ou simplement temporaires.	Le Devoir	https://www.ledevoir.com/societe/actualites/surveillance-et-dociles?utm_medium=SocialMedia
392	mai 11	Un lama pourrait fournir la clé du traitement de la COVID-19	Des chercheurs ont isolé un anticorps du lama qui pourrait grandement contribuer à la lutte contre la maladie.		HuffPost Québec	http://huffp.st/RrtSxnk
389	mai 12	Un vaccin à coup sûr?	Les scientifiques ne peuvent garantir qu'un vaccin contre le coronavirus verra le jour.		Le Devoir	https://www.ledevoir.com/societe/actualites/vaccin-a-coup-sur?utm_medium=SocialMedia

377	avril 24	Trump, coronavirus et désinfectant: confusion et consternation	Ses propos confus sur de possibles injections de désinfectant pour lutter contre le coronavirus ont suscité la stupéfaction.	Des déclarations de Donald Trump jeudi ont consterné la communauté scientifique.	Le Journal de Montréal	https://www.journaldemontreal.com/le-coronavirus-avec-du-desinfectant-trump-accuse-detre-irresponsable-1
375	mai 09	Un médicament québécois contre la COVID-19 au printemps 2021?	Il est entendu dans la communauté scientifique que vaccin pourrait ne voir le jour que dans 12 à 18 mois.	Si le gouvernement subventionne la recherche de Tetra Bio Pharma à hauteur de 14 millions \$, le coronavirus pourrait avoir un ennemi de sa taille.	Le Journal de Montréal	https://www.journaldemontreal.com/medicament-canadien-contre-la-covid-19-au-printemps-2021-1
358	avril 25	Coronavirus : rien ne prouve que les personnes guéries sont immunisées, avertit l'OMS Coronavirus	«Il n'existe aucune preuve scientifique que les personnes qui ont vaincu la COVID-19 et ont des anticorps sont prémunies contre une deuxième infection», écrit l'OMS dans une note scientifique.	L'OMS déclare qu'il n'y a « aucune preuve » que les personnes guéries du coronavirus sont immunisées.	ICI Mauricie Centre-du-Québec	https://ici.radio-canada.ca/nouvelle/1697-infection-maladie-contagion-epidemie

338	avril 24	Le vaccin contre la COVID-19 pourrait venir de Laval Courrier Laval	Dr René Roy et son équipe de chercheurs croient possible de développer dans leurs laboratoires de Laval un vaccin dans les 12 à 15 prochains mois.	«D'ici trois mois, avec les ressources appropriées, nous serons en mesure d'évaluer l'efficacité de notre vaccin et de déterminer s'il est en mesure de détruire le virus de la COVID-19», affirme Dany Valiquette qui préside aux destinées d'une biopharma en incubation au Centre qué...	Courrier Laval	https://courrierlaval.com/vaccin-contre-la-covid-19-pourrait-venir-de-laval/
-----	----------	---	--	---	----------------	---

```
rm(fb_posts)
```

Quelle science?

Maintenant, pour étudier les changements qualitatifs dans la représentation médiatique de la science, nous faisons diverses analyses ci-dessous.

Verbes spécifiques à la période COVID

Nous chargeons un objet produit lors du prétraitement (voir `scripts/04-Morphosyntactic_analysis.R`). Il contient l'analyse morphosyntaxique des paragraphes liés à la COVID. Le prochain graphique distingue les 30 verbes les plus caractéristiques de 2020 en comparaison des trois années précédentes, pour les paragraphes relatifs à la science (voir `figure6.jpg`).

```
load(paste0(data_folder, "dt_tagged_tokens_science.RData"))

dt_lemmas <- dt_tagged_tokens_science[upos == "VERB"]

dt_chi_sqr <- dt_lemmas[lemma != "<unknown>", .N, by = .(lemma, corpus)]
dt_chi_sqr <- merge(dt_chi_sqr[corpus == 1], dt_chi_sqr[corpus == 2],
  by = "lemma", all = TRUE)
dt_chi_sqr[is.na(corpus.x), `:=`(corpus.x = 1, N.x = 0)]
dt_chi_sqr[is.na(corpus.y), `:=`(corpus.y = 2, N.y = 0)]
setnames(dt_chi_sqr, c("corpus.x", "corpus.y", "N.x", "N.y"),
  c("corpus_1", "corpus_2", "Lemma_count_1", "Lemma_count_2"))
#Éviter les divisions par zéro
dt_chi_sqr[, `:=`(Lemma_count_1 = Lemma_count_1 + 1, Lemma_count_2 = Lemma_count_2 + 1)]
```

```

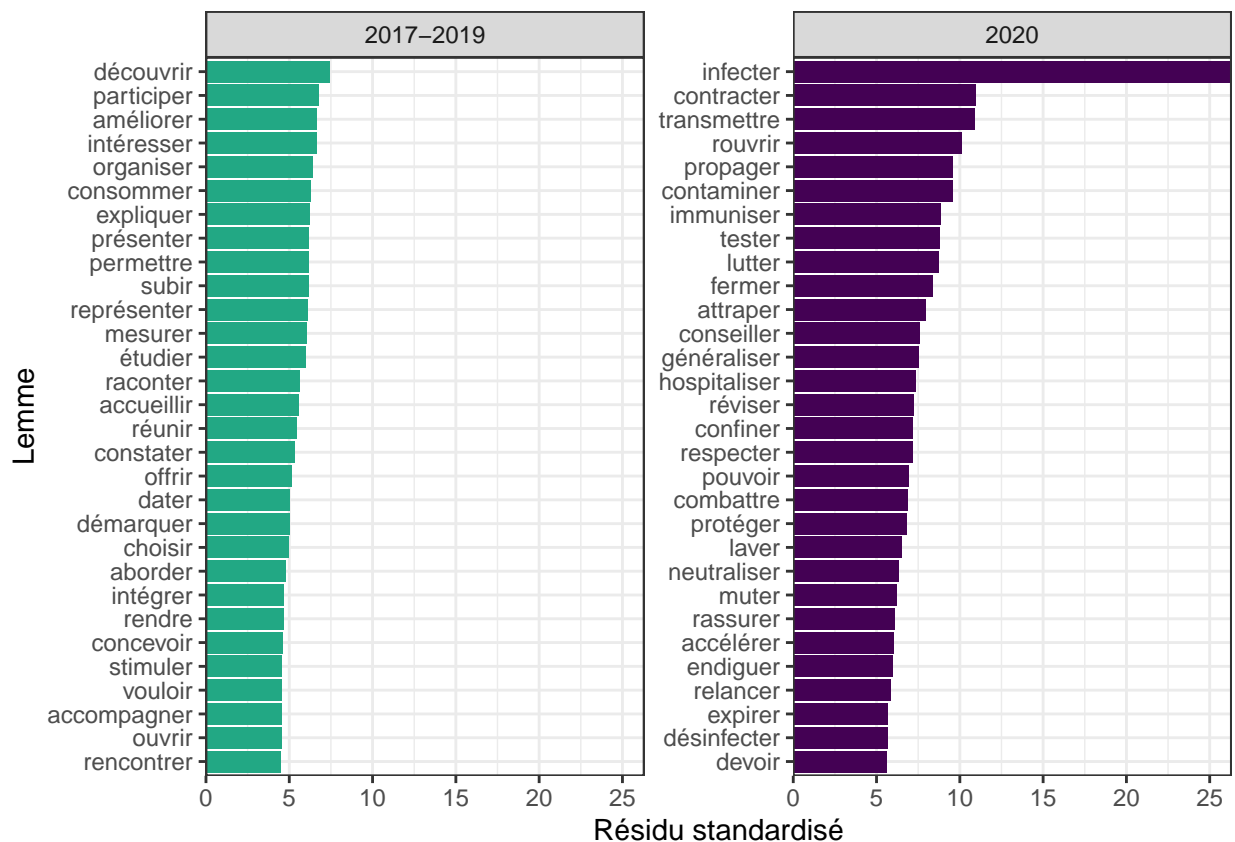
res_chi <- chisq.test(dt_chi_sqr[,list(Lemma_count_1, Lemma_count_2)])

dt_chi_sqr[, residual := res_chi$stdres[,2]]
setorder(dt_chi_sqr, residual)
lemmas_that_are_not_verbs <- c("quir", "unrir", "hydroxychloroquiner", "infecté",
                              "influenzer", "éternir", "immunité", "épidémie",
                              "ne", "professeurer", "âgé", "professeur")
dt_chi_sqr <- dt_chi_sqr[!lemma %in% lemmas_that_are_not_verbs]

nb_words <- 30
dt_plot <- rbind(dt_chi_sqr[, list(lemma, residual)] %>% head(nb_words),
                dt_chi_sqr[, list(lemma, residual)] %>% tail(nb_words))

dt_plot[residual > 0, corpus_title := "2020"][is.na(corpus_title),
                                              corpus_title := "2017-2019"]
dt_plot[corpus_title == "2017-2019", residual := -residual]
ggplot(dt_plot, aes(x = reorder(lemma, residual), y = residual,
                    fill = corpus_title), col = corpus_title) +
  geom_bar(stat = "identity") + ylab("Résidu standardisé") + xlab("Lemme") +
  facet_wrap(~corpus_title, scales = "free_y") +
  scale_x_reordered() +
  scale_y_continuous(expand=c(0,0)) + guides(fill="none") +
  theme_bw() + coord_flip() + scale_fill_viridis(discrete=TRUE, end = .6, direction = -1)

```



```

ggsave(file=paste0(fig_dir, "figure6.jpg") )

```

```
rm(dt_tagged_tokens_science)
```

Moteur de recherche sémantique

Les verbes caractéristiques ci-dessus (figure6.jpg) peuvent être difficiles à interpréter. Nous construisons donc un moteur de recherche pour trouver les paragraphes les plus caractéristiques de l'usage général d'un ou de plusieurs de ces verbes.

```
search_most_similar_paragraph <- function(request_words, corpus_, nombre_article = 5)
{
  dt_lemmas <- dt_tagged_tokens_science[upos %in% c('VERB', 'NOUN', 'ADJ', 'ADV', 'PROPN') &
                                         corpus == corpus_,
                                         list(lemma, art_ID, doc_id)]

  # La fonction itoken de text2vec veut ce format, c'est-à-dire,
  # une liste de token (ou lemme) par document
  dt_lemmas <- dt_lemmas[,.(lemma = list(lemma)), by =doc_id]

  #Structure de donnée qui permet de travailler avec nos données sans faire de copies
  iterateur <- text2vec::itoken(dt_lemmas$lemma,
                                ids = dt_lemmas$doc_id,
                                progressBar = FALSE)

  vocabulaire <- text2vec::create_vocabulary(iterateur,
                                             ngram=c(ngram_min = 1,
                                                       ngram_max = 1))

  vocabulaire <- text2vec::prune_vocabulary(vocabulaire,
                                           term_count_min = 2,
                                           term_count_max = Inf,
                                           doc_proportion_min = 0.0001,
                                           doc_proportion_max = 0.5,
                                           doc_count_min = 1,
                                           doc_count_max = Inf,
                                           vocab_term_max = Inf)

  setDT(vocabulaire)

  #Permet de vectoriser nos documents. Le vectoriseur est utilisé
  # par create_dtm (voir la doc pertinente au besoin)
  vectorizer <- text2vec::vocab_vectorizer(vocabulaire)

  dtm <- text2vec::create_dtm(iterateur,
                              vectorizer,
                              type = "dgCMatrix")

  #On transforme le dtm pour que les cellules contiennent les valeurs tf-idf
  dtm_tfidf_initialisation <- text2vec::TfIdf$new()
  dtm_tfidf <- dtm_tfidf_initialisation$fit_transform(dtm)
  rm(dtm_tfidf_initialisation)

  n_dimensions_M_originale <- ncol(dtm)
```



```

#Initiatilisation vecteur de requête
vecteurRequete <- Matrix(data = 0, nrow = 1, ncol = n_dimensions_M_originale)

#Les colonnes sont nommées avec les mots de notre dtm,
# ce qui facilite la création d'une requête
colnames(vecteurRequete) <- colnames(dtm)

rownames(vecteurRequete) <- "requete"

vecteurRequete[1, request_words] <- 1

similarite_cosinus <- proxyC::simil(x = dtm_tfidf,
  y = vecteurRequete, method = "cosine") %>%
  as.data.frame.matrix()

colnames(similarite_cosinus) <- c("Cosine")
similarite_cosinus['ID'] <- rownames(dtm_tfidf)

setDT(similarite_cosinus)
setorder(similarite_cosinus, -Cosine)
setnames(similarite_cosinus, "ID", "paragraphe_ID")

#On va chercher les paragraphes dans dt_science et on trie par le score cosinus
resultats_recherche_semantique <- merge(similarite_cosinus[Cosine > 0.1], dt_science,
  by = "paragraphe_ID", all.x = TRUE, all.y = FALSE)

resultats_recherche_semantique[order(-Cosine),
  list(Cosine, paragraphe =
    str_replace_all(paragraphe, "\n", " "))] %>%
  unique() %>% head(nombre_article)
}

```

Verbes de 2017-2019

Avec ce moteur de recherche, nous sortons les paragraphes caractéristiques pour les 10 verbes les plus caractéristiques de 2017-2019 (voir figure6.jpg):

```

#Ces deux tables ont été produites dans 04-Morphosyntactic_analysis.R
load(paste0(data_folder, "dt_tagged_tokens_science.RData"))
load(paste0(data_folder, "dt_science.RData"))

corpus_nb <- 1
for( term_to_search in dt_plot[grepl("2017", corpus_title),lemma][1:10] ){
  if(outputFormat == "html"){
    search_most_similar_paragraph(term_to_search, corpus_ = corpus_nb) %>%
      kable(caption = paste0("Paragraphes caractéristiques pour le terme « ", term_to_search,
        " » dans le corpus ", ifelse(corpus_nb ==1,"2017-2019", "2020")),
        ) %>% kable_styling("hover") %>%
      scroll_box(width = "100%", height = "200px") %>% htmltools::HTML() %>% print
      cat("\n")
  } else{
    search_most_similar_paragraph(term_to_search, corpus_ = corpus_nb) %>%

```

Table 15: Paragraphes caractéristiques pour le terme « découvrir » dans le corpus 2017-2019

Cosine	paragraphe
0.7190085	Les chercheurs ont notamment découvert :
0.5414218	Quelques actualités scientifiques à découvrir:
0.5167803	Les chercheurs ont analysé des fossiles découverts au cours des 30 dernières années et qui incluent plus d'espèces, dont quelques-unes ont été découvertes récemment seulement.
0.3865394	Les chercheurs tentent toujours de découvrir ce qui peut expliquer ce type de cancer.
0.3797436	Des chercheurs ont découvert qu'il est possible, en combinant l'âge des patients et les informations découvertes dans de simples analyses de sang, de prédire dans près de 80% des cas leur risque de souffrir d'une maladie chronique.

Table 16: Paragraphes caractéristiques pour le terme « participer » dans le corpus 2017-2019

Cosine	paragraphe
0.5923243	Le chercheur participe au projet BiotaNB depuis ses débuts.
0.3761234	Le premier patient à avoir participé à ce programme, le «patient zéro», a permis aux chercheurs de croire qu'ils étaient sur la bonne voie.
0.3592489	Des chercheurs de la Société de conservation de la faune du Canada ont également participé à cette étude.
0.3552853	Le premier patient à avoir participé à ce programme, le «patient zéro», a permis aux chercheurs de croire qu'ils étaient sur la bonne voie.
0.3515860	Des jeunes participent à la Foire des sciences de l'Université de Montréal. Photo : 24 heures de science

```

kable(caption = paste0("Paragraphes caractéristiques pour le terme « ", term_to_search,
                        " » dans le corpus ", ifelse(corpus_nb ==1,"2017-2019", "2020")
                        ), booktabs = TRUE
      ) %>% kable_styling(full_width = F) %>% column_spec(2, width = "35em") %>% print()
cat("\n")
}
}

```

```
rm(dt_tagged_tokens_science, dt_science)
```

Verbes de 2020

Nous faisons la même chose pour les 10 verbes les plus caractéristiques de 2020:

```

#Ces deux tables ont été produites dans 04-Morphosyntactic_analysis.R
load(paste0(data_folder, "dt_tagged_tokens_science.RData"))
load(paste0(data_folder, "dt_science.RData"))

corpus_nb <- 2
for( term_to_search in dt_plot[grepl("2020", corpus_title),rev(lemma)][1:10] ){
  if(outputFormat == "html"){
    search_most_similar_paragraph(term_to_search, corpus_ = corpus_nb) %>%

```

Table 17: Paragraphes caractéristiques pour le terme « améliorer » dans le corpus 2017-2019

Cosine	paragraphe
0.4198217	Les scientifiques en viennent à la conclusion que l'exercice améliore la vie quotidienne et la mobilité des aînés qui souffrent de la maladie d'Alzheimer; il pourrait aussi améliorer leur équilibre et leur cognition générale.
0.4182645	Les scientifiques en viennent à la conclusion que l'exercice améliore la vie quotidienne et la mobilité des aînés qui souffrent de la maladie d'Alzheimer; il pourrait aussi améliorer leur équilibre et leur cognition générale.
0.3598532	«[La série d'exercices] permet de renforcer les muscles, améliorer l'équilibre, améliorer la santé cardiovasculaire, voire la santé cognitive», souligne la chercheuse Mylène Aubertin-Leheudre, également docteure en kinésiologie.
0.3533776	Le rapport publié par le Conseil des ministres de l'Éducation du Canada (CMEC) précise également que le rendement en mathématiques s'est nettement amélioré dans presque toutes les provinces. Entre 2013 et 2016, le rendement en sciences s'est amélioré quant à lui dans la moitié des 10 provinces canadiennes.
0.3352167	Il rappelle que le gouvernement fédéral a annoncé un investissement de 24 millions \$ dans le programme d'aquaculture durable du Canada, qui vise à simplifier la réglementation, améliorer sa gestion, accroître les connaissances scientifiques, améliorer la prise de décisions axée sur la science et assurer la transparence grâce à la publication de rapports publics.

Table 18: Paragraphes caractéristiques pour le terme « intéresser » dans le corpus 2017-2019

Cosine	paragraphe
0.6475784	Il dit s'être toujours intéressé à la science.
0.3784631	Les chercheurs se sont intéressés à quelque 4200 jeunes âgés de 9 à 11 ans provenant de partout aux États-Unis.
0.3691491	Il souhaite, par cette expérience, emmener le grand public à s'intéresser à la science.
0.3690649	On lui demande si la Terre l'intéresse désormais plus que les étoiles, qui ont longtemps été au centre de ses recherches et de son oeuvre de vulgarisation scientifique. «Je m'intéresse toujours aux deux», dit-il. Mais dans le cas de l'écologie, il y a une «urgence», alors que les étoiles, elles, «peuvent attendre encore un peu».
0.3645930	«Oui, on s'intéresse à la science, mais au-delà de ça, il faut qu'on s'assure que les filles s'intéressent à l'école. C'est pour ça que nous avons aussi un volet d'aide aux devoirs et d'apprentissage du français», explique Valérie Bilodeau.

Table 19: Paragraphes caractéristiques pour le terme « organiser » dans le corpus 2017-2019

Cosine	paragraphe
0.4809659	C'est au gouvernement de s'organiser pour que la culture scientifique soit convenablement financée.
0.4213113	L'activité avait été organisée par Sylvie Latour, enseignante de sciences à l'école.
0.4196470	En janvier 2015, une chercheuse norvégienne, Emilie Oftedal, a étudié pour le compte du Norwegian Defense Research Establishment (FFI) 40 cellules terroristes qui ont, entre 1994 et 2013, organisé ou tenté d'organiser des attentats en Europe.
0.4006068	En janvier 2015, une chercheuse norvégienne, Émilie Oftedal, a étudié pour le compte du Norwegian Defense Research Establishment (FFI) quarante cellules terroristes qui ont, entre 1994 et 2013, organisé ou tenté d'organiser des attentats en Europe.
0.3684599	L'événement est organisé grâce à la collaboration du Conseil de recherches en sciences naturelles et en génie du Canada.

Table 20: Paragraphes caractéristiques pour le terme « consommer » dans le corpus 2017-2019

Cosine	paragraphe
0.4370618	En se basant sur des dizaines de mémoires et études scientifiques déposés à Ottawa ces derniers mois, l'ASPQ arrive à la conclusion que consommer du cannabis ne comporte pas plus de risques pour la santé que consommer de l'alcool. «Selon l'OMS, il est plus risqué de consommer de l'alcool ou d'être en surpoids que de consommer du cannabis, lorsque l'on analyse le nombre d'années perdues dues à la mort et à l'invalidité», affirme l'organisme dans un énoncé de position obtenu par <u>La Presse</u> . Les données provenant des États américains où la consommation récréative a été légalisée tendent à démontrer que la consommation chez les jeunes est restée stable depuis la légalisation. «Au Colorado, une légère augmentation a été observée au début de la légalisation. Par la suite, la consommation de cannabis chez les jeunes a diminué», écrit l'ASPQ.
0.4035406	Ces chercheurs ont en effet identifié une espèce, le merlu argenté, qui ne consomme aucune matière plastique.
0.4008297	L'équipe de chercheurs ignore toutefois si c'est le fait de consommer souvent du pot qui rend plus téméraire ou si c'est le fait d'avoir une personnalité téméraire qui pousse à consommer.
0.3649864	En examinant le lien entre ces cancers et les habitudes de consommation de thé, les chercheurs ont observé une hausse importante du risque (1,5 fois) chez les personnes qui le consommaient à une température supérieure à 60 °C, de même que chez celles qui le buvaient plus rapidement après l'infusion. En mesurant de façon objective la température des boissons consommées, ces résultats renforcent donc le lien existant entre les boissons très chaudes et le cancer de l'œsophage et suggèrent donc fortement que les amateurs de thé et de café ont tout avantage à laisser refroidir quelque peu leur boisson avant de la consommer.
0.3406970	La science déboulonne cet argument-là. Ça serait le même principe que de dire que ceux qui consomment de l'alcool ont d'abord commencé à boire du lait. Oui, la plupart des gens qui consomment des drogues dures ont commencé par le cannabis, mais beaucoup de fumeurs de cannabis ne vont jamais vers les drogues dures.

Table 21: Paragraphes caractéristiques pour le terme « expliquer » dans le corpus 2017-2019

Cosine	paragraphe
0.3351945	«Ces propriétés pourraient expliquer la brillance caractéristique des violons Stradivari», expliquent les chercheurs, dans leur étude publiée lundi dans la revue scientifique américaine Proceedings of the National Academy of Sciences.
0.3235538	Déjà, les résultats se font sentir. La chercheuse explique que les jeunes ont augmenté leurs connaissances en matière de violence amoureuse. «La jalousie n'est pas une preuve d'amour», explique-t-elle.
0.3004652	Cet hydrogène ne peut s'expliquer que par une activité hydrothermale propice à l'existence de la vie, ont expliqué les scientifiques en annonçant cette découverte le 13 avril.
0.2973654	Cet hydrogène ne peut s'expliquer que par une activité hydrothermale propice à l'existence de la vie, ont expliqué les scientifiques en annonçant cette découverte le 13 avril.
0.2814779	Prudents dans leurs analyses, des spécialistes ont expliqué qu'il y avait une base scientifique dans cette hypothèse.

Table 22: Paragraphes caractéristiques pour le terme « présenter » dans le corpus 2017-2019

Cosine	paragraphe
0.3941287	Il se désolé d'ailleurs de la façon dont est parfois présentée la science aux jeunes. «On leur présente ça comme une matière scolaire qui est très académique, alors que je fais léviter des objets grâce à la science! C'est l'fun!» Au final, il doit beaucoup à la science. «C'est ce qui m'a permis de percer dans le milieu, de créer de nouvelles illusions, de les présenter de façon différente.»
0.3600115	Cette journée, présentée dans le cadre de la Semaine de l'emploi au Bas-Saint-Laurent, est l'occasion pour les entreprises à la recherche de main-d'œuvre et les chercheurs d'emploi de La Mitis de se présenter et de faire connaissance.
0.3337001	Une délégation française composée de 24 jeunes et de 6 accompagnateurs et accompagnatrices sera présente tout au long de la finale québécoise. Cette délégation arrivera de Clermont-Ferrand à Auvergne en France. Ils présenteront leur projet scientifique au Québec, puis quatre jeunes originaires du Québec seront sélectionnés lors de la finale afin d'aller présenter leur projet lors de l'Exposciences Auvergne du 21 au 29 mai prochain.
0.3330399	Le chercheur présentait ses résultats dans le cadre d'un colloque sur la régionalisation de l'immigration.
0.3147622	__« __ Les scientifiques présentent quelques idées, mais ce n'est pas fait sur un ton universitaire __, explique Janis Timm-Bottos. __C'est très informel. On choisit aussi des gens avec des profils très diversifiés qui viennent présenter différents savoirs. On ne veut pas que des savoirs occidentaux. On souhaite présenter plusieurs voix et que tout le monde participe à la discussion. Les scientifiques veulent savoir aussi ce à quoi croient les gens et ce qui fonctionne pour eux. __ » __

Table 23: Paragraphes caractéristiques pour le terme « permettre » dans le corpus 2017-2019

Cosine	paragraphe
0.3590936	Permettre l'avancement des connaissances scientifiques
0.3537710	Le site science24heures.com permettra de faciliter vos recherches, selon :
0.3136587	Notre programme MAIN-D'OEUVRE 50+ permet aux gens de 50 ans et plus de s'inscrire sur notre plateforme, ce qui permet de faire le maillage entre les chercheurs d'emplois et les entreprises cibles.
0.2947989	Les experts du projet britannique espèrent que celui-ci permettra aux douanes de disposer de données scientifiques fiables permettant de déterminer rapidement si l'essence et l'origine du bois sont bien celles affichées.
0.2805347	Qui plus est pour John Stuart Mill, l'éducation est à la fois un art et une science. Elle doit à la fois permettre à l'individu de trouver le bonheur tout en lui permettant d'acquérir des compétences qui seront utiles aux autres.

Table 24: Paragraphes caractéristiques pour le terme « subir » dans le corpus 2017-2019

Cosine	paragraphe
0.3466105	Les chercheurs ont conclu que près de 70% des femmes atteintes de cette forme de cancer pourraient ne pas subir de chimiothérapie, en toute sécurité.
0.3466105	Les chercheurs ont conclu que près de 70 pour cent des femmes atteintes de cette forme de cancer pourraient ne pas subir de chimiothérapie, en toute sécurité.
0.3096612	Les aînés qui subissent une commotion cérébrale ont davantage de risque de développer la démence, selon une étude effectuée par des chercheurs à Toronto.
0.3008532	Le groupe de scientifiques a concentré ses recherches sur un gène nommé GRIN2B, connu pour causer l'autisme lorsqu'il subit une mutation.
0.2973033	Le mystère continue aussi de s'épaissir, alors que certaines victimes ont subi des lésions cérébrales dont la nature échappe encore à la science.

Table 25: Paragraphes caractéristiques pour le terme '« infecter » dans le corpus 2020

Cosine	paragraphe
0.4006674	Les anticorps monoclonaux identifiés et produits par des chercheurs de Québec « empêcheront le coronavirus d'aller infecter de nouvelles cellules. Si on les transfère directement dans le sérum des individus infectés, ils vont rapidement se distribuer dans l'organisme et neutraliser les virus qui sont en circulation. Ils auront toutefois peu d'effet sur les cellules qui sont déjà infectées, car l'anticorps ne pénètre pas dans la cellule. C'est la raison pour laquelle il est important d'administrer ce traitement tôt, lorsque l'infection se déclare, afin d'éviter que le virus ait le temps d'infecter beaucoup de cellules. Une fois qu'un grand nombre de cellules ont été infectées, il faut passer à d'autres types de traitement, comme des antiviraux », explique M. Lamarre.
0.3504047	• À lire aussi: Des centaines de Canadiens prêts à se faire infecter pour la science
0.2927341	Mais sans ces avancées scientifiques que sont le vaccin et les tests rapides, un tiers de la population, sinon plus, sera infectée, estime-t-il.
0.2921330	Ensuite, à partir de ces observations et de l'évaluation de l'intervalle de temps s'écoulant entre l'apparition des symptômes d'une première personne et l'apparition des symptômes de la personne qu'elle a infectée chez 77 paires de personnes, dont l'une a infecté l'autre, les chercheurs de la Guangzhou Medical University et de la University of Hong Kong ont pu calculer que les personnes infectées commencent à être contagieuses deux à trois jours avant l'émergence des symptômes et le sont particulièrement plus 0,7 jour avant de devenir symptomatiques. Ils ont également pu estimer que 44 % des transmissions étaient imputables à des personnes présymptomatiques.
0.2795589	Pour le moment, aucune donnée scientifique ne montre que les chiens ou les chats peuvent être infectés — ou transmettre — le virus.

```

kable(caption = paste0("Paragraphes caractéristiques pour le terme '< ", term_to_search,
                        " » dans le corpus ", ifelse(corpus_nb ==1,"2017-2019", "2020")),
      ) %>% kable_styling("hover") %>%
scroll_box(width = "100%", height = "200px") %>% htmltools::HTML() %>% print
cat("\n")
} else{
search_most_similar_paragraph(term_to_search, corpus_ = corpus_nb) %>%
  kable(caption = paste0("Paragraphes caractéristiques pour le terme '< ", term_to_search,
                        " » dans le corpus ", ifelse(corpus_nb ==1,"2017-2019", "2020")
                        ), booktabs = TRUE
      ) %>% kable_styling(full_width = F) %>% column_spec(2, width = "35em") %>% print()
  cat("\n")
}
}

rm(dt_tagged_tokens_science, dt_science)

```

Test de la prévalence des champs sémantiques liés au politique

Pour tester si la science a été plus liée à la régulation politique du gouvernement durant la COVID, voici notre stratégie :

Table 26: Paragraphes caractéristiques pour le terme '« contracter » dans le corpus 2020

Cosine	paragraphe
0.3688283	Les chercheurs analyseront si elles produisent les anticorps du Covid-19, signe qu'elles ont contracté la maladie et se sont rétablies.
0.3183892	Les chercheurs ont ensuite analysé via des tests en laboratoire et l'étude de symptômes combien d'entre eux ont contracté la COVID-19 dans les deux semaines suivantes.
0.2829170	Un article publié en 2019 par des chercheurs de l'université d'Arizona a révélé que les revêtements réduisaient de 36% les infections contractées en milieu hospitalier.
0.2822705	Un article publié en 2019 par des chercheurs de l'université d'Arizona a révélé que les revêtements réduisaient de 36 % les infections contractées en milieu hospitalier.
0.2820585	Une équipe de chercheurs chinois a constaté que le sperme de plusieurs personnes ayant contracté la COVID-19 était positif au virus, y compris chez des patients en rémission, selon une étude publiée jeudi par la revue scientifique américaine Journal of American Medical Association.

Table 27: Paragraphes caractéristiques pour le terme '« transmettre » dans le corpus 2020

Cosine	paragraphe
0.4060318	Et pour le moment, il n'existe aucune preuve scientifique qu'un animal peut transmettre la COVID-19 à l'homme.
0.3615706	Selon les autorités suisses, les scientifiques ont démontré que les jeunes enfants ne transmettent pas le coronavirus.
0.3584773	« Des chercheurs pensent que même avec une respiration normale, on peut transmettre la maladie », ajoute-t-il.
0.3320266	En date du 19 mars, l'état de la science n'indique pas si le virus peut se transmettre sexuellement.
0.3279215	Pour le moment, aucune donnée scientifique ne montre que les chiens ou les chats peuvent être infectés — ou transmettre — le virus.

Table 28: Paragraphes caractéristiques pour le terme '« rouvrir » dans le corpus 2020

Cosine	paragraphe
0.3802014	À Laval, la Maison des arts, dont la salle Alfred-Pellan, va rouvrir le 20 juin. À Joliette, le Musée d'art rouvre mardi. À Victoriaville, la réouverture du Centre d'art Jacques-et-Michel-Auger est prévue pour le 29 juillet. À Sherbrooke, le Musée d'histoire, le Musée des beaux-arts et le Musée de la nature et des sciences rouvriront mercredi. Enfin, à Trois-Rivières, le musée POP accueillera le public dès le 20 juin.
0.3638703	La décision de rouvrir les entreprises sera fondée sur l'opinion du médecin hygiéniste en chef et sur la science, a-t-il répété.
0.3555122	«Toutes ces institutions pourront rouvrir à leur rythme, parce qu'elles devront adapter les lieux à leur nouvelle réalité. Ça doit se faire en ayant comme objectif principal la santé de tous [...]. Il était important pour nous d'offrir de l'espoir», a illustré vendredi la ministre de la Culture, Nathalie Roy. Quatre musées d'État ainsi que 96 institutions muséales et 16 musées à vocation scientifique rouvriront vendredi prochain.
0.3380143	Il existe un consensus scientifique, selon M. Legault, pour rouvrir ces écoles sans que cela ne pose un risque pour le reste de la population si les mesures de distanciation physique sont respectées;
0.3298028	La chercheuse s'inquiète notamment de la possibilité de voir des pays exiger des voyageurs des preuves d'immunité lorsque les frontières vont rouvrir.

Table 29: Paragraphes caractéristiques pour le terme '« propager » dans le corpus 2020

Cosine	paragraphe
0.4459377	D'une part, certains chercheurs ont regardé de grands groupes de malades et analysé par des études épidémiologiques comment le virus se propage dans la population. Une compilation réalisée début mars par l'OMS en Chine sur 75 465 patients n'a pas trouvé de preuve que l'épidémie ait pu se propager par des aérosols.
0.4450131	Les chercheurs tentent de comprendre comment la COVID-19 se propage afin de savoir qui sont les plus grands propagateurs du virus.
0.3649778	Des théories du complot insinuant que le chercheur aurait été attaqué en raison de ses travaux sur le virus se sont pourtant propagées sur les réseaux sociaux.
0.3557681	«Le virus se propagera vite, quelles que soient les conditions climatiques», ajoute la chercheuse.
0.3557681	« Le virus se propagera vite, quelles que soient les conditions climatiques », ajoute la chercheuse.

Table 30: Paragraphes caractéristiques pour le terme '« contaminer » dans le corpus 2020

Cosine	paragraphe
0.3674552	Les scientifiques qui les promeuvent soulignent qu'ils servent à éviter de contaminer les autres plus qu'à se protéger soi-même.
0.3555320	Ces niveaux étaient semblables à ceux détectés chez les humains en voie de rétablissement après avoir été contaminés par le nouveau coronavirus, d'après les chercheurs.
0.3541762	Ces niveaux étaient semblables à ceux détectés chez les humains en voie de rétablissement après avoir été contaminés par le nouveau coronavirus, d'après les chercheurs.
0.3446400	Même si des milliers de personnes sont contaminées à travers le monde, les chercheurs et scientifiques qui s'affairent à comprendre le virus n'ont pas toutes les réponses.
0.2969137	Le coronavirus a fait officiellement plus de 11 000 morts et contaminé près de 170 000 personnes au Brésil, mais des scientifiques estiment ces chiffres largement en dessous de la réalité.

Table 31: Paragraphes caractéristiques pour le terme '« immuniser » dans le corpus 2020

Cosine	paragraphe
0.5255113	On a ouvert les garderies et les écoles primaires pour créer «une immunité collective », c'est-à-dire s'immuniser contre ça. La meilleure façon de s'immuniser, c'est de la pognier la COVID-19. Plate hein? Mais c'est ça. J'ai vu quelque part que de prendre un peu de sang de quelqu'un d'immunisé (qui l'a pogné et guéri) et l'injecter dans le corps de quelqu'un immuniserait la personne qui se l'ai fait injecter... Vieille technique de vaccin, je me trompe peut-être, au pire corrigez-moi. Sauf que si c'est vrai, je ne suis pas médecin ou scientifique, mais il n'y a rien de compliqué là.
0.4469741	Pour lui, l'état actuel de la science indique qu'en théorie une personne qui guérit de la COVID-19 devrait être immunisée pour quelques années.
0.4359405	Normalement, cet homme aurait dû être immunisé à la suite de la première infection. Du moins, c'est ce que pensent de nombreux scientifiques.
0.3585580	Selon l'OMS, les preuves scientifiques ne sont pas suffisantes pour démontrer que les personnes qui ont vaincu la COVID-19 et qui ont des anticorps sont immunisées contre une deuxième infection.
0.3219692	«J'ai peur de reprendre le travail», dit-il à l'AFP. «Un médecin m'a dit que j'étais immunisé, mais ce n'est pas encore prouvé scientifiquement».

Table 32: Paragraphes caractéristiques pour le terme '« tester » dans le corpus 2020

Cosine	paragraphe
0.3833007	"Notre urgence, c'est d'essayer de transmettre ces molécules à des laboratoires qui peuvent les tester, dans un environnement sécuritaire, des gens qui sont capables de travailler avec le coronavirus et de tester si les molécules sont efficaces", lance le professeur à la Faculté de médecine et des sciences de la santé de l'Université de Sherbrooke.
0.3821897	Le chercheur teste deux technologies différentes et se donne quatre mois pour faire le bon choix.
0.2953718	> Dans une situation de pandémie, dès qu'on soupçonne que c'est un cas de > COVID, on le considère comme un cas de COVID. La logique derrière ça c'est > qu'on ne peut pas tester tout le monde. Dans un monde idéal, on testerait > tout le monde, mais c'est impossible de le faire maintenant. > > Chantal Srivastava, journaliste scientifique, Les Années lumière
0.2943462	Ses chercheurs ont aussi testé la résistance du virus à diverses conditions de température et d'humidité en intérieur.
0.2941808	Seuls les chercheurs du Nebraska ont ensuite vérifié si ces aérosols avaient le pouvoir d'être infectieux. Pour cela, ils ont placé les échantillons d'air infecté au contact de cultures de cellules appelées Vero E6, utilisées fréquemment pour tester la réplication de divers virus. Or ils n'ont trouvé aucune trace d'activité virale dans tous les échantillons testés, ce qui est plutôt rassurant.

Table 33: Paragraphes caractéristiques pour le terme '« lutter » dans le corpus 2020

Cosine	paragraphe
0.3498541	Sauf que, jusqu'ici, aucune étude scientifique rigoureuse n'a démontré l'efficacité de ces molécules pour lutter contre l'infection au coronavirus chez l'être humain.
0.3045402	À l'échelle internationale, la Réponse mondiale au coronavirus vise à recueillir plus de huit milliards de dollars américains pour aider les chercheurs à lutter contre la COVID-19.
0.2937553	Décrit par Donald Trump comme «un don du ciel» pour lutter contre la pandémie, la chloroquine ne cesse en fait de diviser la communauté scientifique.
0.2880693	Selon Mme Rang, les chercheurs s'efforcent également de lutter contre le flétrissement bactérien du chêne.
0.2803735	Moduler les mesures de distanciation sociale à long terme pourrait permettre de lutter contre la COVID-19 sans asphyxier l'économie, selon une nouvelle étude réalisée par des chercheurs ontariens.

Table 34: Paragraphes caractéristiques pour le terme '« fermer » dans le corpus 2020

Cosine	paragraphe
0.4761385	« Je ne pense pas que l'on devrait les fermer, mais nous, on s'en remet à la science et aux autorités de santé publique s'il faut fermer des entreprises ou des chantiers de la construction parce qu'on est rendu là, va falloir s'y soumettre », a-t-il conclu.
0.3957177	«Nous ne fermons la porte à aucune mesure supplémentaire, mais nous prendrons ces décisions en fonction de ce que la science nous dit», a-t-il affirmé.
0.3365640	Tous ces comptes ont été fermés, mais leur contenu a été sauvegardé sur une base de données à des fins de recherche scientifique, a précisé Twitter.
0.3221235	> Si nous avons tout fermé dès le début, la situation aurait été différente. > Mais il y a eu beaucoup d'opposition. > > Dr Anthony Fauci, scientifique en chef de la Maison-Blanche
0.2567042	"Pour Montréal, l'idée serait de fermer les points d'entrée, les ponts, les tunnels", détaille Benoit Barbeau, professeur au Département des sciences biologiques de l'UQAM.

1. Récupération du lexique français du système d'analyse sémantique UCREL.
2. Extraction des marqueurs « G1.1 Gouvernement etc. », « G1.2 Politique » et « G2.1 Ordre public ». Ces marqueurs font partie de la catégorie générale « G. Gouvernement et domaine public ». Le marqueur qui nous intéresse principalement est G1.1. Nous prenons les deux autres marqueurs pour nous permettre de tester si le changement dans la prévalence du champ sémantique est large ou s'il est plutôt concentré sur **l'action gouvernementale** (notre hypothèse). En effet, le marqueur G1.2 est davantage lié à la politique partisane et le G2.1 à celui des systèmes légal et pénal. À noter que nous ne prenons pas certains des marqueurs sous la catégorie G1: « G2.2 Éthique générale », « G3 Guerre, défense et armé; armement ». Ceux-ci ont été considérés trop loin de notre hypothèse pour mériter notre attention.
3. Amélioration manuelle du lexique pour chacun de ces marqueurs.
4. Recherche des paragraphes liés à la science qui ont au moins un mot des lexiques de ces marqueurs. Ces paragraphes sont considérés comme étant liés au champ sémantique qui nous intéresse. Ci-dessous, nous donnons les fréquences de chaque item des lexiques dans notre corpus.
5. Calcul de la proportion des paragraphes liés à chaque champ sémantique sur l'ensemble des paragraphes. Ce calcul est fait pour le corpus 2017-2019, d'une part, et pour le corpus 2020 d'autre part.
6. Test de proportions égales pour déterminer si la différence entre les proportions peut être attribuée à du bruit.
7. Test de robustesse du résultat en retirant certains items des lexiques qui peuvent être considérés comme n'étant pas suffisamment liés aux champs sémantiques visés.

Le bloc de code ci-dessous réalise l'étape 2 et une partie de l'étape 3 en retirant des termes manifestement trop loin des champs sémantiques visés (les traductions du lexique de l'anglais vers le français n'étant pas toujours heureuses).

```
# Ce code est roulé seulement une fois puisque nous éditons manuellement
# ensuite les fichiers de sortie pour régler certains problèmes
# avec les lexiques français.

dictionnaire_usas <- read.csv(paste0(data_folder,"dictionnaire_usas.txt"),
                             sep = "\t") %>% data.table()
setnames(dictionnaire_usas, c("X.French.Single.Word.Semantic.Lexicon.", "X", "X.1"),
         c("mot", "upos", "codes_usas"))

#Prendre seulement les trois premiers marqueurs par terme
dictionnaire_usas$codes_usas <- dictionnaire_usas$codes_usas %>%
```

```

str_split(., pattern = " ") %>%
lapply(., function(x) x[c(1:3)])

# Termes liés au gouvernement 'G1.1'
mots_gouvernement <- dictionnaire_usas[grepl("G1.1", codes_usas)]
mots_gouvernement <- mots_gouvernement[!mot %in% c("maison", "vert")]
#fwrite(mots_gouvernement,)

# Termes liés à la politique partisane ou électorale 'G1.2'
mots_politique <- dictionnaire_usas[grepl("G1.2", codes_usas)]
mots_politique <- mots_politique[!mot %in% c("billet", "courir", "démontrer", "rouge")]

# Termes liés au judiciaire/pénal 'G2.1'
mots_crimes <- dictionnaire_usas[grepl("G2.1", codes_usas)]
mots_crimes <- mots_crimes[!mot %in% c("billet", "boutique", "bref", "cas", "courir",
                                     "docteur", "spécial", "station", "volonté")]

mots_gouvernement %>% fwrite(paste0(accessible_data_folder, "mots_gvt.csv"))
mots_politique %>% fwrite(paste0(accessible_data_folder, "mots_pol.csv"))
mots_crimes %>% fwrite(paste0(accessible_data_folder, "mots_crime.csv"))

```

En plus, des modifications dans le bloc de code ci-dessus, certains changements ont été faits manuellement aux trois lexiques avant l'analyse. Les différences entre les fichiers d'origine et les fichiers avec `_edited` peuvent être comparées au besoin dans le dossier `data/`. Les mots ajoutés sont identifiés par un « ajouté par ... » dans la dernière colonne. Ils visent principalement à tenir compte du contexte canadien (p. ex., avec les variantes de « province » dans le lexique de G1.1).

Les tableaux ci-dessous donnent les items des lexiques pour nos trois marqueurs. Ils sont triés en ordre décroissant de présence dans notre corpus. On voit que les termes les plus fréquents pour G1.1 (« gouvernement ») et pour G1.2 (« politique ») sont exactement sur la cible. Par contre, on peut douter de la pertinence d'autres termes avec des fréquences élevées – p. ex., « social » dans G1.1; « proposer » dans G1.2; « entendre » dans G2.1.

Notre stratégie est de tester d'abord avec ces lexiques et d'ensuite faire un lexique retirant ces items fréquents, mais possiblement moins pertinents, pour tester la stabilité des résultats.

```

# Chargement du gros fichier résultant de l'analyse morphosyntaxique
load(paste0(data_folder, "dt_tagged_tokens_science.RData"))

# Chargement des lexiques
mots_gouvernement <- fread(paste0(accessible_data_folder#data_folder
                                   , "mots_gvt_edited.csv"))
mots_politique <- fread(paste0(accessible_data_folder#data_folder
                                , "mots_pol_edited.csv"))
mots_crimes <- fread(paste0(accessible_data_folder#data_folder
                             , "mots_crime_edited.csv"))

# Calcul de la fréquence avec une fonction
fct_compute_freq_lexicon <- function(lexicon){
  dt_count <- dt_tagged_tokens_science[lemma %in% lexicon$mot,
                                         list(corpus, doc_id, lemma)] %>%
  unique() %>%
  .[,list(nb_of_succes = .N), by = .(corpus, lemma)]

```

```

dt_count[corpus == 1, corpus_years := "2017-2019"][
  is.na(corpus_years), corpus_years := "2020"]
dt_count <- pivot_wider(dt_count, id_cols = "lemma", names_from = "corpus_years",
  values_from = "nb_of_succes") %>%

data.table()
dt_count <- merge(dt_count, lexicon, by.x = "lemma", by.y = "mot", all.y = TRUE)
dt_count[is.na(`2017-2019`), `2017-2019` := 0]
dt_count[is.na(`2020`), `2020` := 0]
dt_count[, Total := `2017-2019` + `2020`]
setnames(dt_count, "lemma", "mot")
setorder(dt_count, -Total, mot)
setcolorder(dt_count, c("mot", "Total"))
return(dt_count)
}

mots_gouvernement <- fct_compute_freq_lexicon(mots_gouvernement)
mots_politique <- fct_compute_freq_lexicon(mots_politique)
mots_crimes <- fct_compute_freq_lexicon(mots_crimes)

# Inspection des lexiqes:
kable(mots_gouvernement, caption =
  "Fréquence des termes du lexique français pour 'G1.1 Government'
  (adapté de USAS)", longtable = TRUE, booktabs = TRUE) %>% kable_styling() %>%
  scroll_box(width = "100%", height = "200px")

```

Table 35: Fréquence des termes du lexique français pour 'G1.1 Government' (adapté de USAS)

mot	Total	2017-2019	2020	upos	codes_usas
gouvernement	1828	981	847	noun	G1.1c
social	1666	1060	606	noun	S1.1.1 I1.1/G1.1
pays	1632	927	705	noun	G1.1c W3 F4/M7
ministre	974	596	378	noun	G1.1/S2mf S9/S2mf
président	939	609	330	noun	G1.1/S7.1+/S2mf
état	596	397	199	noun	G1.1c O4.1 Q2.1
ministère	515	385	130	noun	G1.1c S9
citoyen	489	303	186	noun	G1.1/S2mf S5+/S2mf
autorité	477	191	286	noun	G1.1c S7.1+ S7.4+
fédéral	468	306	162	adj	ajouté par JHR
province	449	294	155	nom	ajouté par JHR
agence	425	305	120	noun	I2.2/S5+c G1.1c
loi	385	312	73	nom	ajouté par JHR
commun	367	236	131	nom	ajouté par JHR
commission	363	263	100	noun	A1.1.1 I1.1/I2.2 G1.1
État	317	197	120	noun	ajouté par FC
privé	316	229	87	noun	A10- S5- G1.1-
nation	223	156	67	noun	G1.1c S5+c
officiel	211	121	90	noun	G1.1 S7.4+ G1.1/S2mf
régime	193	172	21	noun	G1.1 S7.1+ B3
budget	190	156	34	noun	I1 G1.1/I1 I1.3-

représentant	186	154	32	noun	A6.2+ G1.1 A4.1
agent	181	131	50	noun	I2.1/S2mf G1.1/X2.2+/S2mf K4/S2mf
frontière	177	95	82	noun	M7/G1.1 M6 O2
municipalité	170	143	27	noun	M7/G1.1
provincial	166	126	40	adj	ajouté par JHR
maire	151	120	31	noun	G1.1/S7.1+/S2mf
municipal	134	96	38	adj	M7/G1.1
élu	107	75	32	nom	ajouté par JHR
approuver	104	64	40	verb	S7.4+ E2+ G1.1
capitale	94	69	25	noun	I1.1 G1.1 A5.1+
représentation	86	70	16	noun	G1.1 C1 Q1.1
vice-président	82	49	33	noun	G1.1/S7.1+/S2mf
exécutif	79	34	45	noun	I2.1/S2mf G1.1c S7.1+/S5+c
officiellement	68	39	29	adv	G1.1 S7.4+
réglementation	61	56	5	noun	G2.1 A1.7+/G1.1
taxe	60	58	2	noun	I1.1/G1.1 A12-
gouverneur	57	36	21	noun	S7.1+/S2mf G1.1/S2mf O2%
impôt	56	44	12	noun	I1.1/G1.1 A12-
représentatif	52	37	15	adj	A6.2+ G1.1 A4.1
intergouvernemental	49	40	9	adj	G1.1/S5+
fiscal	46	36	10	adj	G2.1/S2mf I1/G1.1 I1
traité	42	38	4	noun	G3 G1.1 G2.1
diplomatique	39	34	5	noun	G1.1 S1.2.4+
approbation	36	25	11	noun	S7.4+ E2+ G1.1
ambassadeur	34	20	14	noun	G1.1/S2mf S1.1.1/S2mf
diplomatie	31	22	9	noun	S1.2.4+ G1.1
législatif	30	27	3	adj	ajouté par JHR
présidentiel	28	16	12	adj	G1.1/S7.1+
parlement	26	23	3	noun	G1.1
décret	18	14	4	nom	ajouté par JHR
inspecteur	18	12	6	noun	G2.1/S2mf G1.1/S2mf
royaume	18	14	4	noun	G1.1 M7/S7.1+ A4.1/L2
citoyenneté	16	10	6	noun	G1.1 S5+
constitutionnel	16	13	3	noun	G1.1 M1
délégué	15	14	1	noun	G1.1/S2mf S7.4+ A9-
empire	15	11	4	noun	G1.1 I2.1/S+
nationalité	15	14	1	noun	G1.1c S5+c
législateur	14	7	7	nom	ajouté par JHR
constitution	13	11	2	noun	G1.1 B2 O1/A1.8+
passport	12	3	9	noun	G1.1/Q1.2
privatisation	11	7	4	noun	G1.1-
imposition	9	7	2	noun	I1.1/G1.1
taxer	9	8	1	verb	I1.1/G1.1 A12-
ambassade	5	2	3	noun	G1.1/S7.1+/S5+ G1.1/H1
législature	2	1	1	noun	G1.1c

```
kable(mots_politique, caption =
  "Fréquence des termes du lexique français pour 'G1.2 Politics'
  (adapté de USAS)", longtable = TRUE, booktabs =TRUE) %>% kable_styling() %>%
```

```
scroll_box(width = "100%", height = "200px")
```

Table 36: Fréquence des termes du lexique français pour 'G1.2 Politics' (adapté de USAS)

mot	Total	2017-2019	2020	upos	codes_usas
politique	2256	1565	691	nom	ajouté par JHR
proposer	684	507	177	verb	Q2.2/X7+ X4.1 G1.2
parti	250	196	54	noun	K1/S1.1.3+ K1/S1.1.3+c G1.2c
candidat	240	193	47	noun	A7+ A1.2+ G1.2/S2mf
sondage	201	137	64	noun	G1.2 Q1.1
député	187	143	44	nom	ajouté par FC
élection	179	147	32	noun	G1.2/X7+
conservateur	151	127	24	noun	G1.2 A6.2+ A1.3+
militant	126	86	40	noun	E3-/G1.2 E3-/G1.2/S2mf
politicien	124	72	52	noun	G1.2/S2mf
électoral	123	88	35	adj	G1.2/X7+
libéral	106	77	29	noun	A1.7- G1.2 N5+
démocratique	97	65	32	adj	G1.2
révolution	93	74	19	noun	G1.2 A2.1+ M1@
assemblée	91	82	9	nom	ajouté par FC
démocratie	89	66	23	noun	G1.2
manifester	89	69	20	verbe	ajouté par JHR
manifestation	84	63	21	nom	ajouté par JHR
syndicat	83	55	28	noun	G1.2/I3.1/S5+c S5+c
radical	80	54	26	noun	G1.2/S2mf O1% N5+/A2.1
complot	78	20	58	nom	ajouté par JHR
voter	72	61	11	verb	G1.2/X7
républicain	71	50	21	noun	G1.2 G1.2/S2mf
élire	70	52	18	verb	G1.2/X7+
progressif	62	11	51	adj	G1.2 T3— N3.8-
vote	54	37	17	noun	G1.2/X7
capitalisme	31	26	5	noun	G1.2/I2.1
marginal	28	24	4	noun	A11.2- A7- G1.2/A7-
nazi	27	27	0	noun	G1.2 G1.2/S2mf
sénateur	27	21	6	noun	G1.2/S2mf
république	26	16	10	noun	G1.2 S5+/M7
grève	25	19	6	noun	I3.1-/G1.2 E3- A1.1.1
conspirationniste	22	4	18	nom	ajouté par JHR
progressiste	22	20	2	noun	G1.2 T3— N3.8-
référendum	16	13	3	noun	G1.2
communiste	15	10	5	noun	G1.2 G1.2/S2mf
complotiste	15	0	15	nom	ajouté par JHR
protester	12	11	1	verb	Q2.2 A10+ G1.2
propagande	11	9	2	noun	G1.2
socialiste	6	6	0	noun	G1.2 G1.2/S2mf
suffrage	6	4	2	noun	G1.2/X7
conspiration	5	1	4	nom	ajouté par JHR
rébellion	5	5	0	noun	E3- G1.2/A6.1-
socialisme	4	3	1	noun	G1.2

communisme	2	2	0	noun	G1.2
députer	2	1	1	verbe	ajouté par FC
séparatiste	1	0	1	noun	G1.2 G1.2/S2mf
complotisme	0	0	0	nom	ajouté par JHR
conspirationnisme	0	0	0	nom	ajouté par JHR
qanon	0	0	0	nom	ajouté par JHR

```
kable(mots_crimes, caption =
  "Fréquence des termes du lexique français pour 'G2.1 Crime and Law'
  (adapté de USAS)", longtable = TRUE, booktabs = TRUE) %>% kable_styling() %>%
  scroll_box(width = "100%", height = "200px")
```

Table 37: Fréquence des termes du lexique français pour 'G2.1 Crime and Law' (adapté de USAS)

mot	Total	2017-2019	2020	upos	codes_usas
entendre	609	386	223	verb	X3.2 X2.3+ G2.1@
cellule	461	317	144	noun	B1 G2.1/H2 S9/H2
force	343	227	116	adv	E3- S7.1+ G2.1c
victime	324	263	61	noun	A1.1.2/G2.1/S2mf B2-/S2mf
ordre	286	194	92	nom	ajouté par JHR
juger	269	160	109	verb	G2.1 X6
judiciaire	246	201	45	adj	G2.1c
discipline	229	182	47	noun	G2.1 A1.7+ P1
police	229	163	66	noun	G2.1 G2.2
règle	218	128	90	noun	G2.1 X4.2 A6.2+
policier	209	117	92	noun	G2.1/I3.2/S2.2m
arrêter	199	130	69	verbe	ajouté par JHR
accuser	156	85	71	verb	Q2.2 G2.1
témoigner	154	105	49	verb	G2.1 A5.2+
plaider	136	116	20	verb	Q2.2 G2.1/Q2.2
juge	112	91	21	noun	G2.1 X6 G2.1/I3.2/S2mf
crime	101	88	13	noun	G2.1-
témoignage	95	81	14	noun	G2.1 Q2.2
défense	93	65	28	noun	S8+ G3 G2.1c
avocat	92	68	24	noun	G2.1/I3.2/S2mf
témoin	89	66	23	noun	G2.1/S2mf X3.4/S2mf S9/S2mf
justice	86	73	13	noun	G2.1+ G2.2+
jugement	71	47	24	noun	X2.1 X6 G2.1
tribunal	71	63	8	noun	G2.1 H3 M7/K5.2
condamner	65	47	18	verb	Q2.2/E2- Q2.2/A1.1.2 G2.1
dégât	64	55	9	noun	A1.1.2 I1.1+/G2.1
réglementation	61	56	5	noun	G2.1 A1.7+/G1.1
défenseur	56	41	15	noun	S8+/S2mf G2.1/S2mf G3/S2mf
commettre	54	45	9	verb	A1.1.1 S6+ A1.7+/G2.1
criminel	48	35	13	noun	G2.1- A5.1- G2.1-/S2mf
coupable	43	38	5	noun	G2.1- E4.1-
sanction	42	39	3	noun	S7.4 G2.1
traité	42	38	4	noun	G3 G1.1 G2.1
fraude	39	29	10	noun	G2.1- G2.2-

prison	38	26	12	noun	G2.1/H1
illégal	35	27	8	adj	G2.1-
législatif	30	27	3	noun	G2.1
procureur	28	18	10	noun	G2.1/I3.1/S2mf
privilège	27	19	8	noun	A5.1+ G2.1 S7.1+
terroriste	25	23	2	noun	G2.1-/S2mf
terrorisme	23	19	4	noun	G2.1-
détenu	22	17	5	noun	G2.1/S2mf A1.7+/S2mf
arrestation	19	16	3	noun	G2.1
criminalité	18	14	4	noun	G2.1-
inspecteur	18	12	6	noun	G2.1/S2mf G1.1/S2mf
corruption	16	12	4	noun	G2.2- G2.1- Q3
législation	16	16	0	noun	G2.1
verdict	14	11	3	noun	G2.1 X6+
viol	14	13	1	noun	G2.1-/S3.2 A1.1.2
chaos	13	8	5	noun	A1.7- G2.1- Y1@
pénal	13	9	4	adj	G2.1
culpabilité	11	9	2	noun	G2.2- G2.1-
révolte	8	5	3	noun	G2.1- E3- G3
commandant	7	6	1	noun	S7.1+/S2mf G3/S7.1+/S2mf G2.1/S2mf
innocent	7	6	1	noun	G2.1+ X9.1- G2.2+
accusé	6	6	0	noun	G2.1/S2mf
punir	6	5	1	verb	G2.1
innocence	5	2	3	noun	G2.1+ G2.2+
inculper	4	2	2	verb	I1.3 O3 G2.1
lieutenant	2	2	0	noun	G3/S7.1+/S2mf G2.1/S2mf S7.1+/S2mf
punition	2	1	1	noun	G2.1

Fonction pour les tests de proportion afin de pouvoir changer les lexiques d'entrée:
fct_test_prop_champ_sem <- fonction(mots_gvt, mots_pol, mots_crim){

#Test de proportion pour les termes G1.1

```
test_gouvernement <- dt_tagged_tokens_science[lemma %in% mots_gvt$mot,
                                                list(corpus, doc_id)] %>%
  unique() %>%
  .[,list(nb_of_succes = .N), by = corpus]
test_gouvernement <- merge(test_gouvernement,
  dt_tagged_tokens_science[,list(corpus, doc_id)] %>%
  unique() %>%
  .[,list(total_nb_paragraphe = .N), by = corpus],
  by = "corpus")
test_results_gouvernement <- prop.test(x=test_gouvernement$nb_of_succes,
                                         n=test_gouvernement$total_nb_paragraphe)#,
  # alternative = "less")
```

#Test de proportion pour les termes G1.2

```
test_politique <- dt_tagged_tokens_science[lemma %in% mots_pol$mot,
                                              list(corpus, doc_id)] %>%
  unique() %>%
  .[,list(nb_of_succes = .N), by = corpus]
```

```

test_politique <- merge(test_politique,
  dt_tagged_tokens_science[,list(corpus, doc_id)] %>%
  unique() %>%
  .[,list(total_nb_paragraphe = .N), by = corpus],
  by = "corpus")
test_results_politique <- prop.test(x=test_politique$nb_of_succes,
                                   n=test_politique$total_nb_paragraphe)#,
  # alternative = "less")

#Test de proportion pour les termes G2.1
test_crimes <- dt_tagged_tokens_science[lemma %in% mots_crim$mot,
                                         list(corpus, doc_id)] %>%
  unique() %>%
  .[,list(nb_of_succes = .N), by = corpus]
test_crimes <- merge(test_crimes,
  dt_tagged_tokens_science[,list(corpus, doc_id)] %>%
  unique() %>%
  .[,list(total_nb_paragraphe = .N), by = corpus],
  by = "corpus")
test_results_crimes <- prop.test(x=test_crimes$nb_of_succes,
                                n=test_crimes$total_nb_paragraphe)#,
  # alternative = "less")

p_values <- rbind(data.table(two_proportions_test_on = "government related terms",
                             p_value = test_results_gouvernement$p.value),
  data.table(two_proportions_test_on = "politics related terms",
             p_value = test_results_politique$p.value),
  data.table(two_proportions_test_on = "crime related terms",
             p_value = test_results_crimes$p.value))
p_values[, p_value := round(p_value, digits = 5)]

proportion_tests <- rbind(data.table(`Champ sémantique` = "G1.1 Gouvernement",
  `Proportion avant COVID` = round(test_results_gouvernement$estimate[1],3),
  `Proportion pendant COVID` = round(test_results_gouvernement$estimate[2],3),
  `p_value` = test_results_gouvernement$p.value)
,
  data.table(`Champ sémantique` = "G1.2 Politique",
  `Proportion avant COVID` = round(test_results_politique$estimate[1],3),
  `Proportion pendant COVID` = round(test_results_politique$estimate[2],3),
  p_value = test_results_politique$p.value),
  data.table(`Champ sémantique` = "G2.1 Ordre public",
  `Proportion avant COVID` = round(test_results_crimes$estimate[1],3),
  `Proportion pendant COVID` = round(test_results_crimes$estimate[2],3),
  p_value = test_results_crimes$p.value))
proportion_tests[, `Valeur p` := round(p_value, digits = 4)][,p_value :=NULL]

proportion_tests[, `% de changement` :=
  round(( `Proportion pendant COVID` - `Proportion avant COVID`)/
  `Proportion avant COVID`*100, 1) %>% paste("%")]
return(proportion_tests)
}

```

Table 38: Résultats des tests d'égalité de proportion pour les trois champs sémantiques

Champ sémantique	Proportion avant COVID	Proportion pendant COVID	Valeur p	% de changement
G1.1 Gouvernement	0.227	0.269	0.0000	18.5 %
G1.2 Politique	0.102	0.094	0.0028	-7.8 %
G2.1 Ordre public	0.105	0.093	0.0000	-11.4 %

Table 39: Termes retirés des lexiques

Champ sémantique	item
G1.1 Gouvernement	social
G1.2 Politique	proposer
G2.1 Ordre public	entendre

```
proportion_tests <- fct_test_prop_champ_sem(mots_gouvernement,
                                             mots_politique, mots_crimes)

kable(proportion_tests, caption =
  "Résultats des tests d'égalité de proportion pour les trois champs sémantiques"
) %>% kable_styling("hover", full_width = F)
```

```
rm(dt_tagged_tokens_science)
```

Le test de proportion donne donc des changements statistiquement significatifs pour les trois champs sémantiques : une augmentation pour le champ sur l'action gouvernementale (G1.1), mais des diminutions pour les champs de la politique partisane (G1.2) et de l'ordre public (G2.1).

Pour tester la robustesse de ce résultat, nous enlevons les termes suivants de chaque lexique et faisons le même test à nouveau.

```
# Changement du gros fichier résultant de l'analyse morphosyntaxique
load(paste0(data_folder, "dt_tagged_tokens_science.RData"))

mots_out_gvt <- c("social")
mots_out_pol <- c("proposer")
mots_out_loi <- c("entendre")

mots_out <- data.table(`Champ sémantique` =
  c("G1.1 Gouvernement", "G1.2 Politique", "G2.1 Ordre public"),
  `item` = c(paste(mots_out_gvt, collapse = ", "),
    paste(mots_out_pol, collapse = ", "),
    paste(mots_out_loi, collapse = ", "))
)

kable(mots_out, caption = "Termes retirés des lexiques") %>%
  kable_styling("hover", full_width = F) # %>%

# scroll_box(width = "100%", height = "200px")
```

Table 40: Résultats des tests d'égale proportion pour les trois champs sémantiques

Champ sémantique	Proportion avant COVID	Proportion pendant COVID	Valeur p	% de changement
G1.1 Gouvernement	0.203	0.241	0.0000	18.7 %
G1.2 Politique	0.089	0.084	0.0546	-5.6 %
G2.1 Ordre public	0.095	0.080	0.0000	-15.8 %

```

proportion_tests2 <- fct_test_prop_champ_sem(mots_gouvernement[!mot %in% mots_out_gvt],
  mots_politique[!mot %in% mots_out_pol],
  mots_crimes[!mot %in% mots_out_loi])

kable(proportion_tests2, caption =
  "Résultats des tests d'égale proportion pour les trois champs sémantiques"
) %>% kable_styling("hover", full_width = F)

rm(dt_tagged_tokens_science)

```

Les résultats demeurent qualitativement les mêmes sans ces termes trop peu spécifiques et très fréquents. Comme le lexique semble meilleur sans ces termes, nous rapportons ces derniers résultats pour l'article principal.

La présence de l'incertitude

Maintenant, il s'agit de voir si les paragraphes à propos de la science sont davantage marqués par l'incertitude pendant l'année de la COVID.

Analyse de cooccurrence

Une première analyse est de regarder la centralité des verbes les plus caractéristiques de chaque corpus. Voir figure_A5.jpg et figure_A6.jpg.

```

#Ces deux tables ont été produites dans 04-Morphosyntactic_analysis.R
load(paste0(data_folder, "dt_tagged_tokens_science.RData"))
load(paste0(data_folder, "dt_science.RData"))

nb_mots <- 25

#Pour passer l'analyse de cooccurrence sur les deux corpus, on fait une longue fonction
#pour éviter de copier du code
analyse_cooccurrence <- function(corpus, nb_mots)
{
  dt_lemmas <- dt_tagged_tokens_science[corpus == corpus & upos %in% c('VERB'),
    list(lemma, art_ID, doc_id)]

  # La fonction itoken de text2vec veut ce format, c'est-à-dire,
  # une liste de token (ou lemme) par document
  dt_lemmas <- dt_lemmas[,.(lemma = list(lemma)), by = doc_id]

  #Structure de donnée qui permet de travailler avec nos données sans faire de copies
  iterateur <- text2vec::itoken(dt_lemmas$lemma,
    ids = dt_lemmas$doc_id,

```

```

progressbar = FALSE)

vocabulaire <- text2vec::create_vocabulary(iterateur,
  ngram=c(ngram_min = 1,
  ngram_max = 1))

vocabulaire <- text2vec::prune_vocabulary(vocabulaire,
  term_count_min = 2,
  term_count_max = Inf,
  doc_proportion_min = 0.0001,
  doc_proportion_max = 0.5,
  doc_count_min = 1,
  doc_count_max = Inf,
  vocab_term_max = Inf)

setDT(vocabulaire)

# Permet de vectoriser nos documents. Le vectorizer est utilisé
# par create_dtm (voir la doc si ce n'est pas clair)
vectorizer <- text2vec::vocab_vectorizer(vocabulaire)

dtm <- text2vec::create_dtm(iterateur,
  vectorizer,
  type = "dgCMatrix")

#On transforme le dtm pour que les cellules contiennent les valeurs tf-idf
dtm_tfidf_initialisation <- text2vec::TfIdf$new()
dtm_tfidf <- dtm_tfidf_initialisation$fit_transform(dtm)
rm(dtm_tfidf_initialisation)

#Calculer la somme des valeurs TFIDF pour les documents
vecteur_sumMots <- Matrix::colSums(dtm_tfidf)

# Selection des tops mots (pour le corpus 2, les valeurs sont négatives,
# ce pourquoi on utilise tail)
if (corpus == 1)
{
  top_mots <- dt_chi_sqr[,list(lemma)] %>% head(nb_mots)
}
if (corpus == 2)
{
  top_mots <- dt_chi_sqr[,list(lemma)] %>% tail(nb_mots)
}

tdm_cluster <- Matrix::t(dtm_tfidf)

tdm_cluster_norm <- text2vec::normalize(tdm_cluster, norm = 'l2')

#Calcul du cosinus par produit matriciel de la matrice normalisée avec sa transposée
mot_mot_cosine <- Matrix::tcrossprod(tdm_cluster_norm)

#Réduction à zéro des valeurs pour la matrice triangulaire inférieure
mot_mot_cosine <- Matrix::triu(mot_mot_cosine)

```

```

# La visualisation est dans quanteda, on doit donc créer un objet de son type
mot_mot_cosine <- quanteda::as.fcm(mot_mot_cosine)

# On sélectionne les tops mots
mot_par_mot_filtre_cosinus <- quanteda::fcm_select(mot_mot_cosine,
                                                    pattern = top_mots,
                                                    selection = "keep")

set.seed(123)
p <- textplot_network(mot_par_mot_filtre_cosinus, min_freq = 0.8) +
ggtitle(paste0("Co-occurrences des verbes dans le corpus ",
               ifelse(corpus ==1,"2017-2019", "2020")))
print(p)
ggsave(file=paste0(fig_dir,ifelse(corpus ==1,"figure_A5.jpg", "figure_A6.jpg")))
}

analyse_cooccurrence(1, 25)

```

Analyser cooccurrences cosinus-1.pdf

```
analyse_cooccurrence(2, 25)
```

Analyser cooccurrences cosinus-2.pdf

On voit que « permettre » était le plus central avant la pandémie et que « pouvoir » a pris sa place pendant la pandémie. Ce dernier peut être un marqueur d'incertitude. Nous regardons maintenant les phrases caractéristiques de cas où le verbe le plus central de chaque corpus est coprésent avec certains des verbes périphériques :

```

research_results_2017_2019 <- rbind(search_most_similar_paragraph(
  c("permettre", "expliquer"), 1),
  search_most_similar_paragraph(c("permettre", "améliorer"), 1),
  search_most_similar_paragraph(c("permettre", "démarquer"), 1),
  search_most_similar_paragraph(c("permettre", "étudier"), 1))

caption_txt <- "Corpus pré-COVID: phrases les plus caractéristiques
avec « permettre » et un autre verbe très fréquent"
if(outputFormat == "html"){
  research_results_2017_2019[order(-Cosine)] %>%
  kable(caption = caption_txt) %>% kable_styling("hover") %>%
  scroll_box(width = "100%", height = "200px")
} else{
  research_results_2017_2019[order(-Cosine)] %>%
  kable(caption = caption_txt, booktabs = TRUE) %>%
  kable_styling(full_width = F) %>%
  column_spec(2, width = "45em")
}

```

```

research_results_2020 <- rbind(search_most_similar_paragraph(
  c("pouvoir", "transmettre"), 2),
  search_most_similar_paragraph(c("pouvoir", "attraper"), 2),
  search_most_similar_paragraph(c("pouvoir", "combattre"), 2),
  search_most_similar_paragraph(c("pouvoir", "protéger"), 2))

```

Table 41: Corpus pré-COVID: phrases les plus caractéristiques avec « permettre » et un autre verbe très fréquent

Cosine	paragraphe
0.4993374	Vandal étudiera en sciences politiques.
0.4093563	# Les étudiants du Cégep se démarquent dans les concours scientifiques
0.3668156	Dans un communiqué publié lundi, l'Ifri s'est démarqué des propos tenus par le chercheur.
0.3578947	L'utilisation de ce format de CV permet aux chercheurs d'emploi de se démarquer des autres documents plus conventionnels. « Une cliente, après avoir été embauchée, s'est fait dire qu'elle avait un beau CV, qu'il se démarquait et qu'il était visuellement plus accessible », rapporte Mme Fortin.
0.3432108	Les fonds consentis permettront de soutenir pour les deux prochaines années la recherche scientifique et d'améliorer les installations de l'Observatoire.
0.3364744	«[La série d'exercices] permet de renforcer les muscles, améliorer l'équilibre, améliorer la santé cardiovasculaire, voire la santé cognitive», souligne la chercheuse Mylène Aubertin-Leheudre, également docteure en kinésiologie.
0.3335018	Le Canada se démarque à l'échelle du monde par les travaux de ses chercheurs sur les cellules souches qui permettent des avancées dans la lutte contre le cancer. Crédit : iStock
0.3237468	# Des étudiants du Cégep Beauce-Appalaches se démarquent dans les concours scientifiques
0.3186786	Annoncé comme une «romance de science-fiction», Zoe racontera l'histoire de deux collègues dans un laboratoire qui travaillent sur une nouvelle technologie qui doit permettre d'améliorer les relations romantiques.
0.2968588	Les scientifiques en viennent à la conclusion que l'exercice améliore la vie quotidienne et la mobilité des aînés qui souffrent de la maladie d'Alzheimer; il pourrait aussi améliorer leur équilibre et leur cognition générale.
0.2957577	Les scientifiques en viennent à la conclusion que l'exercice améliore la vie quotidienne et la mobilité des aînés qui souffrent de la maladie d'Alzheimer; il pourrait aussi améliorer leur équilibre et leur cognition générale.
0.2690983	Les résultats de cette étude permettent de jeter une nouvelle lumière sur le processus de reconnaissance visuelle de l'humain, expliquent les chercheurs.
0.2678008	Pour arriver à ces conclusions, les chercheurs ont étudié les effets de l'aspirine chez près de 130 000 personnes, entre 1980 et 2012.
0.2665186	\- Elle étudie actuellement en science politique à l'Université du Québec à Montréal (UQAM).
0.2664503	L'étude ne permet pas de conclure à un lien entre végétarisme et dépression. La littérature scientifique ne va généralement pas dans ce sens.
0.2653978	Shweta Todkar, qui est venue de l'Inde pour réaliser son rêve d'étudier dans une université canadienne, un geste difficile à poser pour une femme dans son pays précise-t-elle, est infirmière et étudie en sciences biomédicales pour la deuxième année. Les retards occasionnés vont l'obliger à prolonger ses études. Elle avait une bourse de 50 000 \$ pour venir étudier ici. Elle devra en demander une autre.
0.2584660	Il a expliqué qu'un effort «colossal» par les chercheurs, les travailleurs, les ingénieurs et les scientifiques a permis de réaliser une «vraie percée».
0.2541694	Cette démarche scientifique participative permet du coup à la personne d'en apprendre plus sur ce qu'est la science, à l'image de ceux qui participent aux recherches cliniques. Parce que, explique Sonya Anvar, se familiariser avec la manière dont la science se construit permet à cette personne de mieux comprendre ce qu'est vraiment la science.
0.2539175	Permettre l'avancement des connaissances scientifiques
0.2501539	Le site science24heures.com permettra de faciliter vos recherches, selon :


```
caption_txt <- "Corpus COVID: phrases les plus caractéristiques avec
« pouvoir » et un autre verbe très fréquent"
if(outputFormat == "html"){
  research_results_2020[order(-Cosine)] %>%
  kable(caption = caption_txt) %>% kable_styling("hover") %>%
  scroll_box(width = "100%", height = "200px")
} else{
  research_results_2020[order(-Cosine)] %>%
  kable(caption = caption_txt, booktabs = TRUE, longtable = TRUE) %>%
  kable_styling(full_width = F) %>%
  column_spec(2, width = "45em")
}
```

Table 42: Corpus COVID: phrases les plus caractéristiques avec « pouvoir » et un autre verbe très fréquent

Cosine	paragraphe
0.4345656	« Il n'y a aucune science, encore, qui dit que si on a attrapé la COVID-19 une fois, on est protégés de l'attraper une deuxième fois », a dit M. Trudeau.
0.4099765	Et pour le moment, il n'existe aucune preuve scientifique qu'un animal peut transmettre la COVID-19 à l'homme.
0.3650073	« Il n'y a aucune science encore qui dit que si on a attrapé la COVID-19 une fois, on ne peut l'attraper une deuxième fois, a-t-il soutenu. Tous les plans des provinces sont basés sur la protection. Personne ne se base sur l'immunité comme une mesure qui fait partie de la marche à suivre pour les prochains mois », a avancé Justin Trudeau.
0.3619600	« Des chercheurs pensent que même avec une respiration normale, on peut transmettre la maladie », ajoute-t-il.
0.3476097	"C'est toujours ma crainte, affirme Dominic Cardy. On peut voir aujourd'hui que les messages, la désinformation, le <code>_fake news_</code> , c'est aussi un type de virus qui peut se partager rapidement. On a besoin de le combattre. On a besoin d'une voix beaucoup plus forte des médecins, des scientifiques, des politiciens, de tout le monde dans la société. [...] La seule manière qu'on peut combattre les mensonges, c'est avec la vérité."
0.3464986	Dr William Li : Pour le moment, les scientifiques s'accordent à dire que les animaux de compagnie ne peuvent pas attraper la maladie ni nous la donner. Bonne nouvelle!
0.3352523	En date du 19 mars, l'état de la science n'indique pas si le virus peut se transmettre sexuellement.
0.3311074	Pour le moment, aucune donnée scientifique ne montre que les chiens ou les chats peuvent être infectés — ou transmettre — le virus.
0.3241021	Pour leur défense, les chercheurs chinois affirment qu'ils n'ont pu déterminer que courant janvier, étude scientifique à l'appui, que le virus pouvait se transmettre au sein de l'espèce humaine.
0.3236961	Se rangeant du côté de l'OMS, Justin Trudeau affirme « qu'il n'y a aucune science encore qui affirme que si on a attrapé la COVID-19 une fois, on ne peut l'attraper une deuxième fois ». La Dre Theresa Tam, administratrice en chef de la santé publique du Canada, croit même que « l'idée de générer une sorte d'immunité naturelle ne devrait pas être considérée ».
0.3193200	Selon le directeur de la santé, il n'est pas prouvé scientifiquement que les personnes qui ont été atteintes du coronavirus pourraient être protégées pour une deuxième vague:
0.3145192	Se pourrait-il, se demandent plusieurs scientifiques, que la mémoire immunitaire générée par ces coronavirus bénins aide à combattre le SARS-CoV-2 ?
0.3107638	Depuis le début, les autorités politiques dans le monde disent s'appuyer sur les données de la science pour combattre la pandémie.
0.2946508	La science ne peut pas donner un pourcentage de probabilité précis d'attraper le virus pour chaque activité : il est impossible d'envisager absolument tous les cas de figure qui peuvent se présenter en tenant compte de tous les paramètres qui pourraient avoir une influence. De plus, les modes de transmission du virus ne sont pas encore parfaitement élucidés.

0.2913032	"Les premières données scientifiques démontraient que de 20 à 25 % des personnes infectées étaient des cas asymptomatiques. Les gens se sont mis alors à porter des masques. Bien sûr, pour se protéger, mais aussi pour protéger les autres. Il y a eu une prise de conscience; on a compris qu'une personne pouvait en contaminer une autre, sans le savoir, en l'absence de symptômes", explique Lee Hyuk-Min, qui est également épidémiologiste.
0.2809443	Selon lui, ce n'est pas parce que le Chine a été le premier pays à déclarer l'apparition du coronavirus qu'elle est à l'origine de la pandémie. Il faut plutôt laisser les scientifiques faire leur travail pour qu'ils puissent mieux connaître le virus et le combattre, a-t-il affirmé
0.2763274	Les chercheurs enquêtent, car ils croient que les enfants détiennent un secret qui pourrait aider à protéger ou traiter les plus âgés.
0.2719048	Pour sa part, le magazine «Forbes» révèle que des chercheurs s'intéressent au plasma équestre pour combattre la pandémie.
0.2588521	« Considérant que des preuves scientifiques confirment que des personnes sans symptômes peuvent transmettre le virus de la COVID-19, il est maintenant indiqué de protéger les autres à proximité », peut-on lire dans le communiqué du Ministère de la Santé et des Services sociaux diffusé vendredi après-midi.
0.2444896	La vaccination demeure le moyen le plus efficace de se protéger contre certaines maladies, en plus d'être reconnue comme sécuritaire par la communauté scientifique, a récemment soutenu la ministre de la Santé, Danielle McCann. Plusieurs vaccins doivent être donnés durant l'enfance, certains dès l'âge de 2 mois, afin que les petits soient protégés rapidement.

Cette analyse ne permet évidemment pas de tirer de conclusions même si les phrases de 2020 semblent généralement exprimer plus d'incertitude. Nous utilisons une stratégie plus systématique dans la section suivante.

Proportion des phrases exprimant de l'incertitude

Nous cherchons maintenant à savoir si la proportion de phrases exprimant de l'incertitude dans les paragraphes liés à la science est similaire dans nos deux corpus d'articles (avant et pendant la pandémie). Toute phrase incluant au moins une instance respectant une des conditions suivantes est considérée comme exprimant de l'incertitude :

- Un verbe au mode conditionnel
- Le verbe pouvoir, sauf à l'infinitif
- Les lemmes suivants : possible, possiblement, probable, probablement, improbable, suspect (comme adjectif), suspicion
- Un de ces verbes : penser, croire, espérer, suggérer, questionner, soupçonner

Une chose que cette technique simple ne capture pas est la négation qui annule l'incertitude. Par exemple, "possible" exprime de l'incertitude, mais "pas possible" exprime de la certitude. Nous devrions avoir des règles plus complexes pour détecter cette annulation de l'incertitude. Notre présupposition est que la proportion d'annulations de l'incertitude sur l'incertitude apparente totale n'est pas plus élevée dans un de nos deux corpus. Si ce présupposé est adéquat, la simplicité de notre règle nous fait surestimer la proportion de phrases exprimant l'incertitude, mais ce biais vers le haut affecte les deux corpus et n'invalide donc pas notre test de différence de proportion.

```
#### TEST de la proportion des phrases avec marqueurs d'incertitude ###

test_for_uncertainty <- function(tagged_tokens)
{

  data_for_test <- tagged_tokens[, list( nb_sentences_w_uncertainty =
    .SD[ grepl("Mood=Cnd",feats) # verbs or auxiliaries in the conditional mood
```

```

      # verb "pouvoir" (for can, may, could...) except at the infinitive
      | (lemma == "pouvoir" & upos == "VERB" & token != "pouvoir")
      | lemma == "possible" | lemma == "possiblement" # around the possible
      # around the probable:
      | lemma == "probable" | lemma == "probablement" | lemma == "improbable"
      | (lemma == "suspect" & upos == "ADJ") | lemma == "suspicion" # something suspicious
      # verbs of uncertainty:
      | (upos == "VERB" & (lemma == "penser" | lemma == "croire" | lemma == "espérer" |
        lemma == "suggérer" | lemma == "questionner" | lemma == "soupçonner"))
      , paste0(doc_id, "_", sentence_id) %>% unique() %>% length(),
      nb_total_sentences = paste0(doc_id, "_", sentence_id) %>% unique() %>% length()
    ), by = since_COVID]

prop.test(data_for_test$nb_sentences_w_uncertainty,
          n = data_for_test$nb_total_sentences, alternative = "less")
}

#Test on article corpus
load(paste0(data_folder, "dt_tagged_tokens_science.RData"))
dt_tagged_tokens_science[corpus == 1, since_COVID := TRUE][
  is.na(since_COVID), since_COVID := FALSE]
#To get rid of some sentences that are misidentified by udpipe, we keep those upos only.
dt_tagged_tokens_science <- dt_tagged_tokens_science[
  upos %in% c("NOUN", "ADJ", "VERB", "AUX", "PROPN", "X", "ADV", "PRON", "DET")]
articles_uncertainty_results <- test_for_uncertainty(dt_tagged_tokens_science)
articles_uncertainty_results

##
## 2-sample test for equality of proportions with continuity correction
##
## data: data_for_test$nb_sentences_w_uncertainty out of data_for_test$nb_total_sentences
## X-squared = 226.34, df = 1, p-value < 2.2e-16
## alternative hypothesis: less
## 95 percent confidence interval:
## -1.00000000 -0.03259453
## sample estimates:
## prop 1 prop 2
## 0.1474629 0.1842299

rm(dt_tagged_tokens_science)

```

Session R utilisée

```

sessionInfo()

## R version 4.2.2 Patched (2022-11-10 r83330)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 18.04.6 LTS
##
## Matrix products: default
## BLAS: /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.7.1

```

```

## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.7.1
##
## locale:
## [1] LC_CTYPE=C.UTF-8      LC_NUMERIC=C           LC_TIME=fr_CA.UTF-8
## [4] LC_COLLATE=C.UTF-8    LC_MONETARY=C.UTF-8    LC_MESSAGES=C.UTF-8
## [7] LC_PAPER=C.UTF-8      LC_NAME=C              LC_ADDRESS=C
## [10] LC_TELEPHONE=C        LC_MEASUREMENT=C.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] parallel stats      graphics grDevices utils      datasets methods
## [8] base
##
## other attached packages:
## [1] quanteda.textplots_0.94.1 kableExtra_1.3.4
## [3] directlabels_2021.1.13    viridis_0.6.1
## [5] viridisLite_0.4.0         mgcv_1.8-41
## [7] nlme_3.1-160              sandwich_3.0-1
## [9] stargazer_5.2.3           text2vec_0.6.1
## [11] Matrix_1.5-1             splitstackshape_1.4.8
## [13] udpipe_0.8.9             lubridate_1.7.10
## [15] zoo_1.8-10               doParallel_1.0.17
## [17] iterators_1.0.13         foreach_1.5.1
## [19] tidytext_0.3.1          knitr_1.41
## [21] tm_0.7-8                 NLP_0.2-1
## [23] topicmodels_0.2-12       anytime_0.3.9
## [25] fuzzyjoin_0.1.6          ggrepel_0.9.1
## [27] stringr_1.4.0            tidyr_1.1.3
## [29] data.table_1.14.0        mongolite_2.6.1
## [31] dplyr_1.0.9              ggplot2_3.3.5
##
## loaded via a namespace (and not attached):
## [1] webshot_0.5.2           httr_1.4.4              SnowballC_0.7.0
## [4] tools_4.2.2            utf8_1.2.2              R6_2.5.1
## [7] DBI_1.1.1              colorspace_2.0-2        withr_2.5.0
## [10] tidyselect_1.1.1       gridExtra_2.3           compiler_4.2.2
## [13] cli_3.4.1              rvest_1.0.1            lgr_0.4.3
## [16] xml2_1.3.2             rsparse_0.5.0          slam_0.1-48
## [19] scales_1.1.1          quadprog_1.5-8         systemfonts_1.0.4
## [22] digest_0.6.30         rmarkdown_2.18         svglite_2.0.0
## [25] RhpcBLASctl_0.21-247.1 pkgconfig_2.0.3        htmltools_0.5.2
## [28] fastmap_1.1.0         rlang_1.0.6            rstudioapi_0.13
## [31] generics_0.1.2        jsonlite_1.7.2         tokenizers_0.2.1
## [34] magrittr_2.0.1        modeltools_0.2-23      Rcpp_1.0.7
## [37] munsell_0.5.0         fansi_0.5.0            lifecycle_1.0.1
## [40] stringi_1.7.4         yaml_2.2.1             grid_4.2.2
## [43] mlapi_0.1.1           crayon_1.4.1          lattice_0.20-45
## [46] splines_4.2.2         pillar_1.6.2          stopwords_2.2
## [49] codetools_0.2-18     stats4_4.2.2          fastmatch_1.1-3
## [52] glue_1.6.2           evaluate_0.18         RcppParallel_5.1.4
## [55] float_0.3-0          vctrs_0.4.1          quanteda_3.1.0
## [58] gtable_0.3.0         purrr_0.3.4          assertthat_0.2.1
## [61] xfun_0.35            janeaustenr_0.1.5     tibble_3.1.4
## [64] ellipsis_0.3.2

```