

Title : Kittens and Jesus : What would remain in a newsless Facebook?

Author : Jean-Hugues Roy, École des médias, Université du Québec à
Montréal, roy.jean-hugues@uqam.ca

Abstract : This paper examines what would remain on Facebook if news content was removed, like the company temporarily did in Australia, in February 2021. Using a corpus of 3.3 million Facebook posts published in French in 2020 in four countries (Belgium, Canada, France and Switzerland), it compares media content to non-media content by submitting the text of the posts to three computational analyses : basic n-gram comparison, χ^2 residuals and topic modeling. Two distinct spheres are defined within Facebook content : a “public interest” sphere, made up of media pages, and a “public’s interest” sphere, made up of non-media pages. Religious content and “inspirational” “feel good memes” are found to be most characteristic of a newsless Facebook.

Keywords : Facebook, news, journalism, Belgium, Canada, France, Switzerland, text mining, computational methods

Licence :

```
<a rel="license"
href="http://creativecommons.org/licenses/by-nc-sa/4.0/"></a><br />This work is licensed under a <a rel="license"
href="http://creativecommons.org/licenses/by-nc-sa/4.0/">Creative
Commons Attribution-NonCommercial-ShareAlike 4.0 International
License</a>.
```

Introduction : The Australian Standoff

It was probably one of the biggest gambits in Internet history. On February 17, 2021, Facebook removed all Australian news content from its platform. Facebook users all over the world could no longer access articles from Australian media organizations, nor could they share URLs from those publications on their personal profiles, or on groups or pages they participate in. Even *The Conversation*, an international website used by academics to publish research findings, was barred from the platform because it is headquartered in Melbourne. The ban also worked the other way around as Facebook's Australian users could no longer see or share any news, wherever it originated from in the world.

Facebook was reacting to Canberra's [News Media and Digital Platforms Bargaining Code](#). The legislation forces Google and Facebook to negotiate deals with Australian media companies to compensate for news content in search results or users' posts. While it aims to help finance Australian journalism, the Code has been denounced by both web giants as "unworkable" (Facebook, Inc., 2020b, p. 6; Google, Inc., 2020, p. 23).

In January, 2021, Google threatened, “if the Code were to become law in its current form”, to “have no real choice but to stop making Google Search available in Australia” (Silva, 2021). But it refrained from a total blackout after negotiating changes in the Code with Australian Treasurer Josh Frydenberg (Leaver, 2021).

Facebook, on the other hand, surprised everyone by nuking the news in Australia. The ban, denounced by Prime Minister Scott Morrison and others, was however short lived. On Feb. 22, the Palo Alto company announced it too had convinced Australian lawmakers to amend their Code before enacting it. Basically, both platforms agreed to sign deals with Australian media companies before the Code became law.

In the end, everybody claimed they have won. Government and news businesses were happy Google and Facebook diverted some of their revenue to support journalism. Both web behemoths were happy to escape the more constraining aspects of the Code.

While the conflict has been resolved in Australia, other countries contemplating similar legislation wonder if they would face the same kind of retaliation from platforms. Facebook, in particular, has said it would not hesitate to remove news content in Canada, even if it was as a last resort (Reynolds, 2021; Roy, 2021a).

Which leads to wonder what would Facebook look like if news content was permanently removed? What would remain on Facebook users'...

"News Feed" if it was devoid of news? This is the research question this article tries to answer.

Methods : Using four French-language countries as an example

As a journalism prof working in Montreal, the most natural "terrain" to conduct this research project is French-language Canada. But to establish whether my findings are restricted or not to my home area, I needed to examine what a news ban would look like in more than one country. France, Belgium and Switzerland, nations with sizeable French-language mediascapes, were selected. In all four countries, between 38 and 55 percent of the population mentions social media as a source for news. Everywhere, Facebook is the top social media they use (Newman *et al.*, 2021, pp. 67, 77, 107, 119).

Sampling

I used CrowdTangle to download a sample of Facebook posts from all four countries. CrowdTangle is a Facebook-owned tool to discover public content on social media platforms. It is made available to academics through a [partnership between Facebook and Social Science One](#).

One of the ways to access CrowdTangle data is by using a web-based dashboard. It enables country-specific¹ searches, but for content published on Facebook pages only. For any given search, CrowdTangle can return, as of early 2021, as much as 300,000 posts.

I thus used a CrowdTangle dashboard to download the 300,000 posts made on public pages which generated the most interactions in all four countries covered by this research, for each and every month of the year 2020. After removing duplicates, this gave me an initial sample of 13.4 million posts. I then filtered it three times (see Table 1, below).

First, I only kept posts in French using langId (Lui, 2016), langDetect (Danilak, 2020) and polyglot (Al-Rfou, 2016), three python language-detection libraries. When two or three agreed on a given language, the post was classified in that language, otherwise it was classified as "unknown". This reduced my sample to 5.3 million posts, which is not surprising given that French is not spoken by the majority in Canada or Switzerland, and given that even in France many pages publish content in English to appeal to an international audience.

Second, because I wanted to analyse the text of those posts (see Findings, below), I only kept those which had 100 characters or more.

This includes memes, as CrowdTangle is able to extract text from images using OCR. After running this filter, my sample was further decreased to 4.2 million posts.

Third, I wanted to reduce the number of pages in my sample because I had to manually classify media vs. non-media pages. This work being so tedious, I only kept pages which had 100 posts or more in my initial sample. This left me with a final sample of approximately 3.3 million posts.

Table 1 - Sample filtering

Country	Posts in initial sample	French-language posts	Posts with 100 characters of text or more	Posts in pages with 100 posts or more (final sample)
Belgium	3,597,336	1,346,988	1,126,524	901,931
Canada	3,599,699	621,509	528,750	398,071
France	3,599,310	2,779,406	2,031,998	1,550,768
Switzerland	2,648,595	529,032	488,653	420,671
All four	13,444,940	5,276,935	4,175,925	3,271,441

Classification

The next step was to classify media and non-media pages. I certainly could not trust Facebook's definition. "Consider [...] the absurdly wide scope of the 'Media/News' category on Facebook", writes Columbia's Emily Bell (2021). She quotes Gordon Crovitz, former *Wall Street Journal* publisher and cofounder of NewsGuard, a system that rates the credibility of news sources. According to him, platforms

have demonstrated their “fundamental failure to understand the core concepts of journalism”. So I needed to craft my own criteria. I used the following. Media pages were :

- General news dailies and their regional or thematic pages (science, opinion, etc.) if any;
- Regional or local general weeklies;
- Newsmagazines;
- Talk radio programs, stations or networks, along with their regional or thematic pages, if any;
- Television news programs, stations or networks, along with their regional or thematic pages, if any;
- National, local or hyperlocal general news websites;
- Media or programs specializing in politics, business/economy, culture, science or sports.

I excluded, and thus classified in the non-media group, pages for :

- Individual journalists;
- Media or programs specializing in only one topic (soccer as opposed to all sports, literature as opposed to all culture or mining as opposed to all business);
- Media or programs specializing in celebrities, lifestyle, cooking or fashion;
- Viral content agregators or pages;
- Corporate headquarters of media;
- Non-news television programs (sitcoms, series, comedy);
- Music radio programs, stations or networks;
- Pseudomedia used by corporations or special interests;
- News parody;
- Reinformation (sites repurposing authentic news with a specific political agenda, as defined by Blanc (2016));

- Misinformation, disinformation or fake news;
- All other non news pages.

Table 2 - Media and non-media subcorpora, by country

Country	Non-media		Media			
	Posts	Pages	Posts	% of total posts	Pages	% of total pages
Belgium	609,869	1,654	292,062	32.4%	107	6.1%
Canada	274,195	705	123,876	31.1%	108	13.3%
France	1,178,885	3,089	371,883	24.0%	318	9.5%
Switzerland	270,411	598	150,260	35.7%	56	8.6%
All four	2,333,360	6,046	938,081	28.7%	589	8.9%

The result of this classification process is presented in Table 2.

The percentage of media pages in Canada is greater perhaps because I am more familiar with the news ecosystem there. In other countries, when I was unable to establish the news character of a page, I left it out. This probably exclude smaller or local media. But the main news organizations are included. The full list of pages, both media and non-media, can be found in the Technical appendix.

It produced eight subcorpora : a subset of media and non-media Facebook posts for each country. On average, posts from media pages represent 28.7 percent of the total 3.3 million posts in my final sample.

To answer the research question at the heart of this article, I needed to compare, within each country, media and non-media corpora. To do so, I performed three different textual analyses.

Analysis 1 : n-grams

Using spaCy (Honnibal and Montani, 2017, version 3.0.6), a natural language processing python library, and its [fr_core_news_md](#) model for French, I first did a preprocessing of the textual elements of the posts (stopword removal, lemmatization, lowercasing, etc.). It was followed by basic n-gram extractions for single lemmas, bigrams and trigrams to see which are most frequent in each subcorpus. The total number of lemmas in all preprocessed subcorpora was more than 106-million.

This first analysis is based on the “bag of words” approach which has been used for decades in text mining. It assumes that “words appear independently [in a document] and their order is immaterial” (Huang, 2008, p. 50). The weight of each word is its frequency, “which means terms that appear more frequently are more important and descriptive for the document” (Ibid.). The same applies to bigrams and trigrams.

The approach has however been perfected by different weighing techniques to achieve better performance, according to the task at

hand (Li et al., 2016, p. 1592). A corpus made of Facebook posts gives us a novel way of weighing n-grams : the total number of interactions generated by the posts they appear in.

This method also better reflects how Facebook users have reacted to a given content. Measuring only frequencies could lead to make certain terms misleadingly salient. For example, an election campaign team or a marketing firm could publish hundreds of posts on Facebook, making occurrences of a given politician or a given product much more common. In the context of Facebook, a "space that celebrates the primacy of the emotional, the impulsive, over argumentation" (Cepernich, 2016), a market of emotions where the engagement of users is what gives meaning to content (and value to Facebook), it appears methodologically justified to use interactions rather than simple frequencies.

Analysis 2 : χ^2 residuals

But there are other ways of comparing two corpora that go beyond simply counting n-grams. For the last 20 years, or so, computational linguists have been tackling the issue. In this quest, Adam Kilgarriff's work stands out as he has considered many statistical methods. He concluded that " χ^2 [chi-squared] is [...] a suitable measure for comparing corpora, and is shown to be the best measure of

those tested" (2001, p. 258). The χ^2 test tells us whether the observed frequency of a given n-gram in one corpus is greater than its expected frequency.

A few years later, Oakes and Farrow pushed χ^2 further by calculating the standardized residual (2007, p. 89) using this formula :

$$(\text{observed} - \text{expected}) / \sqrt{\text{expected}}$$

Dividing the difference between observed and expected frequencies by the square root of the expected frequency, this formula produces a positive or a negative value for each n-gram, by corpus. The greater the positive value, the more characteristic this n-gram is to that corpus. A great negative value means the opposite and a value close to zero means the word is equally present in both corpora.

Table 3 gives an example of how to calculate standardized residuals for three words taken from the media and non-media Facebook posts in Canada : *travail* (work), *pandémie* (pandemic) and *recette* (recipe).

Table 3 - Standardized residuals calculations for three words from the Canadian subcorpus.

word	observed		sum	expected		residuals	
	media	non-media		media	non-media	media	non-media

<i>travail</i>	4,414	12,037	16,451	4,473.40	11,977.60	-0.89	0.54
<i>pandémie</i>	14,476	9,887	24,363	6,624.85	17,738.15	96.46	-58.95
<i>recette</i>	1,373	20,995	22,368	6,082.36	16,285.64	-60.38	36.9
...				
sum of all words	3,904,931	10,455,525	14,360,456				

Table 2 shows how the expected frequency is computed : the word *pandémie*, for example, appears 24,363 times in both subcorpora. Given that 27.3 percent of all words are in the media subcorpus (3.9 million / 14.4 million), the expected frequency of *pandémie* in this subcorpus is 27.3 percent of 24,363, or 6,624.85.

Table 2 also shows that the standardized residuals indicate that the word *pandémie* is more distinctive of the media corpus, that the word *recette* rather defines the non-media corpus and that the word *travail*, with values very close to zero in both subcorpora, is characteristic of neither.

Standardized residuals are a “statistical tool for exploratory research that allows for the identification of words that deserve deeper analysis, and not [...] an instrument of confirmatory analysis”, warns Bestgen (2014). Poudat and Landragin also point out that the χ^2 test is based on the normal distribution. Insofar as most words are infrequent in most corpora, the χ^2 test, and its residuals, are thus an “unreasonable” way to analyse text... unless you have “an

important volume of data" (Poudat and Landragin, 2017, p. 169). Given that a corpus containing more than 1 million words can be considered a "large corpus" (Ibid., p. 11), that the smallest of my subcorpora contains 3.6 million terms (trigrams from Canadian media pages), I will consider reasonable to use the χ^2 residuals to compare media and non-media corpora as part of the second textual analysis in the Findings section, keeping in mind that they are only "indicator of the potential interest of each of the numerous vocabulary differences between the corpora" (Bestgen, 2014).

I also chose to weigh the χ^2 residuals by the number of interactions. Instead of calculating them using the frequency of each term (lemma, bigram or trigram), I've used the sum of the interactions of the posts in which they were found. As discussed above, this gives a measure that is more adapted to the corpus used (Facebook posts).

Analysis 3 : topic modeling

Finally, in order to further explore the differences between both supcorpora, the third textual analysis I did was topic modeling (Lin, 1995; Ramage *et al.*, 2009). This technique clusters words which often appear together in documents, therefore revealing that they are probably somewhat semantically related. The models used also score each word to weigh a given topic relative to others. In a way, topic

modeling lets the corpus speak by itself... even though the model doesn't label topics for us. It is our job, as researchers, to name them.

I used the BERTopic algorithm (Grootendorst, 2020), based on the recently released Bidirectional Encoder Representations from Transformers (BERT) model (Devlin et al., 2019). BERTopic provides two sets of embedding models : one for the English language and the other for multilingual documents. None are specifically adapted to French-language documents.

It is however possible to ask BERTopic to use other models for word embeddings. I used three : the aforementioned "fr_core_news_md" model by spaCy, which includes a French regional newspaper (*L'Est républicain*) in its training corpus; the "flaubert_large_cased" model from the FlauBERT project (Le et al., 2020), which includes some belgian newspapers in its training corpus (Tiedemann, 2012); and the "camembert-base" model from the CamemBERT project (Martin et al., 2020), trained on the French subcorpus of the OSCAR corpus, which does not seem to include any news content.

I ran BERTopic four times on each of all eight subcorpora with different parameters. The first run was using spaCy's model and asking BERTopic to provide 20 topics with 20 bigrams each. The

remaining three runs (once with each model) asked BERTopic to provide 12 topics with 8 single words or bigrams. The use of three different models (spaCy's, FlauBERT's and CamemBERT's) leads me to believe the resulting analysis will be all the more robust.

This analysis was performed on the "raw" text of Facebook posts without lemmatization or removal of stopwords. Also, I did not weigh by interactions the terms used by the algorithm. However, instead of running the algorithm on single words, which is a limitation in most topic modeling approaches (Boumedyen Billami *et al.*, 2020, p. 156), I used it with bigrams on my first run with the spaCy model. In the following three runs, I let BERTopic the possibility of using unigrams or bigrams. Joining names (such as "emmanuel macron") or entities (like "états unis" or "real madrid") often provided richer topics, facilitating their interpretation.

Topic modeling is extremely memory intensive for computers. With subcorpora as large as 37-million bigrams, in the case of non-media pages in France, I had to divide each by month, otherwise my code would crash. This strategy, though it meant each run took several hours, enriched the analysis, in the end, because it enabled me to explore one month at a time and tell the story of how topics in both media and non-media "spheres" within Facebook evolved over the year 2020.

Findings

The media loves Facebook

The first surprise is the sheer amount of content francophone news media publish on Facebook. Table 4 shows the top 50 pages that published the most posts in 2020 in the four countries covered by this study; 42 of them are media pages!

Table 4 – Top-50 French-language pages by number of posts in Belgium, Canada, France and Switzerland (2020); media pages shown with grey background.

Page	Facebook ID	Country	Nb of posts	Type
RTL Info	RTLInfo	Belgium	20,318	media
Le Soir	lesoirbe	Belgium	17,656	media
BFMTV	BFMTV	France	16,384	media
Le Parisien	leparisien	France	16,290	media
DH.be	dhnet	Belgium	15,049	media
La Libre	lalibre.be	Belgium	14,673	media
Le Nouvelliste	nouvelliste	Switzerland	14,466	media
ArcInfo	arcinfofanpage	Switzerland	13,554	media
La Parole Vivante	laparolevivante	Canada	13,512	religion
RTBF Info	rtbfinfo	Belgium	13,338	media
Le Vif	Leviflexpress	Belgium	12,834	media
lavenir.net	lavenir.net	Belgium	12,292	media
Sudinfo.be	sudpresse	Belgium	11,718	media
Le Journal de Montréal	jdemontreal	Canada	11,660	media
LN24	LN24LesNews24	Belgium	11,163	media
TVA Nouvelles	TVAnouvelles	Canada	11,041	media
7sur7.be	7sur7.be	Belgium	10,083	media
La Côte	LaCoteJournal	Switzerland	9,998	media

La Provence	laprovence	France	9,726	media
24heures	24heures.ch	Switzerland	9,615	media
Le Figaro	lefigaro	France	9,582	media
Foot Mercato	footmercato	France	9,464	football
20 Minutes	20minutes	France	9,160	media
Franceinfo	franceinfo	France	8,762	media
Radio-Canada Information	radiocanada.info	Canada	8,740	media
Epoch Times Paris	EpochTimesParis	France	8,537	misinformation
LeMatin.ch	lematin.ch	Switzerland	8,277	media
Metro Belgique	metrobelgique	Belgium	8,256	media
La Presse	LaPresseFB	Canada	8,252	media
RTBF Sport	RTBFSport	Belgium	8,167	media
Tribune de Genève	tdg.ch	Switzerland	8,145	media
Ouest France	ouestfrance	France	8,046	media
Le Monde	lemonde.fr	France	8,011	media
Heidi.news	Heidi.news	Switzerland	7,916	media
beIN SPORTS France	beINSPORTSFrance	France	7,789	media
CNEWS	CNEWSofficiel	France	7,788	media
FRANCE 24	FRANCE24	France	7,689	media
Walfoot	Walfoot	Belgium	7,583	football
L'EQUIPE	lequipe.fr	France	7,453	media
RT France	RTFrance	France	7,417	misinformation
Le Journal de Québec	JdeQuebec	Canada	7,338	media
L'Express	LExpress	France	7,245	media
Nordpresse	nordpressed	Belgium	6,690	parody
RTSinfo	RTSinfo	Switzerland	6,590	media
TF1 Le JT	TF1leJT	France	6,345	media
RFI	RFI	France	6,293	media
France Bleu	reseau.francebleu	France	6,139	media
Lions de l'Atlas	lionsdelatlas.net	Belgium	6,099	football
Horoscope du jour	horoscopedujour.ch	Switzerland	5,803	astrology
20 minutes online	20minutesonline	Switzerland	5,792	media

This might be explained by the fact media organizations have complete teams dedicated to repurposing content for Facebook and other platforms (Bell et al., 2017). But so do record companies, viral

content producers, government departments and other organizations keen on maintaining a regular presence on Facebook. So it seems fair to say that media are the single most active type of organization in the French-speaking areas of Facebook.

Nechushtai (2018, p. 1052) has argued journalism organizations have been “infrastructurally captured” by Facebook. Not only do they depend on the Palo Alto company for reaching their readers, but they are driven to adapt “news production more broadly [...], to comply with the logics, norms, or business strategies of external platforms” (ibid.). In the United States, more recent studies report media organizations are distancing themselves more and more from digital platforms. Tech giants’ increasing power, the greater scrutiny on their actions and the growing calls for regulation has translated into caution, even distrust, towards platforms. “Publishers are attempting to regain control over the future of their business” (Rashidian *et al.*, 2020, p. 8). This movement is however not at all apparent in my subcorpora of French-language media posts. The capture of francophone journalism by Facebook seems still strong and, dare I say, borders on rapture.

First glance at top non-media pages

Let's now take a look at what type of pages would remain if media pages were gone. For each country, I sorted non-media pages by number of posts, by number of interactions and by average interactions by post. All twelve tables would be too long to reproduce here (tables 5 to 8, with the top-20 pages by average interactions, by country, are presented below). But what stands out of this exercise is that comedy and humor pages, artists and fan pages (for musicians most of the time) and what I'd call "feel good meme pages" are the most common.

That is not surprising, given that these types of content are the bread and butter of viral pages. The fact they appear on top of tables sorted by interactions is expected. That's why more thorough textual analyses are needed to dig deeper into these corpora. Before doing that, though, I want to discuss three elements worthy of notice in the top non-media pages.

Table 5 - Top-20 non-media French-language pages in Belgium (with more than 1 post per day), by average interactions per post.

Page	Facebook ID	Nb of posts	Interactions	Interactions /post	Type
Merveille du monde	merveilledumondejaninevero	1,565	21,344,119	13,638.4	Feel good memes
Le Grand Cactus - RTBF	LeGrandCactus	371	4,232,858	11,409.3	TV show
The Voice Belgique -	thevoicebelgique	370	4,179,007	11,294.6	TV show

RTBF					
Brigade des nurses	lesinfirmieres	579	3,630,483	6,270.3	Comedy
Jérôme de Warzée	jeromedewarzee	406	2,481,480	6,112.0	Artist
Guillaume Corpard	GuillaumeCorpard.officiel	514	2,556,520	4,973.8	Artist
Gregory Lemarchal	gregorylemarchal112	519	2,461,591	4,742.9	Fan page
Pairi Daiza	JardinDesMondes	1,032	4,637,333	4,493.5	Animals
Tarmac	TarmacRTBF	836	3,624,764	4,335.8	TV show
Bisoutendresse	bisoutendresse	2,045	8,666,761	4,238.0	Feel good memes
Météo-Mons	Meteo.mons	720	2,939,261	4,082.3	Weather
Samuel Movie	ActeurSamuelMovie	1,128	4,527,459	4,013.7	Artist
Pape François	PapeFrancoisVatican	711	2,808,529	3,950.1	Religion
David Antoine	DavidAntoine	906	3,249,951	3,587.1	Artist
Woman and material & Aimee Virgile Makougoum	Artisteaucameroun	475	1,356,705	2,856.2	Artist
Permavenir	Permavenir	683	1,655,625	2,424.0	NGO
PTB	ptbbelgique	904	2,138,859	2,366.0	Political party
<i>Fuck Love</i>	IHateYouAndOnlyU	391	887,657	2,270.2	Comedy
Djanii Alfa	Djanii.X	447	963,807	2,156.2	Artist
Cristiano Ronaldo - France	Ronaldo7France	987	1,942,312	1,967.9	Fan page

Table 6 - Top-20 non-media French-language pages in Canada (with more than 1 post per day), by average interactions per post.

Page	Facebook ID	Nb of posts	Interactions	Interactions/post	Type
Justin Trudeau	JustinPJTrudeau	480	6,973,554	14,528.2	Political figure
François Legault	FrancoisLegaultPremierMinistre	647	5,366,866	8,295.0	Political figure
African Heroes	africanheroesmagazine	1,171	7,013,702	5,989.5	Special interest
Indochine officiel	Indochineofficiel	402	1,597,326	3,973.4	Artist
FFL -	FFLose	916	3,083,841	3,366.6	Comedy

Fédération Française de la Lose					
Éric Duhaime	eduhaime	385	1,263,369	3,281.5	Political figure
La parfaite maman cinglante	laparfaitemamancinglante	1,153	3,604,682	3,126.4	Special interest
La solution est en vous	lasolutionestenvous	970	2,935,008	3,025.8	Feel good memes
Ministère Paul Mukendi	ministerepaulmukendi	418	1,223,633	2,927.4	Religion
L'Anti-Média	LAntiMedia	1,017	2,639,922	2,595.8	Misinformation
Occupation Double	occupationdouble	621	1,088,219	1,752.4	TV show
Gabriel Nadeau-Dubois	GNadeauDubois	421	720,253	1,710.8	Political figure
Queen Fumi	queenfumiofficiel	419	712,755	1,701.1	Artist
EMCI TV	emcitr	1,316	2,207,163	1,677.2	Religion
Québec Niaiserie	quebecniaiserie1	598	951,948	1,591.9	Comedy
La Parole Vivante	laparolevivante	13,512	21,379,356	1,582.2	Religion
Ricardo Cuisine	ricardocuisine	494	752,572	1,523.4	Food
LES ETHNIES DE LA COTE D'IVOIRE ET D'AFRIQUE	ethniesdeCI	1,101	1,643,412	1,492.7	Special interest
District 31	ICIDistrict31	868	1,252,896	1,443.4	TV show
Le Revoir	JournalLeRevoir	900	1,298,617	1,442.9	Parody

Table 7 - Top-20 non-media French-language pages in France (with more than 1 post per day), by average interactions per post.

Page	Facebook ID	Nb of posts	Interactions	Interactions/post	Type
Le Meilleur du Football	LemeilleurduFootball92	3,691	146,312,371	39,640.3	Football
Mathieu Rivrin • Photographe de Bretagne	Mathieu.Rivrin.photographies	368	7,120,390	19,348.9	Artist
Les marseillais W9	LesMarseillaisW9	368	6,434,885	17,486.1	TV show

Nostalgies 60'-70'-80'	Nostalgies607080	692	12,095,023	17,478.4	Nostalgia
30 Millions d'Amis (Officiel)	30millionsdamis	663	10,400,103	15,686.4	Animals
Marine Le Pen	MarineLePen	604	8,944,603	14,808.9	Political figure
Demotivateur	demotivateur	5,711	80,736,401	14,137.0	Comedy
Né pour brûler la gomme	Nepourbrulerlagomme	516	7,230,470	14,012.5	Special interest
Imineo	imineoTV	505	6,939,097	13,740.8	Special interest
One Voice	onevoiceanimal	799	10,309,383	12,902.9	Animals
Madame Connasse	Madameconnasse0987654321	2,879	36,341,190	12,622.9	Comedy
La vraie démocratie	lavraiedemocratie	1,071	13,300,634	12,418.9	Reinformation
Corsica ile Magique	CorsicaileMagique	669	8,296,473	12,401.3	Tourism
L214 Ethique et Animaux	l214.animaux	383	4,663,064	12,175.1	Animals
Génération 80's	mageneration80	842	10,143,087	12,046.4	Nostalgia
M6	M6	870	10,106,366	11,616.5	TV channel
Nostalgie Du Football	NostalgieDuFootball	524	6,074,626	11,592.8	Football
Le Pépère	lepepere	752	8,554,827	11,376.1	Comedy
JEAN-MARIE BIGARD PAGE OFFICIELLE	levraiJeanmariebigard	1,197	13,259,818	11,077.5	Artist
Ina.fr	Ina.fr	1,408	15,450,584	10,973.4	Institution

Table 8 - Top-20 non-media French-language pages in Switzerland (with more than 1 post per day), by average interactions per post.

Page	Facebook ID	Nb of posts	Interactions	Interactions/post	Type
Les temps sont durs pour les rêveurs	bureaudesrevesbrises	410	2,682,193	6,541.9	Comedy
Trust My Science	TrustMyScience	2,579	5,641,255	2,187.4	Special interest
La Bible	Bibles	2,459	2,995,389	1,218.1	Religion
Vanessa_beauti	vanessabeauti	440	526,606	1,196.8	Business
Darius	dariusrochebin	371	320,604	864.2	Journalist

Rochebin					
Forum Économique Mondial	weffrancais	2,643	1,171,794	443.4	International organization
Anthony Desonpere	ADeSonPere	560	228,334	407.7	Political figure
Hôpitaux Universitaires de Genève	hopitaux.universitaires.geneve	456	174,366	382.4	Institution
Te rappelles-tu	rappelletoi	5,194	1,934,293	372.4	Nostalgia
Les souvenirs de Dakar	lessouvenirsdusenegal	520	189,883	365.2	Special interest
Ambassade de Suisse en France	AmbassadeSuisseParis	408	147,891	362.5	Institution
Ville de Neuchâtel	neuchatelville	724	259,582	358.5	Institution
FCBarcelone French	FCBarceloneFre	611	216,280	354.0	Football
Ville de Genève - Officiel	villegeneve.ch	397	122,243	307.9	Institution
Animal mon Egal	animal.mon.egal	947	260,594	275.2	Animals
Objectif Fitness	ObjectifFitnessOfficiel	473	124,453	263.1	Feel good memes
Armée suisse	armee.ch.fr	576	138,559	240.6	Institution
Servette FC	servettefootballclub	973	185,946	191.1	Football
228 & AfrikaExclu	228exclu	1,243	214,458	172.5	Special interest
loisirs.ch	www.loisirs.ch	1,174	180,057	153.4	Tourism

First, few misinformation or “reinforcement” pages are found. The only exceptions are “Epoch Times Paris”, “RT France” and “La vraie démocratie” (in France), as well as “L’Anti-Média” (in Québec).

Between August and November 2020, Facebook said it has removed “about 3,200 Pages [...] for violating our policy against militarized social movements [and] about 3,000 Pages [...] for violating our policy against QAnon” (Facebook, Inc., 2021). It said it has also removed

pages for otherwise “coordinated inauthentic behavior”. Many such pages have been publishing content in French in the four countries covered by this study (Belga, 2020; Facebook, Inc., 2020a; Gleicher and Agranovitch, 2020; Loiseau, 2020). It must however be said that the sixth highest ranked page in Canada is Éric Duhaime’s. Mr. Duhaime is a controversial talk radio host, head of Québec’s Conservative Party since April, 2021. He has been known to propagate misinformation on his social networks (one of his videos has been removed by Youtube (Bellerose, 2021)) and on his radio show where he famously said, “Bad information is better than no information” (Hachey, 2019).

Second, apart from clickbait and artists/celebrities pages, religious content appears to be what generates the most engagement among francophone Facebook users. For example, “La Parole vivante”, publishing biblical quotes and memes, is the page generating the most interactions in French Canada, media or non-media categories alike. Throughout 2020, Facebook users have shared, reacted on or commented its content more than 21 million times. That’s two interactions every three seconds. A page on pope Francis, created in Belgium, has attracted almost 4,000 interactions per post through the year. The fact that Pastor Paul Mukendi has been convicted of sexual assault in 2019, condemned to eight years behind bars in 2020 and is awaiting a second trial, still for sexual assault (Bergeron, 2020), does not

seem to have deterred his followers as his page is the ninth most engaging non-media page in French Canada! We'll see later that religious content is not only common in, but characteristic of the remainder of the non-media corpora.

Third, political figures are completely absent from the top non-media pages. The only exception is Canada where the official pages of Justin Trudeau and François Legault, respectively Prime ministers of Canada and of Québec, occupy the first two spots in terms of interactions per post. This might be because both politicians have held almost daily publicly transmitted press conferences during the COVID-19 pandemic, something heads of State or government in the other three countries have seldom done.

No news is feel good news on Facebook

After extracting n-grams for all eight subcorpora, I sorted them by the total number of interactions generated by the posts they were found in. I kept only the top-1,000 in each and then put them all back together in three lists, each made of 8,000 lemmas, bigrams or trigrams which have been associated with the content garnering the most attention in French-language Facebook pages in 2020.

What is most relevant, here, are not the lists themselves. They tell us, for example, that “cristiano ronaldo” is, after “timeline photo” and “photo from”, the top bigram in the French non-media corpus, appearing in 7,370 posts having elicited close to 54 million interactions in 2020.

More interesting is to inquire what n-grams present in non-media are absent from media. Bear in mind that we have limited ourselves to the top-1,000, so “absent” only means absent from that particular list, not from Facebook. But here is precisely why this exercise is relevant. We have kept only the most significant expressions according to Facebook’s own logic, where shares, likes and comments are used to elicit users’ emotions and keep them coming back. We can thus say that we are beginning to draw light on what would be characteristic of a newsless Facebook. Tables 9 and 10 give some examples.

Table 9 – Top-5 lemmas from non-media corpora, absent from media corpora, and present in all four countries

Lemma	Posts	Interactions
<i>dieu</i>	75,193	122,997,101
<i>aime</i>	57,190	111,981,655
<i>heureux</i>	35,653	61,630,964
<i>bonheur</i>	31,164	60,600,496
3	36,531	59,923,266

Table 10 - Top-5 bigrams from the non-media corpora, absent from the media corpora, and present in all four countries

Bigram	Posts	Interactions
<i>bon journée</i>	10,133	22,199,695
<i>bel journée</i>	8,518	20,319,814
<i>bon dimanche</i>	4,753	9,531,666
<i>chaîne youtube</i>	5,492	9,504,084
<i>jésus christ</i>	8,279	9,493,523

In the case of 3-grams, the top of the list is occupied by “*bon week end*” (have a good weekend). This, and the expressions “*bon/bel* (lemmatized “*belle*”) *journée*” (have a nice day), “*bon dimanche*” (wishing someone a nice day, but on a Sunday), “*aime*” (love at the first or third person, singular), “*heureux*” (happy), “*bonheur*” (happiness) and “3” (emoji code for a heart, “<3”, stripped from the less-than punctuation character in the preprocessing phase), reflect how much Facebook is built on feel good memes of people wishing their loved ones a good day with teddy bears, sparkling hearts and rainbows on pink backgrounds. Figure 1 gives a typical example.

Figure 1 - Feel good meme post by the “Une pomme par jour” (An apple a day) page posted Dec. 8, 2020. The message reads, “Life is short. Spend it with those who make you happy. Have a nice day.”



Facebook calls these “inspirational posts”. In the Discussion, below, we’ll see why the Palo Alto company has publicly favored this type of post over news content since 2018.

Two other keywords that stand out in the top non-news content are "dieu" (god) and "jésus christ". The prevalence of religious content is one of the main surprises of this study, given the secular tradition of France, which also permeates French-language societies in Canada, Belgium and, to a lesser extent, Switzerland (Willlaime, 2017). We will discuss it further below, as the other textual analyses performed on our corpora will continue to reveal the contents of a newsless francophone Facebook.

God is everywhere

Twelve computations of weighed χ^2 residuals were executed, one for each of the n-gram types extracted (lemmas, bigrams and trigrams) in every country. The top-50 results for each were then joined in two tables, one for those most characteristic of media pages, the other for those most characteristic of non-media pages. If a result appeared in two countries or more, it was included in Figures 2 and 3.

Figure 2 - Most characteristic terms in posts from French-language media pages appearing in two countries (Belgium, Canada, France and Switzerland) or more, by sum of χ^2 residual score weighed by interactions (2020). The number right of the bar is the number of countries this term was found in.

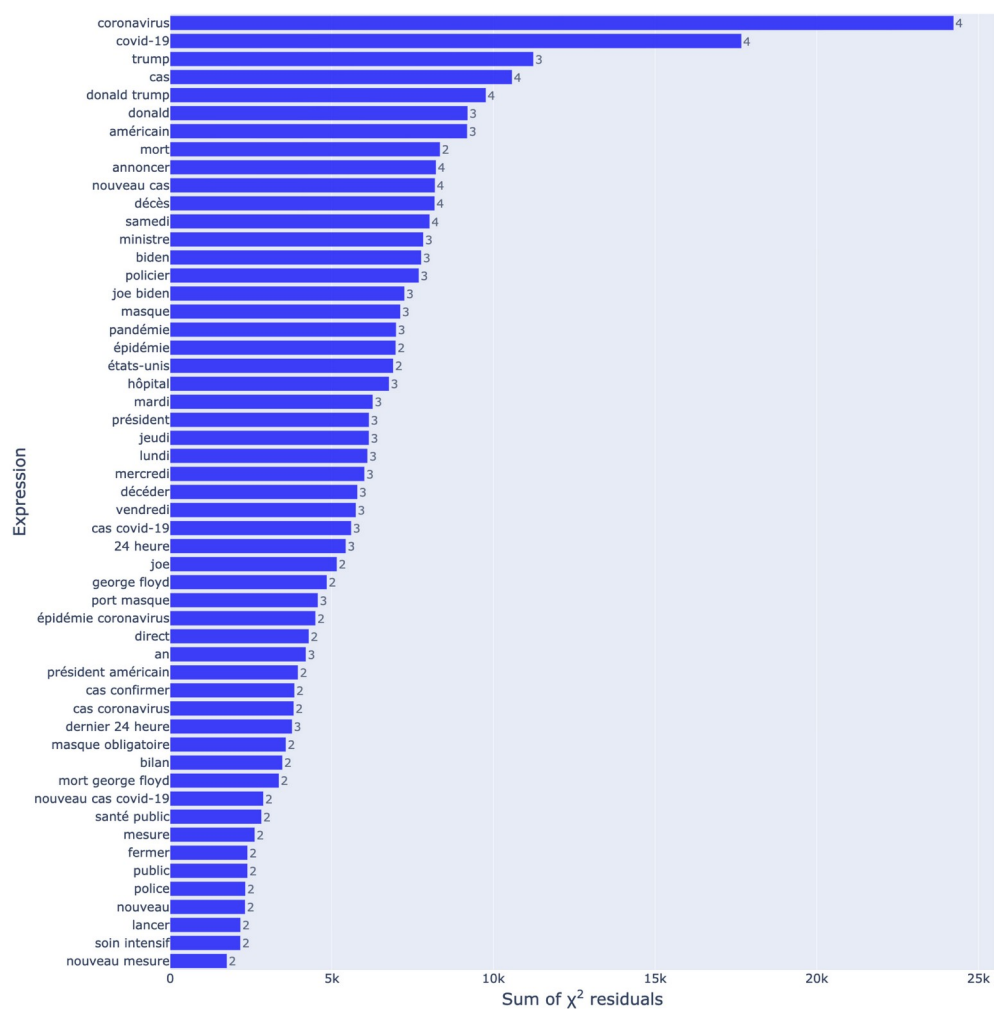
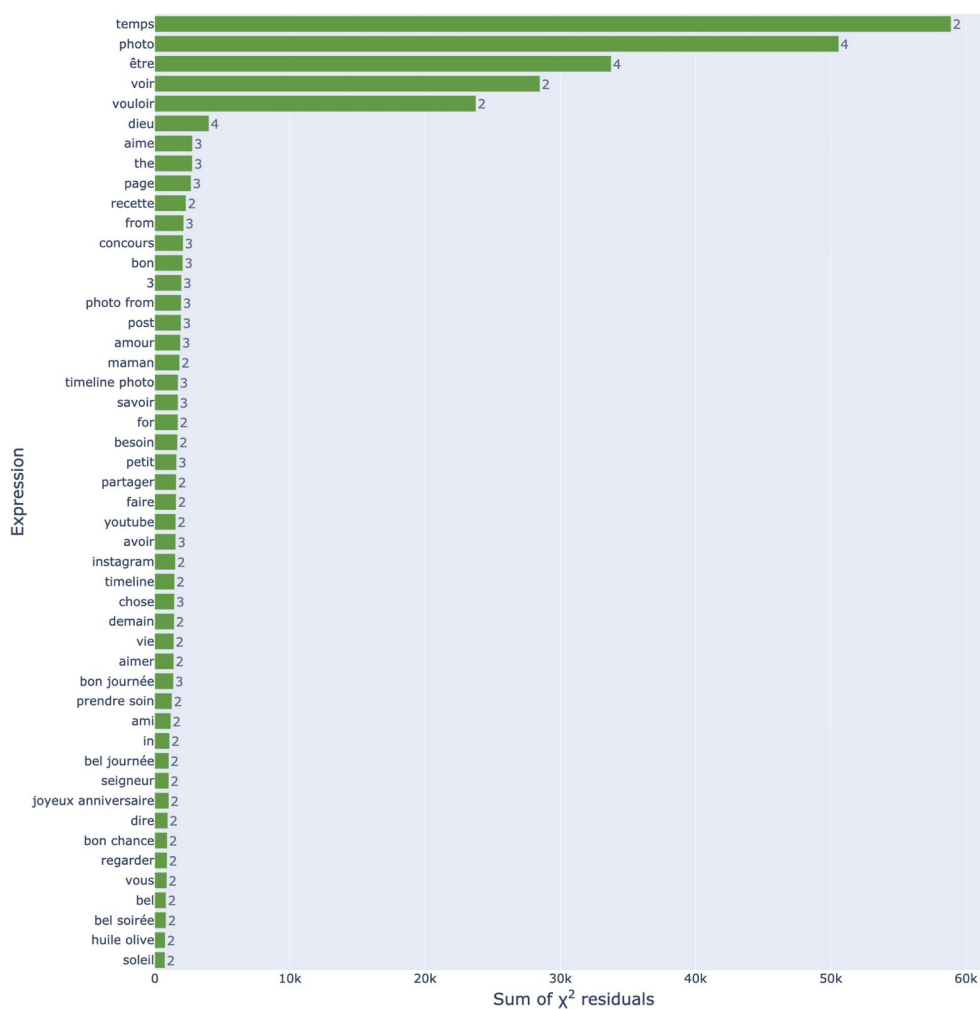


Figure 3 – Most characteristic terms in posts from French-language non-media pages appearing in two countries (Belgium, Canada, France and Switzerland) or more, by sum of χ^2 residual score weighed by interactions (2020). The number right of the bar is the number of countries this term was found in.



First, it might look surprising that the events generating the most attention in French-language media, apart from the COVID-19 pandemic, have happened in the United States. Terms related to the George Floyd murder, and ensuing demonstrations, and to the presidential election are among the most characteristic of media pages in the Francophone areas of Facebook. But that is predictable, given that “global news flows are dominated by Anglo-Saxon media” (Marthoz, 2018, p. 95). The dominance of news from the United States has also been observed in a

content analysis of Instagram posts by Francophone news organizations from 2011 to 2020 (Roy, 2021b).

Unsurprisingly, "*coronavirus*" and "*covid-19*" are by far the top two terms most characteristic of media pages. This is line with the formidable amount of coverage devoted to the pandemic in 2020 by journalists worldwide. "*Coronavirus*" is the number one lemma characteristic of media pages in France, Belgium and Switzerland, while "*covid-19*" is number one in Canada.

This result also means COVID-19 has not been characteristic of non-media pages. That is a little more unexpected, given the conventional wisdom according to which misinformation about the pandemic has been rampant on Facebook. Figure 3 shows no term related to any conspiracy theory having circulated in 2020 (vaccines, chloroquine, 5G technology, China, Bill Gates, etc.). Some related expressions are found in individual countries' top-50 terms. The bigram "*coronavirus confinement*" is the 49th most characteristic of non-media pages in Belgium. In Canada, "*presse covid-19*" and "*anti media*" are found in the 38th and 39th positions. But these are the only traces of probable misinformation in content having generated the most interactions in French-speaking Facebook.

What is much more typical of non-media pages is content of little public interest (we will define this notion in the Discussion, below). After “*photo*”, the highest scoring term appearing in all four countries is “*être*” (to be). The dominance of that verb, the most common in the French language, is difficult to interpret. If it was as common in media pages, it would not even show up in either Figures 2 or 3. Many reasons may explain why it is more characteristic of non-media pages.

One hypothesis : this indicates that news content is more impersonal and relates less often to well-being than what can be found in non-media Facebook pages. It may also stem from the fact that Facebook and other social networks answer people’s need to self-represent (Nadkarni and Hofmann, 2012). In a study with more than 15,000 participants, Bastard *et al.* have developed a taxonomy of Facebook users. Only 17 percent are what the authors call “non active users” who prefer to lurk instead of participating in conversations with others or publishing content themselves. “For these users, Facebook isn’t a tool for self-expression or self-representation” (2017, p. 70). Most people use Facebook to project of whom they wish *to be*. Thus, a second hypothesis : non-media content caters to that need.

The words “*recette*” (recipe) and “*concours*” (contest) appearing rather high in Figure 3, along with the bigrams “*bon chance*” (good

luck, with "*bonne*" being lemmatized) and "*huile olive*" (olive oil), give another indication of the type of content characterizing non-media pages on Facebook. Recipes and contests are surely of value to the social network's users, but of little public interest value.

The word "*temps*", which can either mean time or weather, has the highest χ^2 residual score even though it appears in the top-50 list in only two countries (France and Belgium). In which sense is it used in non-media pages? Weather information is a staple of news content. Some weather-related terms should therefore appear in the n-gram lists specific to media pages. Yet there are none. This means that weather-related terms are found in proportionate amounts in non-media pages and that they are not specific of non-news Facebook content. So it is when it means "time" that "*temps*" is characteristic of non-media pages.

Besides, "*prendre temps*" (take [your, the] time) is the 34th in the list of non-media bigrams in Canada. This expression is another fixture of feel good memes, along with "*joyeux anniversaire*" (happy birthday), "*prendre soin*" (take care) and others mentioned earlier in this section.

Further terms of little value outside of Facebook that seem to characterize non-media pages are "*partager*" (to share), "youtube",

"instagram", and other calls to action. Also, terms that describe a given post and that are untranslated, like "photo from" and "timeline photo", appear in my sample only because they are automatically generated when an administrator creates a post on a Facebook page.

More significantly, "God" is a lemma that has shown up in all four countries' list of terms most specific of non-news content. It has been in posts generating close to 123-million interactions in 2020. Only "coronavirus" has generated more in media pages... during the year the COVID-19 pandemic began! "*Seigneur*" (Lord) is also there, having made the top-50 list in two countries. Not appearing on in Figure 3, "*allah*" is the 26th most specific lemma in non-media content in Belgium.

The word "*dieu*" is part of many common French expressions.

"*Mon dieu!*" is as prevalent as "Oh my god" in English. But closer inspection reveals that it is indeed in faith-based pages that the word is used most often and where posts collect the most interactions. Many of those pages have over a quarter million subscribers : [Pastor Yvan Castanou's](#) or "[Un miracle chaque jour](#)" (A miracle a day) in France, "[Darifton & compagnie](#)", a Muslim community in Belgium, "[La Bible](#)", a Swiss page carpet bombing the social network with biblical memes, or "[EMCI-TV](#)", a Christian evangelical

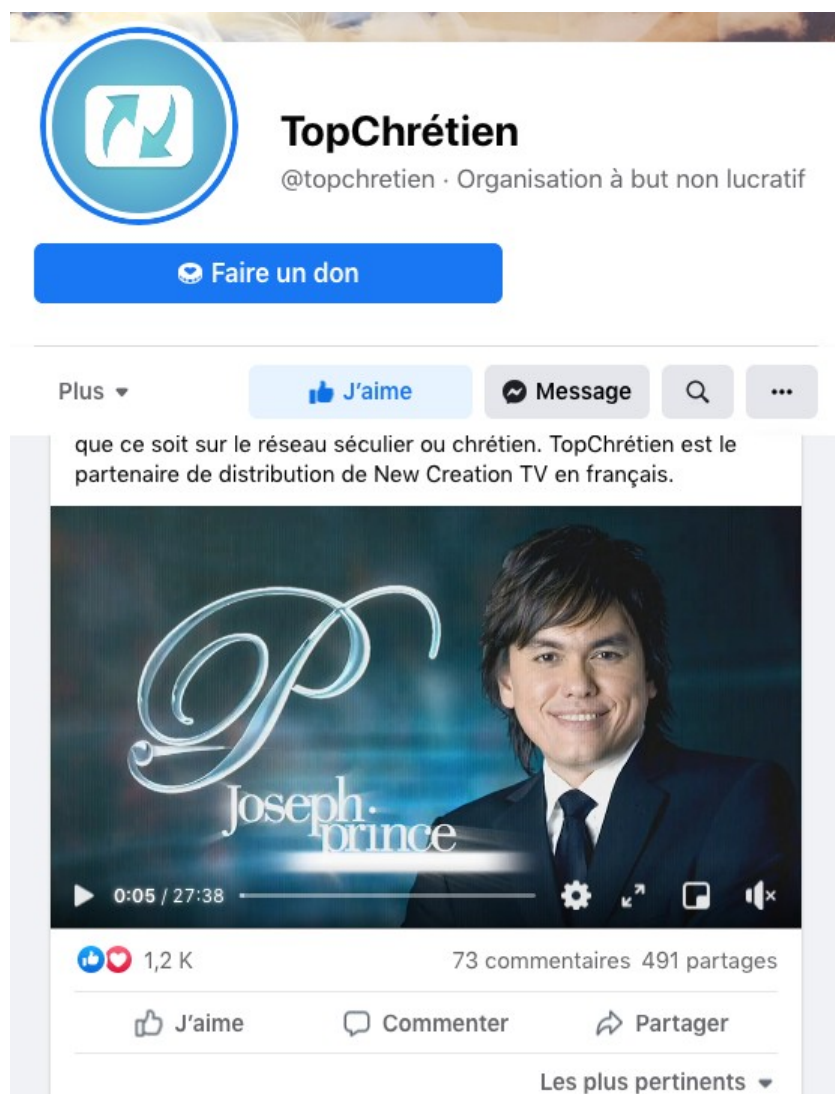
television network based in Waterloo, Québec, broadcasting in Europe and Africa.

The internet is a "mission field" (Brubaker and Haigh, 2017, p. 2) for which Facebook is well tailored. "Religious traditions [...] encourage the use of Facebook for proselytizing" (Ibid., 8).

Facebook's algorithm may thwart these efforts by enclosing its users in filter bubbles (Pariser, 2012), making it difficult to preach to those not already converted. Yet, Brubaker and Haigh have found in a survey they conducted in 2017 that "Facebook use for religious purposes is primarily motivated by the need to minister to others" (Ibid.). More than simply sharing info about themselves, devout Facebook users are motivated to share information about their religion : "They reach out to and uplift others by engaging in faith-based conversations and uploading faith-based messages" (Ibid.).

Figure 4 gives one example. It was posted by [TopChrétien](#), a page based in France with more than a million subscribers. Posted in April, this video is a half-hour sermon by Singapore preacher Joseph Prince dubbed in French. It had 60,000 views at the time of data collection.

Figure 4 - Repost of an English-language evangelical sermon by a French-language christian page.



Churches can also use Facebook to expand. Kgatle Mookgo (2018) has documented the use of the social network in the emergence of prophetic churches in southern Africa and found it plays a “major role” (Ibid., 4). Apart from sharing messages, it provides a platform to organize and advertise online events and services, as well as a way for the faithful to attend those services virtually live or later on Watch.

We will see in the next subsection that non-media corpora is more varied, but that religious content is still very much present.

Topic modeling : the complete diet of a newsless Facebook

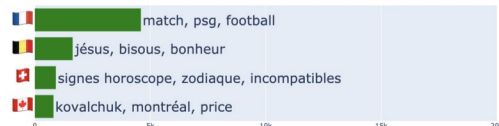
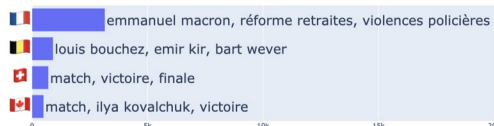
The fact topic modeling was applied to monthly subcorpora provides a finer, more granular analysis of the content posted in the four countries studied during the year. It generated approximately 72,000 terms in more than 5,000 topics.

Figure 5 - Selection of terms appearing in the clearest topics identified per country every month during 2020, in both media and non-media subcorpora; length of bar shows frequency, or number of posts included in a given topic.

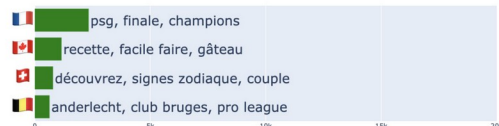
media

non-media

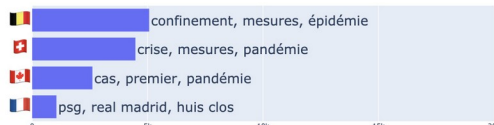
Jan



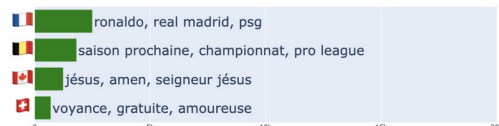
Feb



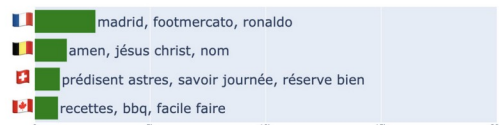
Mar



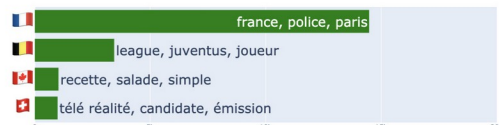
Apr



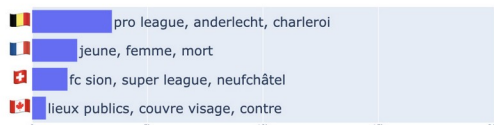
May



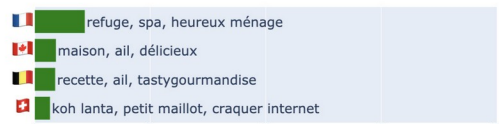
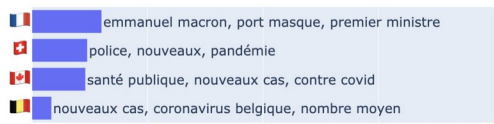
Jun



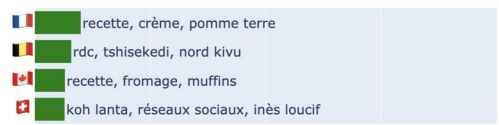
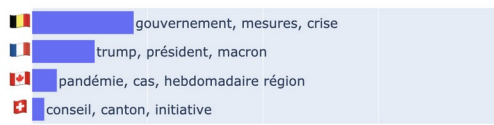
Jul



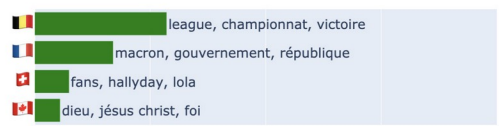
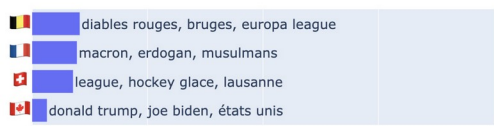
Aug



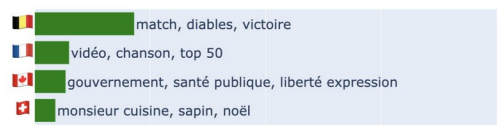
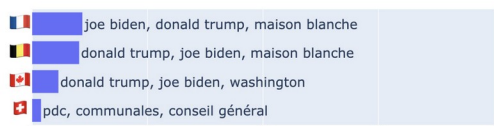
Sep



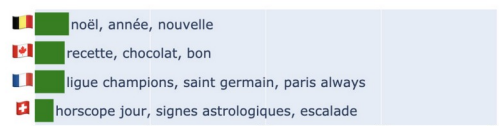
Oct



Nov



Dec



frequency

frequency

Figure 5 synthesizes the main interpretable² topics by month and country in both media and non-media subcorpora. It paints a broad picture where content found in media pages generally relates to current events, mainly the COVID-19 outbreak, while what is published on non-media pages seems to tout Facebook users with easy to produce and enticing content (recipes, astrology, self-help, among others).

However, the difference between media and non-media is not as clear cut as what was revealed by the χ^2 residuals analysis. This may be the main takeaway from the topic modeling phase of this study. The models used have produced topics common to both supcorpora, most notably sports which seem almost as commonplace in media pages as they are in non-media pages.

Topic modeling has also shown that the pandemic was not confined to media pages. It was discussed in non-media pages throughout the year, most particularly in March. That being said, it seemed that the non-media pages dealing with the pandemic were mostly government agencies as the terms were related to rules, like "*masque obligatoire*" (mandatory mask) or "*rassemblements*" (gatherings), along with the name of an agency, such as "*conseil fédéral*" (Federal Council, the executive branch of government in Switzerland), or invitations to call a number to be informed on sanitary measures. Sometimes, covid topics in non-media pages were more personal in tone, inviting

Facebook users to "*prendre soin*" and "*soin autres*" (to take care of others), giving them advice, such as "*nettoyer légumes*" (wash vegetables), or telling them how to "*vivre normalement*" (live normally) despite the confinement.

Occasionally, a few other current events topics appeared in the non-media subcorpora. Apart from sports, as mentioned above, extreme weather events are what appeared most frequently. Storms Ciara and Dennis, in northern Europe, along with the fires in Australia, probably because of all the cute koala pictures ("*koala*" being a term also present in those topics), top the list. Three other topics have generated so much attention worldwide that they have been included in a few topics in the non-media subcorpora : the U.S. presidential election; the Black Lives Matter demonstrations following the murder of George Floyd; and the beheading of teacher Samuel Paty, in a northern suburb of Paris, for having shown cartoons of Muhammad in class.

Were the posts associated with those topics denouncing BLM demonstrations or supporting them? Were they condemning the assassination of Samuel Paty, or condoning it? The terms picked up by the models do not enable us to tell what were the sentiments associated with those posts. Only in Donald Trump's case is it clearer. The terms "*voler élection*" (steal election), "*doit vivre*"

(must live) and "*monde parallèle*" (parallel world) indicate the posts seemed to ridicule the former president's allegations of voter fraud.

Celebrity news is another topic that has proved a standard of both media and non-media pages, even though it couldn't be classified as "current events". French rock singer Johnny Hallyday was present in many topics (he died in 2017!). So were reality TV stars from shows such as "Occupation Double" in Canada or "Koh Lanta" in Europe. According to one Swiss non-media topic in August, the "*petit maillot*" (small bathing suit) of a participant of the latter program makes "*craquer internautes*" (breaks the internet).

Most topics, though, were relatively unique to non-media pages. Recipes have been found in topics in all countries almost every month. Typical terms are "*recette facile*" (easy recipe) with ingredients, like "*beurre*" (butter) or "*pommes terre*" (potatoes), to make "*repas Noël*" (Christmas meal) or "*délicieux gâteau*" (delicious cake). Some self-help topics, mostly to "*perdre poids*" (lose weight), have appeared once in a while, but less often.

In Switzerland, astrology has been surprisingly dominant in non-media pages. Perhaps this comes the fact the non-media page having posted the most in French in this country, last year, is "[*Horoscope du jour*](#)" (Today's horoscope). Terms such as "*compatibles amour*" (compatible

zodiac signs for love) or "*prédisent astres*" (what are the stars predicting), for example, have defined one to five topics (out of 12) almost every month. In January alone, more than 1,950 Swiss non-media posts mention "horoscope". Of course, not a single one predicted the COVID-19 pandemic.

But recipes and horoscopes sometimes made up topics in media pages, albeit rarely. Only two types of content were unique to the non-media francophone Facebook, thus confirming the findings of the first two textual analyses : faith-based and, to a lesser extent, "feel good" content.

To be more precise, it is Christian and Islamic content that has turned up regularly in topics, mostly in the Canadian and Belgian subcorpora. The expression "*jésus christ*", for example, has appeared in no less than 71 topics throughout the year in the non-media subcorpus, not once in the media subcorpus. Terms like "*nom jésus*" (name of Jesus), "*seigneur jésus*" (lord Jesus) or "*prière*" (prayer) also appear in dozens of topics unique to non-media subcorpora. Those topics are sometimes the most salient of the month in some Canadian subcorpora. In some topics, the expression "*lavictoiredelamour*" is present. "La victoire de l'amour" is the name of a dominical television program on TVA, Québec's most watched private tv network. The popularity of the network itself does not explain why religious

content has been highlighted so frequently by topic modeling. One more probable hypothesis could be that content posted by this program has been shared and reposted by many other pages on Facebook and therefore appears more often in my final sample.

In Belgium, Muslim content has regularly been in top topics throughout the year, with terms adapting to events in the Islamic calendar, like “*tarawih*” and “*iftar*” during Ramadan, for example. In some topics, terms appear in Arabic, such as “عليه وسلم” (peace be upon Him).

In the “feel good” category, terms such as “*gros bisous*” (hugs and kisses), “*beaux rêves*” (sweet dreams), along with expressions highlighted in χ^2 residuals such as “*bonne journée*” (have a good day), turned up in many topics of all non-media subcorpora.

One last element that stands out is that only two, maybe three topics could be attributed to disinformation, which is very little. In June, in the Canadian non-media subcorpus, one topic included terms such as “donald trump”, [French president] “emmanuel macron” along with “oui avez”, “avez bien” and “bien lu” (yes, you’ve read correctly). The use of the second person, engaging the reader by talking directly to him or her, is a telltale sign of misinformation (Rashkin et al., 2017, p. 2933). After swearing, it is the linguistic feature most

often associated with fake news. It appears almost seven times as much in fake news articles as it does in trusted sources. In November, in the non-media Switzerland subcorpus, another topic contained terms related to vaccines along with "*expérience incroyable*" (incredible experience or experiment). Rashkin and her colleagues have also found that the use of superlatives could be a sign of misinformation, although not as clear as the use of the second person.

In March and November, the terms "chloroquine" and "raoult" appeared in French and Swiss non-media subcorpora. Didier Raoult, a Marseille-based doctor, has become a household name in much of the French-speaking world after publishing an article in the International Journal of Antimicrobial Agents. It claimed that "hydroxychloroquine treatment is significantly associated with viral load reduction/disappearance in COVID-19 patients and its effect is reinforced by azithromycin" (Gautret *et al.*, 2020). It stirred a debate in the French press, amplified by the subsequent endorsement of the treatment by U.S. president Donald Trump. But it is difficult to attribute these topics to misinformation as Raoult's name also appeared in three topics from the media subcorpus during the year. On closer analysis, it appears Mr. Raoult has been mentioned in almost 4,600 posts from the French subcorpus, 63 percent of which were posted in non-media pages. Among those pages, many are known for

misinformation, such as “RT France”, “Epoch Times Paris”, “Gilets Jaunes Infos” or “L'Eveilleur Quantique”. One of the posts generating the most attention is a video (Figure 6) posted by “La vraie démocratie”, a page considered “one of the most influential francophone misinformation pages on Facebook” (Conspiracy Watch, 2020). The video has more than 178,000 total interactions and almost 3.5 million views as of July 2021. But not all pages mentioning Mr. Raoult can be labeled misinformation. Besides, the page having quoted him most often in 2020 (161 times) is Marseille daily *La Provence*. His name also appears in 34 posts by *Le Monde*. Raoult may certainly be synonymous with debate and controversy, but not necessarily with disinformation.

Figure 6 – Post by reinformation page “La vraie démocratie” on October 28, 2020. It is a video excerpt of an interview of Dr. Didier Raoult aired on all-news channel LCI the day before. The all-caps text above the still frame reads: “You’ve all gone bonkers!” In the message text above, Raoult is quoted as saying: “In the end, will we all be locked up for the rest of our lives because there are viruses outside?”



This study should not be interpreted as demonstrating there is no disinformation on Facebook. All it does is show disinformation is not a salient topic in its Francophone areas, even when one subtracts media content from the social network. This does not mean there are no fake news in French on Facebook, it only means they are not one of its main ingredients.

A lot of the literature on fake news has been focusing on the United States. Perhaps it has skewed the perceptions. Humprecht *et al.* have found that it is "the most vulnerable country regarding the spread of online disinformation" (2020, p. 14). In their cross-national study of the "resilience to online disinformation", they have clustered 18 countries in three groups. The U.S. is in a group all by itself, "characterized by high levels of populist communication, polarization, and low levels of trust in the news media" (ibid.). France was not included in this study, but Belgium, Canada and Switzerland, countries "with consensus political systems, strong welfare states, and pronounced democratic corporatism" (ibid.), have been classified in the cluster most resilient to disinformation. The findings of this study support the conclusions of Humprecht and her colleagues.

Discussion and conclusion

All three textual analyses converge. The differences they've helped pin point between media and non-media content on Facebook boil down to a classical opposition between "public interest" and "the public's interest".

The concept of "public interest" can be synonymous with "common good", "public service" or "general welfare". It is something many

professions vow to defend : public servants, politicians, lawyers, physicians, journalists, to name just a few.

In journalism, public service is a core ideal-typical value and a "powerful component of journalism's ideology" (Deuze, 2005, p. 447). Many journalists see themselves as working for the public first and foremost, the media organization employing them being second in the order of their loyalties. That their "primary commitment is to the public" is a "deeply felt tradition" (Kovach and Rosenstiel, 2014).

But it isn't only for the public's sake that journalists strive to serve it. The public service ideal goes hand in hand with the democratic ideal. "Journalists think of journalism as a service in the public interest, one that is shaped with an eye toward the needs of healthy citizenship" (Zelizer, 2005, p. 72). It can be argued that Western journalism sees "the role of the press as serving the public interest by providing the informational needs necessary for [a] self-governing republic" (Painter, 2019, p. 7). In doing so, "Journalists earn public trust and protection by loyally performing their public service duty to citizens" (ibid.).

To fulfill this duty, normative as it may be, media organizations have historically found ways to reach the public they aim to serve. They've done so by selling newspapers, by broadcasting radio and

television news programs, by publishing websites. In the last decade, social networks have become an increasingly important channel to reach the public. It is a channel media organizations don't control. It is a channel where Facebook products (the original Facebook platform, Instagram, as well as messaging applications Messenger and WhatsApp) are by far the most used by citizens throughout the world to access news according to the latest editions of the Reuters Institute Digital News Report (Newman *et al.*, 2019, 2020, 2021).

Facebook's role goes beyond that of a neutral, passive platform. Beginning in 2013, it has actively courted news organizations, inviting them not only to share content on Facebook, but to create content specifically designed for it (Mattelart, 2020). Without ever explicitly recognizing it, Facebook has taken the public service duty to inform citizens upon itself too.

It seemed to embrace that role in a 2017 manifesto following the 2016 U.S. presidential election. The social network's founder pledged that his company's "next focus will be developing the social infrastructure for community" (Zuckerberg, 2017). Among the five goals he described for the future of Facebook was that it should be "informing us".

Although Zuckerberg wrote that “a strong news industry is [...] critical to building an informed community”, many decisions taken afterwards have run counter to the 2017 manifesto. Barely a year later, in a bid “to encourage meaningful social interactions with family and friends”, the CEO announced his users will “see less public content, including news [...]”. After this change, we expect news to make up roughly 4% of News Feed -- down from roughly 5% today (Zuckerberg, 2018).

In February 2021, while the storm was the fiercest in Australia, Facebook announced it wished to reduce political content on its’ users News Feed. “We’ll temporarily reduce the distribution of political content in News Feed for a small percentage of people in Canada, Brazil and Indonesia this week, and the US in the coming weeks” (Gupta, 2021a). Content from news media could be labeled political, according to a *New York Times* story. “Under the new test, a machine-learning model will predict the likelihood that a post – whether it’s posted by a major news organization, a political pundit, or your friend or relative – is political” (Roose and Isaac, 2021).

Two months later, the Palo Alto company began surveying its users “to understand which posts they find inspirational [...] People have told us they want to see more inspiring and uplifting content in News Feed because it motivates them [...]”. For example, a post featuring a

quote about community can inspire someone to spend more time volunteering" (Gupta, 2021b).

Less news, more "inspirational" content. Facebook's idea of an informed community is one where individual expression around personal interests and experiences are merely aggregated by algorithms. It is not the public sphere of deliberation where professionals gather and report news in the public interest an informed citizenry would be in right to expect (Carignan, 2021).

The kind of content Facebook pushes to its users is a self-serving version of the public interest. Carol W. Lewis would call it "a fabricated version of the common good" (2006, p. 698). Removing public interest content from Facebook helps us understand what kind of content the Palo Alto giant favors : sports, celebrities, "inspirational" "feel good memes"; fast food satisfying in the short term, but dangerous in the long run for the health of a democratic society... or what Mr. Zuckerberg calls a "civically-engaged community".

"It is not true that news has value for Facebook", Kevin Chan, head of public policy for Facebook, Inc., in Canada, told me earlier this year (Roy, 2021a). I beg to differ. News content is extremely valuable for the company. It confers what Jane B. Singer calls

"societal value" by conveying "information people can trust" (2006, p. 24). That trust, she writes, "is best established and nurtured by those with an existential commitment to social responsibility", i.e., journalists. Trust in news media is low in many countries, but trust in social media is much lower (Edelman, 2021).

Would Facebook users come back if they could no longer find trustworthy information on the platform? Maybe not, because as Singer points out : "the public needs some means of differentiating between what is valuable to society as a whole and what is less so; otherwise, the notion of a coherent 'public' falls apart as each individual seeks out whatever seems most personally appealing at the moment" (ibid.).

In other words, a newsless Facebook would be a socially useless Facebook.

Technical appendix

The full code and most data (in accordance with CrowdTangle's Terms of Service) are available in a repository on the author's github account : <https://github.com/jhroy/facebook-franco>

About the author

Jean-Hugues Roy teaches data and computational journalism at Université du Québec à Montréal (UQAM). A journalist for 23 years, he now studies information flows online and media business models using computational methods.

Notes

1. In fact, when searching for content in Canada, for example, CrowdTangle will return content published by pages whose administrators are mainly from Canada. This explains why certain pages whose content is related to a given country are classified in another country. British singer Dua Lipa's official page, for example, has 208 administrators in 31 countries! Because the majority of them (77) are in Canada, the page is considered a Canadian page.

2. Four types of topics produced by the models used were ignored because irrelevant or insignificant : those joining Facebook-specific ("timeline photo") or link-specific ("http www"; "blaquesquebec com") vocabulary; those made up of stopwords ("qu ils"; "2020 06"); those including calls to action by Facebook pages, media or not ("cliquez ici" [click here]; "envoyez nous" [send us]; "chance gagner" [enter to win]); or those made up of words not in French (mostly English, sometimes Dutch or Arabic).

References

- Al-Rfou, R. (2016). *Polyglot is a natural language pipeline that supports massive multilingual applications*. (16.7.4) [Python]. <https://github.com/aboSamoor/polyglot>
- Bastard, I., Cardon, D., Charbey, R., Cointet, J.-P., & Prieur, C. (2017). Facebook, pour quoi faire ? : Configurations d'activités et structures relationnelles. *Sociologie*, 8(1), 57. <https://doi.org/10.3917/socio.081.0057>
- Belga. (2020, October 12). *Facebook interdit les contenus négationnistes*. Lavenir.net. https://www.lavenir.net/cnt/dmf20201012_01519459/facebook-interdit-les-contenus-negationnistes
- Bell, E. J. (2021). *Off-label. How tech platforms decide what counts as journalism*. Columbia Journalism Review : The Existential Issue. <https://existential.cjr.org/who/tech-platforms-labels/>
- Bell, E. J., Owen, T., Brown, P. D., Hauka, C., & Rashidian, N. (2017). *The Platform Press: How Silicon Valley Reengineered Journalism*. <https://doi.org/10.7916/D8R216ZZ>
- Bellerose, P. (2021, August 2). COVID-19: YouTube retire une vidéo d'Éric Duhaime. *Le Journal de Québec*. <https://www.journaldequebec.com/2021/08/02/covid-19-youtube-retire-une-video-deric-duhaime>
- Bergeron, Y. (2020, July 29). *Le pasteur Mukendi devra subir un autre procès*. Radio-Canada.ca; Radio-Canada.ca. <https://ici.radio-canada.ca/nouvelle/1723056/pasteur-paul-mukendi-nouveau-proces>

- Bestgen, Y. (2014). Inadequacy of the chi-squared test to examine vocabulary differences between corpora. *Literary and Linguistic Computing*, 29(2), 164-170. <https://doi.org/10.1093/llc/fqt020>
- Blanc, C. (2016). Réseaux traditionalistes catholiques et « réinformation » sur le web: Mobilisations contre le « Mariage pour tous » et « pro-vie ». *tic&société*, Vol. 9, N° 1-2, Article Vol. 9, N° 1-2. <https://doi.org/10.4000/ticetsociete.1919>
- Boumedyen Billami, M., Bortolaso, C., & Derras, M. (2020). Extraction de thèmes d'un corpus de demandes de support pour un logiciel de relation citoyen. *Actes de La 6e Conférence Conjointe Journées d'Études Sur La Parole*, 155-163.
- Brubaker, P. J., & Haigh, M. M. (2017). The Religious Facebook Experience: Uses and Gratifications of Faith-Based Content. *Social Media + Society*, 3(2), 2056305117703723. <https://doi.org/10.1177/2056305117703723>
- Carignan, R.-Y. (2021). *Le journalisme et la représentation des rapports socio-numériques* [Maîtrise en communication]. UQAM.
- Cepernich, C. (2016). Emotion in Politics. In G. Mazzoleni (Ed.), *The International Encyclopedia of Political Communication* (pp. 1-10). John Wiley & Sons. <https://doi.org/10.1002/9781118541555.wbiepc238>
- Conspiracy Watch. (2020, August 17). La vraie démocratie. *Conspiracy Watch | L'Observatoire du conspirationnisme*. <https://www.conspiracywatch.info/la-vraie-democratie>
- Danilak, M. M. (2020). *Language detection library ported from Google's language-detection*. (1.0.8) [Python; OS Independent]. <https://github.com/Mimino666/langdetect>

- Deuze, M. (2005). What is journalism? Professional identity and ideology of journalists reconsidered. *Journalism*, 6(4), 442-464. <https://doi.org/10.1177/1464884905056815>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv:1810.04805 [Cs]*. <http://arxiv.org/abs/1810.04805>
- Edelman. (2021). *Country Report. Trust in Canada. Edelman Trust Barometer 2021* (p. 37). Edelman. https://www.edelman.ca/sites/g/files/aatuss376/files/trust-barometer/2021%20Canadian%20Edelman%20Trust%20Barometer_0.pdf
- Facebook, Inc. (2020a, August 6). July 2020 Coordinated Inauthentic Behavior Report. *About Facebook*. <https://about.fb.com/news/2020/08/july-2020-cib-report/>
- Facebook, Inc. (2020b). *Facebook response to the Australian Treasury Laws Amendment (News Media and Digital Platforms Mandatory Bargaining Code) Bill 2020* (p. 26). https://www.accc.gov.au/system/files/Facebook_0.pdf
- Facebook, Inc. (2021, January 19). An Update to How We Address Movements and Organizations Tied to Violence. *About Facebook*. <https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence/>
- Gautret, P., Lagier, J.-C., Parola, P., Hoang, V. T., Meddeb, L., Mailhe, M., Doudier, B., Courjon, J., Giordanengo, V., Vieira, V. E., Tissot Dupont, H., Honoré, S., Colson, P., Chabrière, E., La Scola, B., Rolain, J.-M., Brouqui, P., & Raoult, D. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19:

Results of an open-label non-randomized clinical trial.
International Journal of Antimicrobial Agents, 56(1), 105949.
<https://doi.org/10.1016/j.ijantimicag.2020.105949>

Gleicher, N., & Agranovitch, D. (2020, December 15). Removing Coordinated Inauthentic Behavior from France and Russia. *About Facebook*. <https://about.fb.com/news/2020/12/removing-coordinated-inauthentic-behavior-france-russia/>

Google, Inc. (2020). *Draft News Media and Digital Platforms Mandatory Bargaining Code. Submissions in Response* (p. 55).
https://www.accc.gov.au/system/files/Google_0.pdf

Grootendorst, M. (2020). *BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics*. (v0.5.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.4430182>

Gupta, A. (2021a, February 10). Reducing Political Content in News Feed. *About Facebook*.
<https://about.fb.com/news/2021/02/reducing-political-content-in-news-feed/>

Gupta, A. (2021b, April 22). Incorporating More Feedback Into News Feed Ranking. *About Facebook*.
<https://about.fb.com/news/2021/04/incorporating-more-feedback-into-news-feed-ranking/>

Hachey, I. (2019, September 19). Le paradoxe de la tolérance. *La Presse+*. https://plus.lapresse.ca/screens/afc3baaf-b226-4d5d-b541-abea5a7ece2b__7C__0.html

Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*. <https://spacy.io/>

- Huang, A. (2008). *Similarity measures for text document clustering*.
- Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to Online Disinformation: A Framework for Cross-National Comparative Research. *The International Journal of Press/Politics*, 25(3), 493-516.
<https://doi.org/10.1177/1940161219900126>
- Kgatle Mookgo S. (2018). Social media and religion: Missiological perspective on the link between Facebook and the emergence of prophetic churches in southern Africa. *Verbum et Ecclesia*, 39(1), 1-6. <https://doi.org/10.4102/ve.v39i1.1848>
- Kilgarrieff, A. (2001). Comparing Corpora. *International Journal of Corpus Linguistics*, 6(1), 97-133.
<https://doi.org/10.1075/ijcl.6.1.05kil>
- Kovach, B., & Rosenstiel, T. (2014). *The elements of journalism: What newspeople should know and the public should expect* (Revised and updated third edition). Three Rivers Press.
- Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbé, B., Besacier, L., & Schwab, D. (2020). FlauBERT: des modèles de langue contextualisés pré-entraînés pour le français. *Actes de La 6e Conférence Conjointe Journées d'Études Sur La Parole (JEP, 31e Édition), Traitement Automatique Des Langues Naturelles (TALN, 27e Édition), Rencontre Des Étudiants Chercheurs En Informatique Pour Le Traitement Automatique Des Langues (RÉCITAL, 22e Édition). Volume 2: Traitement Automatique Des Langues Naturelles*, 268-278.

- Leaver, T. (2021). Going Dark: How Google and Facebook Fought the Australian News Media and Digital Platforms Mandatory Bargaining Code. *M/C Journal*, 24(2), Article 2.
<https://doi.org/10.5204/mcj.2774>
- Lewis, C. W. (2006). In Pursuit of the Public Interest. *Public Administration Review*, 66(5), 694-701.
<https://doi.org/10.1111/j.1540-6210.2006.00634.x>
- Li, B., Zhao, Z., Liu, T., Wang, P., & Du, X. (2016). Weighted Neural Bag-of-n-grams Model: New Baselines for Text Classification. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 1591-1600.
- Lin, C.-Y. (1995). Knowledge-based Automatic Topic Identification. *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL)*, 308-310.
<https://www.aclweb.org/anthology/P95-1046.pdf>
- Loiseau, C. (2020, October 7). Les complotistes iront ailleurs. *Le Journal de Montréal*.
<https://www.journaldemontreal.com/2020/10/07/facebook-supprime-la-page-radio-quebec-de-sa-plateforme>
- Lui, M. (2016). *Langid.py is a standalone Language Identification (LangID) tool*. (1.1.6) [Python].
<https://github.com/saffsd/langid.py>
- Marthoz, J.-P. (2018). *Journalisme international*. De Boeck Supérieur.
- Martin, L., Muller, B., Suárez, P. J. O., Dupont, Y., Romary, L., Sagot, B., & Seddah, D. (2020). Les modèles de langue contextuels CAMEMBERT pour le français: Impact de la taille et de l'hétérogénéité des données d'entraînement. *Actes de la 6e*

conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 2: Traitement Automatique des Langues Naturelles, 54-65.

Mattelart, T. (2020). Comprendre la stratégie de Facebook à l'égard des médias d'information. *Sur le journalisme, About journalism, Sobre jornalismo, 9(1), 24-43.*

Nadkarni, A., & Hofmann, S. G. (2012). Why do people use Facebook? *Personality and Individual Differences, 52(3), 243-249.*
<https://doi.org/10.1016/j.paid.2011.11.007>

Nechushtai, E. (2018). Could digital platforms capture the media through infrastructure? *Journalism, 19(8), 1043-1058.*
<https://doi.org/10.1177/1464884917725163>

Newman, N., Fletcher, R., Kalogeropoulos, A., & Nielsen, R. K. (2019). *Digital News Report 2019* (p. 156). Reuters Institute.
https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-06/DNR_2019_FINAL_0.pdf

Newman, N., Fletcher, R., Schulz, A., Andi, S., & Nielsen, R. K. (2020). *Digital News Report 2020* (p. 112). Reuters Institute.
https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

Newman, N., Fletcher, Richard, Schulz, A., Andi, S., Robertson, C. T., & Nielsen, R. K. (2021). *Digital News Report 2021* (10th edition; p. 164). Reuters Institute.

- Oakes, M. P., & Farrow, M. (2007). Use of the Chi-Squared Test to Examine Vocabulary Differences in English Language Corpora Representing Seven Different Countries. *Literary and Linguistic Computing*, 22(1), 85-99. <https://doi.org/10.1093/llc/fql044>
- Painter, C. (2019). Public Service Role of Journalism. In *The International Encyclopedia of Journalism Studies* (pp. 1-8). American Cancer Society.
<https://doi.org/10.1002/9781118841570.iejs0093>
- Pariser, E. (2012). *The Filter Bubble: How the New Personalized Web is Changing what We Read and how We Think*. Penguin Books.
- Poudat, C., & Landragin, F. (2017). *Explorer des données textuelles: Méthodes - pratiques - outils* (1re. édition). De Boeck supérieur.
- Ramage, D., Rosen, E., Chuang, J., Manning, C. D., & McFarland, D. A. (2009). Topic Modeling for the Social Sciences. *Workshop on Applications for Topic Models, NIPS*, 4.
- Rashidian, N., Tsiveriotis, G., Brown, P., Bell, E. J., & Hartstone, A. (2020). *Platforms and Publishers. The End of an Era* (p. 62). Tow Center for Digital Journalism.
<https://academiccommons.columbia.edu/doi/10.7916/d8-sc1s-2j58>
- Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2931-2937.
<https://doi.org/10.18653/v1/D17-1317>
- Reynolds, C. (2021, March 29). Facebook Canada head warns news posts could be blocked as last resort. *The Globe and Mail*.

<https://www.theglobeandmail.com/canada/article-facebook-canada-head-warns-news-posts-could-be-blocked-as-last-resort/>

Roose, K., & Isaac, M. (2021, February 10). Facebook Dials Down the Politics for Users. *The New York Times*.

<https://www.nytimes.com/2021/02/10/technology/facebook-reduces-politics-feeds.html>

Roy, J.-H. (2021a, February 23). *Facebook vs. Australia-Canadian media could be the next target for ban*. The Conversation.

<http://theconversation.com/facebook-vs-australia-canadian-media-could-be-the-next-target-for-ban-155728>

Roy, J.-H. (2021b). Instagram: La une de l'ère mobile. *Les Cahiers du journalisme, Nouvelle série*(6), R69-R97.

[https://doi.org/10.31188/CaJsm.2\(6\).2021.R69](https://doi.org/10.31188/CaJsm.2(6).2021.R69)

Silva, M. (2021, January 6). *Open letter-Update on the News Media Bargaining Code in Australia-Google*. About Google.

<https://www.google.com/google-in-australia/jan-6-letter/>

Singer, J. B. (2006). The Socially Responsible Existentialist. *Journalism Studies*, 7(1), 2-18.

<https://doi.org/10.1080/14616700500450277>

Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, 2214-2218.

Willaime, J.-P. (2017). Polarisé autour du religieux en francophonie ? Les nouveaux défis de la laïcité. *Francophonies d'Amérique*, 44-45, 137-162. <https://doi.org/10.7202/1055908ar>

- Zelizer, B. (2005). Definitions of journalism. In G. Overholser & K. H. Jamieson (Eds.), *The Institutions of American Democracy: The Press* (pp. 66-80). Oxford University Press.
- Zuckerberg, M. (2017, February 16). *Building Global Community*. Facebook. <https://www.facebook.com/notes/3707971095882612/>
- Zuckerberg, M. (2018, January 19). Continuing our focus for 2018 to make sure the time we all spend on Facebook is time well spent... *Facebook*.
<https://www.facebook.com/zuck/posts/10104445245963251>