The Evolution of Bias in French News Media: How Does Political Orientation Affect Semantic Change?

Matej Martinc[1], Adelie Laruncet[2], Clara Bordier[2], Marceau Hernandez[2], Senja Pollak[1], Gaël Lejeune[2]

[1]Jozef Stefan Institute - {name}.{surname}@ijs.si

[2]CERES Sorbonne Université - {name}.{surname}@sorbonne-universite.fr

Abstract

In this study, we explore the hypothesis that political bias can significantly influence the temporal change in the usage of specific words. By conducting a temporal analysis on the corpus of French news media articles covering the topics of freedom of speech and conspiracy theory, we investigate how the coverage of these two topics has changed through the years, and point out the main differences in the evolution of the coverage in the left- and right-wing media. By identifying the words whose meaning evolved most differently in media with different political bias and by automatic labelling of the main usages of these words, we show that media's political bias has a substantial influence on temporal semantic shift.

Keywords: semantic change detection, media analysis, bias detection.

1. Introduction

All living languages are continually evolving and this change has become an important research topic, since it reflects changes in the political and cultural sphere. The studies of language evolution can focus on long-term shifts in the meaning of a word, or on short-term evolutionary changes in the usage of a word, where, e.g., a word suddenly appears in a new context, while keeping its meaning unchanged. Recently, several studies explored these short-term evolutionary phenomena in news media and showed that specific events can trigger drastic changes in word usage (Martinc et al., 2019; Martinc et al. 2023; Montariol et al., 2021). However, none of these studies investigated if political bias of the media affects word's diachronic semantic shift and if it does, what are the main differences in temporal language evolution across the political spectrum. To tackle this research gap, in this study we explore the hypothesis that political bias can significantly influence the temporal change in the usage of specific words and try to identify the main differences in evolution in the coverage of specific topics in the left-wing and right-wing media.

The temporal analysis is conducted on the corpus of French news media articles extracted from the Europresse database. The articles from the period between January 2012 and November 2023 cover the topics of freedom of speech (FOS) and conspiracy theory (CT). The first topic, which has a long tradition in the French news media, was chosen, since recently, especially since the Charlie Hebdo attack, any controversy related to the FOS elicits visceral reactions and significant polarisation in society. The second topic of CT was on the other hand selected due to the constant increase in coverage, especially from the 2010s onwards, making CT a real problem in today's society.

To identify words that evolved differently in the left- and right-wing media, we leverage the state-of-the-art natural language processing (NLP) technique for detection of word's semantic change proposed in Montariol et al. (2021). Finally, we interpret the identified differences in evolution by labelling the main usages of the most semantically changed words in different time periods and in media with different political bias.

The main contributions of the study are the following:

- We propose a new methodology for detection of differences in temporal evolution in media with different biases and show that political bias has a substantial influence on short-term temporal evolution of word usage.
- The study contributes to the research of media coverage of freedom of speech (FOS) and conspiracy theory (CT) by highlighting the variations in the way right-wing and left-wing media approach these issues, and by linking the changes in coverage to specific events.
- As far as we are aware, this is the first study that employs NLP techniques for detection of diachronic semantic change in French media.

2. Related Work

When it comes to detection of political bias, there are several general news analysis techniques that can be employed. One of the first systems capable of detecting semantic differences for the purpose of viewpoint analysis of political and media discourse was developed by Azarbonyad (2017). More recently, a study by Spinde et al. (2021) tried to identify biased terms in news articles by comparing news media outlet specific word embeddings. Some studies also tried to determine the main differences between "liberal" and "conservative" topics. For example, a study of 3.8 million Twitter messages (Sterling, 2019) showed some distinctions between liberal and conservative representations of the "good society". Several studies also determined that LGBTIQ+ topics are covered differently in left- and ring-wing media. For example, the study by Martinc et al. (2020) employed a technique for detection of most differently used words to find the main differences in the coverage of LGBTIQ+ topics in Slovenian media with different political bias.

Recently, several models for semantic change detection, which analyse temporal changes in usage of words, have been proposed. In order to construct temporal representations, contemporary work employs two distinct methodologies, i.e. the use of either static or contextual embeddings. In the approach employing static embeddings, you train a distinct static embedding model for each temporal slice of the corpus and then make these temporal representations comparable by employing either *incremental updating* (Kim et al., 2014) or *vector space alignment* (Hamilton et al. 2016). On the other hand, with the rise of contextual embeddings such as BERT (Devlin et al., 2019), several approaches for semantic shift have been proposed that rely on vector representations that consider context and therefore also allow to model polysemy. In these approaches, the information from token embeddings is aggregated into temporal representations by, e.g., simple averaging (Martinc et al. 2019), by a pairwise comparison of vectors from different time slices (Kutuzov et al., 2020), or by clustering of token embeddings (Giulianelli et al., 2020; Montariol et al., 2021). The latter clustering approach is also employed in our work.

We are not aware of any studies that would address changes in media reporting about FOS and CT. While the research into CT has flourished in recent years, it is mainly theoretical or about the circulation of CT and analysis of the culture built around them. When it comes to FOS, most existing research focuses primarily on its historical or legal context (Korolitski, 2015). On the other hand, we take a more practical discursive constructionist approach, which sees the concepts of FOS and CT as social constructs (Bratich, 2020) and examines the role of the media in this construction.

3. Data

The temporal analysis is conducted on the corpus of French language news media articles extracted from the Europresse database (https://www.europresse.com). As stated in Section 1, the articles from the period between January 2012 and November 2023 cover the topics of FOS and CT. To obtain a corpus on the topic of FOS, we extracted articles written by four biggest daily French general-interest newspapers with a well-known political bias that contained words "liberté d'expression". To obtain a corpus on the topic of CT, we extracted articles containing words "conspirationnisme" and "conspirationniste". Due to lack of available data, we expanded the search to cover not only the main French daily newspapers, but all French language media available in the Europresse media. Since this search returned news articles from several news media that no longer exist and/or were hard to verify, we additionally decided to limit the corpus to 20 newspapers with the most retrieved articles and a well defined and known media bias. The final corpus contains the biggest mainstream French and Canadian newspapers and also some smaller media sources known for less factual reporting.

We split the corpora into four distinct chunks according to two axes, political and temporal. We used Wikipedia to split the media in the corpora into left- and right-wing and discarded media close to the centre or media with unknown bias. We also opted to split the media into two temporal chunks, one containing articles between 2012 and 2019 and the other containing articles between 2020 and November 2023. The split coincides with the beginning of the Covid pandemic in Europe in January 2020, which has drastically affected media reporting and also caused several changes in usage of specific words (Montariol et al., 2021). The corpora statistics are described in Tables 1 and 2.

Tables 1 and 2: Number of articles in FOS corpus (above) and CT corpus (below).

Source	Political bias	2012-2019	2020-2023	All
La Croix	right	556	343	899
Le Figaro	right	930	719	1649
Les Echos	right	305	279	584
All right	right	1791	1341	3132
Libération	left	958	487	1445
l'Humanité	left	645	340	985
All left	left	1603	827	2430
All	/	3394	2168	5562

Source	Political bias	2012-2019	2020-2023	All
Courrier International	left	12	45	57
HuffPost	left	16	47	63
ICI Radio-Canada	left	40	135	175
L'Obs	left	62	116	178
La Presse	left	15	26	41
La Presse+	left	34	109	143
Le Soir	left	31	49	80
Le Soleil	left	12	35	47
Libération	left	119	197	316
Télérama	left	28	37	65
l'Humanité	left	19	28	47
All left	left	388	824	1212
Atlantico	right	46	14	60
L'Express	right	67	123	190
La Croix	right	26	68	94
Le Figaro	right	177	204	381
Le Journal de Montréal	right	6	56	62
Le Journal de Québec	right	8	65	73
Le Point	right	119	85	204
Les Echos	right	32	40	72
Valeurs Actuelles	right	38	42	80
All right	right	519	697	1216
All	/	907	1521	2428

4. Methodology

4.1. Measuring semantic changes

We employ a system originally employed for diachronic shift detection (Montariol et al., 2021) and modify the pipeline to work for French corpora and to consider political bias. In the first step, instead of splitting corpora just into temporal slices, each corpus is split into four chunks according to temporal and political axes, as explained in Section 3. We lowercase the corpus, and use the spaCy library (https://spacy.io/) for lemmatization and part-of-speech tagging. Note that in contrast to the original study by Montariol et al. (2021), we only consider nouns in this study, in order to reduce the noise in the results due to the smaller size of the available corpora.

For each noun lemma (i.e., target word) that is not considered a stopword and appears more than 20 times in each corpus slice, we generate a chunk specific set of contextual embeddings.

More specifically, in the first step, we employ the French language model CamemBERT (Martin et al., 2019) to obtain one contextual embedding representation for each appearance of a word in the corpus. The model is fed 256 tokens long sequences in batches of 8 sequences at once. We generate sequence embeddings by summing the last four encoder output layers. Next, we split each sequence into 256 subparts to obtain a separate contextual embedding of size 768 for each token. Since one token does not necessarily correspond to one word due to byte pair tokenization, we average embeddings for each byte-pair token constituting a word to obtain embeddings for each occurrence of a word.

In the next step, we check for each obtained embedding whether it corresponds to a noun lemma. If it does not, the embedding is discarded. The embeddings that are not filtered out are then aggregated into lists according to the corresponding lemma and the specific political/temporal chunk they belong to. A the end of this step, we obtain one list of contextual embeddings per noun lemma in a specific temporal/political corpus chunk.

These lists from different chunks that contain embeddings corresponding to the same target words are concatenated (i.e., we obtain one list of embeddings per each noun lemma in the corpus) and the k-means clustering algorithm (with k=5) is ran on the concatenated list. We also employ an additional cluster filtering and merging step described in Montariol et al. (2021). After obtaining the cluster labels for each embedding, the concatenated list is again split into chunk specific lists of embeddings with assigned cluster labels. We count the number of embeddings belonging to a specific cluster in each of these lists. Finally, we normalize the cluster counts in each list in order to obtain a cluster distribution reflecting main usages of a specific lemma for each chunk.

In contrast to the original study by Montariol et al. (2021), the cluster distributions are not compared only across the temporal axis to obtain temporal semantic shifts, but also across the bias axis. More specifically, we are interested in the following word usage comparisons:

- **PreCOVID bias (PCB) comparison:** Word usages from left wing media articles between 2012 and 2019 (*2012/19 left*) are compared to word usages from right wing articles between 2012 and 2019 (*2012/19 right*).
- **AfterCOVID bias (ACB) comparison:** Word usages from left wing media articles between 2020 and 2023 (*2020/23 left*) are compared to word usages from right wing articles between 2020 and 2023 (*2020/23 right*).
- **Temporal left (TL) comparison:** Word usages from left wing media articles between 2012 and 2019 (*2012/19 left*) are compared to word usages from left wing articles between 2020 and 2023 (*2020/23 left*).
- **Temporal right (TR) comparison:** Word usages from right wing media articles between 2012 and 2019 (*2012/19 right*) are compared to word usages from right wing articles between 2020 and 2023 (*2020/23 right*).

As in Giulianelli et al. (2020), cluster distributions are compared across slices by employing the Jensen-Shannon Divergence (JSD), a measure of similarity between two probability distributions. All words in the vocabulary are ranked according to the JSD distance and it is assumed that this ranking resembles a relative degree of usage change (i.e., a larger distance between distributions in two slices reflects stronger semantic change). We use JSD scores obtained in comparisons between distributions described above in four distinct equations that measure different aspects of political bias's influence on temporal semantic change:

$$Just\ temporal\ shift = \frac{(\text{TL JSD} + \text{TR JSD}) - (\text{PCB JSD} + \text{ACB JSD})}{2} \tag{1}$$

$$\textit{Just left shift} = \text{TL JSD} - \text{TR JSD} \tag{2}$$

$$\textit{Just right shift} = \text{TR JSD} - \text{TL JSD} \tag{3}$$

$$\textit{Similar to different} = \text{ACB JSD} - \text{PCB JSD} \tag{4}$$

$$\textit{Different to similar} = \text{PCB JSD} - \text{ACB JSD} \tag{5}$$

Using the *Just temporal shift* equation*,* we rank words in the corpus vocabulary according to the unbiased temporal shift, which allows us to identify words that have changed a lot between two distinct time periods and were (mostly) unaffected by the media bias. On the other hand, using the *Just left shift* equation, we identify words with significant change in usage between time periods in the left-wing media and rather constant usage in the right-wing media. Vice versa, with the *Just right shift* equation, we can identify words that only changed usage in right-wing media. By employing the *Similar to different* formula, we can find words that have a very similar usage in the left- and right-wing media before the COVID pandemic, but are used differently in the left- and right-wing media after the beginning of the COVID pandemic. Finally, with the *Different to similar* equation we obtain words with very different usage before the pandemic and very similar usage after the start of the pandemic.

4.1. Interpretation of semantic changes

Once the most changed words are identified, we need to interpret how their usage differs in distinct corpus chunks. Assuming that k-means clusters of CamemBERT contextual embeddings resemble word usages of a specific word, we need to assign labels to the clusters that reflect these usages. Since these clusters may consist of several hundreds of word usages (i.e. sentences), manual inspection of these usages would be time-consuming. To avoid it, we employ the procedure proposed in Montariol et al. (2021) and automatically extract the most discriminating unigrams, bigrams and trigrams for each cluster by computing the term frequency - inverse document frequency (TF-IDF) score of each n-gram in the cluster. We obtain up to 10 n-grams for each cluster and remove n-grams that appear in more than 70% of clusters to ensure keyword diversity. The final output is a ranked list of (English translated) keywords for each cluster and the top-ranked keywords (according to TF-IDF) are used as cluster labels that allows us to interpret the specific usage of a word in the cluster.

4. Results

In Table 3, we present the word ranking results for our experiments on the FOS and CT corpora. When just temporal shift or just semantic shift in the right-wing media is considered, the most

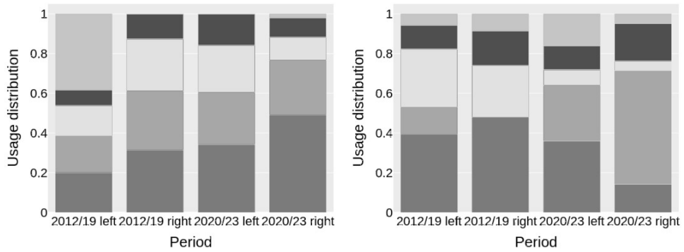
Table 3: Ranking of words in the CT and FOS corpora according to "Just temporal shift", "Just left shift" and "Just right shift" criteria. Words exposed to more semantic change are ranked better.

	Conspiracy theory			Freedom of speech		
Rank	Just temporal shift	Just left shift	Just right shift	Just temporal shift	Just left shift	Just right shift
1	fuite(leak)	banque(bank)	geste(gesture)	tueur(killer)	village(village)	tueur(killer)
2	géant(giant)	vaccin(vaccine)	info(information)	entrepreneur(entrepreneur)	dégradation(degradation)	écriture(writing)
3	geste(gesture)	reprise(reprise)	sol(ground)	assassin(assassin)	émeute(riot)	lanceur(launcher)
4	professeur(teacher)	aide(help)	surprise(surprise)	biais(bias)	incident(incident)	réalisateur(director)
5	pièce(piece)	pression(pressure)	primaire(primary)	transition(transition)	parcours(course)	affiche(poster)
6	reprise(reprise)	religion(religion)	examen(exam)	retraite(retirement)	moitié(half)	complice(partner in crime)
7	mesure(measure)	obsession(obsession)	piste(track)	séance(session)	bas(down)	torture(torture)
8	impact(impact)	détail(detail)	policier(police officer)	réfugié(refugee)	investissement(investment)	conversation(conversation)
9	rassemblement(gathering)	caractère(character)	ton(your)	milliardaire(billionaire)	distinction(distinction)	mineur(minor)
10	opposition(opposition)	secret(secret)	désinformation(disinformation)	complot(conspiracy)	discipline(discipline)	salaire(salary)
11	vaccin(vaccine)	fuite(leak)	construction(construction)	conseiller(advise)	conseiller(advise)	plateau(plateau)
12	voiture(car)	carte(map)	prison(prison)	compétition(competition)	statue(statue)	hiérarchie(hierarchy)
13	sol(ground)	signe(sign)	abonné(subscriber)	lanceur(launcher)	troupe(troop)	voyage(journey)
14	euro(euro)	nature(nature)	université(university)	nez(nose)	nez(nose)	autrui(others)
15	surprise(surprise)	position(position)	dollar(dollar)	directive(directive)	perspective(perspective)	accusé(accused)

changed word in the FOS corpus is *tueur (killer)*. On the other hand, the word that changed its usage the most in the left-wing media is *village*. In the CT corpus, when considering just the diachronic shift, the most changed word is *fuite (leak)*. *Banque (bank)* is at the top of the ranking of words that experienced a significant shift in usage in the left-wing media, while remaining mostly unchanged in the right-wing media. *Geste (gesture)*'s usage on the other hand changed the most in the right-wing media and remained relatively constant in the left-wing media.

Figure 1 gives us insight into the specific usage changes to which the two highest ranked words in the FOS corpus were exposed. The four columns in the figure are presenting usage distributions across political and temporal axes, namely usages in left-wing media before (*2012/2019 left*) and after (*2020/23 left*) the COVID pandemic, and usages in the right-wing media before (*2012/2019 right*) and after (*2020/23 right*) the pandemic. The word *tueur (killer)*, which is the most changed word if "just temporal shift" and "just right shift" criteria are applied, ranked highly due to the extensive coverage of the "Taiwanese church shooting" in California on May 15, 2022 (this usage is represented by the cluster "Taiwan, theory, fate…"), especially by the right-wing media. In the preCOVID period, this word was mostly used in reference to the Charlie Hebdo Bataclan terrorist attacks. The word that changed the most in the left-wing media is village. We can observe a disappearance of the cluster "author, humanity, meeting…" in the left-wing media after the pandemic, which mostly covers the usage of the word village in a context of the "village du livre" (book town), i.e., a town or village with many used book or antiquarian bookstores.

Additionally, we observe an enlargement of the "moment, global village…" cluster in both left- and right-wing media. This cluster covers usages in the context of a "global village", a concept expressing the idea that people throughout the world are interconnected through the use of new media technologies, and also usages connected to the Elon Musk's referral to Twitter as a "virtual village square", which requires safeguarding from the perils of content moderation policies, which he perceives as a threat to free speech.



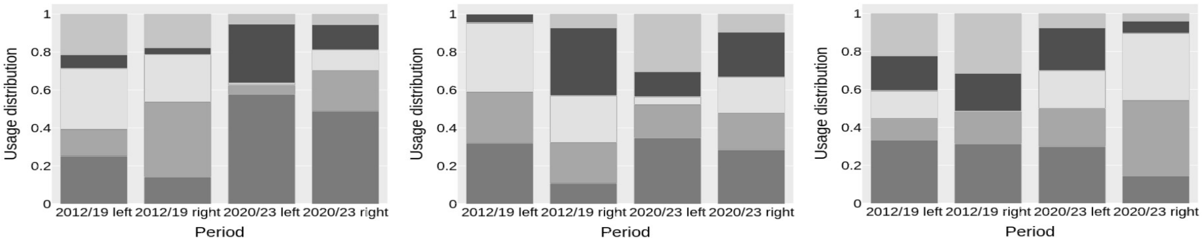
#	Village
	author, humanity, meeting, Greek, karelle, health, Nicolas, literary village, Jack Ralite
	head, plane, income, cemetery, countryside, time, 19th century, see disembark, see flock, home visit
	work, classify, army, small village, destroy, theater, revival, village province, new, Beg Meil
	young, small village, pass, freedom, value, tell, father, vital, Turkey, die
	moment, global village, father, arrive, want, village water, family, twitter, young, perfectly

#	Tueur (killer)
	seem, count, killer virus, killer happen, deputy sheriff, reveal today, thunderous laughter, power obviously, power in, point name
	synagogue, against, attack, terrorist, evidence, time, protect, Norwegian, extremism, expression of power
	Bataclan, reference, large, France, religion, share, appear, note, murder, jihadist
	Taiwan, theory, fate, march, woman, David Chou, learn, write, crime, year
	Charlie, have to, find, against, refuse, never, nice, designer, community

Figure 1: Word usages of highest ranked words according to "Just temporal shift" (Tueur), "Just right shift" (Tueur), and "Just left shift" (Village) criteria in the FOS corpus.

Figure 2 gives us insight into specific usage changes to which the three highest ranked words in the CT corpus were exposed. The word *leak* was mostly used in the context of "information leak" (see cluster "forum, document, hacking…") before the pandemic, and in the context of a "laboratory virus leak" (see cluster "laboratory, hypothesis, virus…") in the period after the pandemic. This change in usage was consistent in both left-wing and right-wing media. On the other hand, the word *banque (bank)* obtained a new usage only in the left-wing media after the pandemic, where a significant new cluster (see cluster "green hope, green recovery…") appeared, presenting the usage of bank in the context of "green transition in Europe". The word *geste (gesture)* experienced a shift in usage mostly in the right-wing media. One can see the increase in usage of the word gesture after the pandemic in the context of "mask wearing" and in relation to the Youtube platform (see cluster "mask, Youtube…"), which is most likely connected to the phrase "geste barrière" used in the government (Youtube) campaign for mandatory mask wearing during the COVID pandemic and the fact that this campaign was advertised on Youtube by the famous entertainer Macfly&Carlito to reach younger audiences.

In Table 4, we present results of word ranking in both corpora, when "Similar to different" and "Different to similar" criteria are used. On top of the ranking in the Conspiracy theory corpus are the words *Info (information),* which was used in a very similar context by the left- and right-wing media before the pandemic, and is used very differently after the pandemic, and the word *taux (rate),* which was used in different context by media with opposite political bias before the pandemic, and in a very similar context after the pandemic. In the FOS corpus, the best ranked words according to the "similar to different" and "different to similar" criteria are words *horizon* and *village*, respectively. While we discussed the specific changes in usages for the word *village* above (see Figure 1), we present the specific changes in usage of the other three words in Figure 3.



#	Fuite (leak)
	penny, China, massive leak, prepare, Danish island, possible leak, elected republican, William Bourton, WIFI, suspect arrest
	laboratory, hypothesis, possibility, scientific, pandemic, affirm, virus, transmission, probable, WHO
	forum, document, hacking, may, hack accessible, Hillary Clinton, English 4chan, accessible penny, 2017, national
	policeman, place, times, force, case, riot, title, shooter, dye, Sarah Halimi
	laboratory, hypothesis, virus, origin, transmission, pandemic, natural, accidental leak, theory, scientific

#	Banque (bank)
	green hope, green recovery, European advocacy, European central bank, banking theory, technology, law, deep state, infinite energy, globalist elite
	government, world, conspiracy, union, France, financial, euro, Rothschild bank, power bank, law
	Goldman, american bank, become bank, jewish emigrant, see logically, tool, month, jew, enemy finance, emblematic figure
	hear, know, power, percent, thing, ask, china, united state, discover, director
	resume, Wuhan, currency, grace, people, capitalist bank, thriving economy, tumor, Sunni, adequate reaction

#	Geste (gesture)
	surprise, left, denounce, theory, gesture signify, post, pretend brutality, state, Renaud writer, want to do
	nurse, novel, origin, justice, point, account, act, stranger, come, system
	behavioral, see, distance, start, violent gesture, measure, risk, minister, mask, statue
	mask, Youtube, repair, small gesture, child, protester, human gesture, train, hear, last
	Sunday, quenelle, affirm, selfie, ignore, consider, explain, supremacist, emphasize

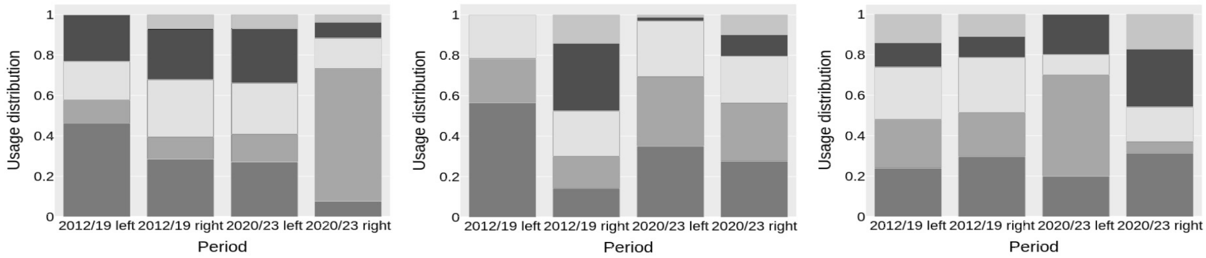
Figure 2: Word usages of highest ranked words according to "Just temporal shift" (Fuite), "Just left shift" (Banque) and "Just right shift" (Geste) criteria in the CT corpus.

Table 4: Similar words that evolved into different and different words that evolved into similar.

	Conspiracy theory		Freedom of speech	
Rank	Similar to different	Different to similar	Similar to different	Different to similar
1	info(information)	taux(rate)	horizon(horizon)	village(village)
2	photo(photo)	courant(fluent)	sang(blood)	distinction(distinction)
3	assassinat(assassination)	champ(field)	animateur(animator)	maladie(disease)
4	écran(screen)	promotion(promotion)	croyant(believer)	foulée(stride)
5	mode(fashion)	réflexion(reflection)	chroniqueur(columnist)	directrice(director)
6	camp(camp)	fil(thread)	entretien(interview)	augmentation(increase)
7	application(application)	fan(fan)	entourage(entourage)	concurrence(competition)
8	sol(ground)	saison(season)	destruction(destruction)	troupe(troop)
9	terrain(ground)	tension(tension)	café(coffee)	plateau(plateau)
10	geste(gesture)	nature(nature)	écrit(writing)	désinformation(disinformation)
11	évidence(evidence)	bataille(battle)	littérature(literature)	convention(convention)
12	remplacement(replacement)	piste(track)	séance(session)	paysage(landscape)
13	militaire(military)	policier(police officer)	rappel(reminder)	ingérence(interference)
14	conférence(conference)	comportement(behavior)	énergie(energy)	capitalisme(capitalism)
15	passé(pass)	roman(novel)	contradiction(contradiction)	cabinet(office)

For the word *info,* we observe a distinct change in usage in the right-wing media after the pandemic (see cluster "tool, Jones, conspiracy…"), with an extensive coverage of Alex Jones', an alt-right radio show host and prominent conspiracy theorist, view on information and fake news. Vice versa, the word *taux (rate)* in the CT corpus is used in very different contexts before the pandemic (i.e. in the context of health by the left-wing media and in the economic and financial context by the right-wing media), but becomes used in the same (mostly health and pandemic) context after the pandemic.

In the FOS corpus, the word *horizon* obtains a new majority usage in the left-wing media after the pandemic (see cluster "speech, watch…"). A closer manual inspection of this somewhat messy cluster revealed that the word *horizon* is mostly discussed in the context of a need for a new common horizon for all (young) French people, in order to repel them from "l'horizon apparemment aussi terriblement enchanteur que macabre de l'État islamique" [the horizon apparently as terribly enchanting as it is macabre of the "Islamic State"] (l'Humanité) and nationalist movements.



#	Info (information) - in conspiracy theory
	enter, superior, law, red carpet, Russia info, electoral period, concern, professor Raoult
	Nordpresse, becoming, internet user, media, fake news, documentary, video, theory, enter, year
	info here, Figaro, culture, Tiktok, Lionel Vighier, Fox news, conspiracy, completely, real info, info publish
	tool, Jones, conspiracy, times, fake news, right, broadcast, theory, French, conspiracy
	know, establish, TV news, remember, Tuesday, day, free, gesture, exceptional news, antenna

#	Taux (rate) - in conspiracy theory
	interest, bank, export, central, lose, monetary, Europe, part
	short rate, interest, long rate, financial, everywhere, term, know how to adjust, find, start again
	last, September, France, definition, vaccination, microcephaly, positive, confirmed diagnosis, study, United States
	unemployment, illness, understand, too much, publish, gGeat britain, reach, arrive, study, public
	cholesterol, against, number, statin, cardiovascular, publish, average, incidence, patient, million

#	Horizon - in freedom of speech
	testify, child, wellness, sexual violence, feeling, resume, affair
	grow, voter, globalize world, cognitive horizon, society, share, recently emerge, separate element, populist electorate, we dream
	tower, social, citizen, situation, european, present, small, war, good, project
	speech, watch, student discover, Islamic state, nationalist wave, Thomas Canto, terribly enchanting
	society, world, glow, enter, far, dimension, reach, too, country, spiritual

Figure 3: Word usages of highest ranked words according to the "Similar to different" (Info and Horizon), and "Different to similar" (Geste) criteria in the CT and FOS corpora.

We also observe more usages of the word "horizon" in the context of 'globalization' (see purple cluster) in right- and (to a lesser extent) left-wing media. Here, "horizon" has a different meaning and refers to people's "background", e.g. "On est dans un monde globalisé , avec des gens d' horizons différents , qui ne se comprennent pas toujours."[We are in a globalized world, with people from different backgrounds, who do not always understand each other.] (l'Humanité).

5. Conclusion

We have tested the hypothesis that political bias of news media can have a significant effect on diachronic semantic shift of some words. The experiments conducted on the French news corpora, which focused on short-term changes in usage before and after the COVID pandemic, suggest that there are indeed several words with very different evolution in right-wing and left-wing media, confirming the hypothesis.

When it comes to the specific topic of CT, one can see a distinct evolution in the right-wing media, which have substantially increased the coverage of themes related to the supposed "fake news". The study's findings might also suggest that right-leaning media tend to pay more attention to FOS in connection with external threats such as terrorism, while the left-leaning media appear to still be more focused on internal French issues. In the future, we will expand the research to other languages and topics, to further test the generality of the proposed hypothesis. We will also test whether this method is applicable to larger corpora. Finally, we will experiment with other systems for semantic change detection, to determine whether the results are consistent across different methods.

Acknowledgements

We acknowledge the financial support from the Slovenian Research Agency (ARRS) core research programs Knowledge Technologies (P2-0103), programme Proteus (French-Slovene Scientific Research Support Scheme), as well as projects CANDAS (Computer-assisted multilingual news discourse analysis with contextual embeddings, No. J6-2581), SOVRAG (Hate speech in contemporary conceptualizations of nationalism, racism, gender and migration, No. J5-3102) and EMMA (Slovenian Research and Innovation Agency research project Embeddings-based techniques for Media Monitoring Applications, No. L2-50070).

Bibliography

Hosein Azarbonyad et al.. (2017). Words are malleable: Computing semantic shifts in political and media discourse. *In Proceedings of the 2017 CIKM Conference*, pp. 1509-1518.

Jack, Bratich. (2020). Civil Society Must Be Defended: Misinformation, Moral Panics, and Wars of Restoration. Communication, Culture & Critique. doi:10.1093/ccc/tcz041

Jacob Devlin et al.. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171-4186.

Mario Giulianelli, Marco Del Tredici, and Raquel Fernández. (2020). Analysing lexical semantic change with contextualised word representations. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3960-3973.

William L. Hamilton, Jure Leskovec and Dan Jurafsky. (2016). Diachronic word embeddings reveal statistical laws of semantic change. *In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1489-1501.

Yoon Kim et al.. (2014). Temporal analysis of language through neural language models. *In Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pp. 61-65.

Ulysse Korolitsky. (2015). Punir le racisme. Liberté d'expression, démocratie et discours racistes. CNRS éditions.

Andrey Kutuzov et al.. (2020). UiO-UvA at SemEval-2020 task 1: Contextualised embeddings for lexical semantic change detection. *Proceedings of the 14th SemEval Workshop*, pp. 126-134.

Louis Martin et al.. (2019). CamemBERT: a tasty French language model. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7203-7219.

Matej Martinc, Petra Kralj Novak and Senja Pollak. (2019). Leveraging contextual embeddings for detecting diachronic semantic shift. *Proceedings of the 13th LREC conference*. pp. 4811-4819.

Matej Martinc et al.. (2020). EMBEDDIA hackathon report: Automatic sentiment and viewpoint analysis of Slovenian news corpus on the topic of LGBTIQ+. *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation*, pp. 121-126.

Matej Martinc et al.. (2023). A meaty discussion: quantitative analysis of the Slovenian meat-related news corpus. *Proceedings of the Slovenian KDD Conference*, pp. 50-53.

Syrielle Montariol, Matej Martinc and Lidia Pivovarova. (2021). Scalable and interpretable semantic change detection. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2021, pp. 4642-4652.

Timo Spinde, Lada Rudnitckaia and Felix Hamborg. (2021). Identification of biased terms in news articles by comparison of outlet-specific word embeddings. *In Proceedings of the 16th International Conference (iConference 2021)*.

Joanna Sterling et al.. (2019). Liberal and Conservative Representations of the Good Society: A (Social) Structural Topic Modeling Approach. SAGE Open, 9(2).