A Vocabulary Analysis of News Articles in Relation to the Political Orientation of Their Source and their Thematic

Laurène CAVE, Gaël LEJEUNE
Sorbonne Université
Maison de la Recherche, rue Serpente, 75006 Paris, France
laurene.cave@etu.sorbonne-universite.fr, gael.lejeune@sorbonne-universite.fr

Abstract

Understanding how political orientation influences lexical choices is essential for detecting bias and framing in news media. In this paper, we present a computational framework for identifying nouns whose interpretation varies across politically divergent newspapers. Using a large corpus of French news articles published in 2024, we categorize texts by topics and political orientation. We use contextual embeddings to cluster occurrences of nouns to detect semantic variations and dissimilarity among sources. This allows us to map semantic distances between newspapers and identify polarized or editorially marked lexical choices. Our results show that topics, polysemy, and editorial priorities contribute differently to lexical divergence. We discuss these findings and highlight how contextual embeddings can help reveal semantic biases that would remain invisible through frequency-based methods. We conclude by outlining perspectives for improving topic classification and the clustering method, exploring alternative divergence measures, conducting a qualitative analysis of our results, and extending the framework to other languages or genres.

Keywords: Semantic Shift, Contextual embeddings, Media Bias, Framing, Political Discourse, News Corpora

1. Introduction

Word choices in political debates both reflect and shape ideological positions. The same word or phrase can convey different meanings depending on the speaker's or writer's political stance, as well as the audience's own perspective. For instance, terms such as *freedom*, *immigration*, or *climate policy* often carry ideological undertones that vary across the political spectrum. Understanding how political orientation influences lexical choice and interpretation is therefore crucial for the study of media discourse and online political debates.

In recent years, the increasing availability of large-scale online text corpora has enabled systematic investigations of language use in political and media contexts. Computational methods have been employed to map ideological differences in word usage (Gentzkow and Shapiro, 2010; Kulkarni et al., 2018), detect framing and bias in news reporting (Card et al., 2015), and measure polarization in online discourse (Budak et al., 2016; Demszky et al., 2019). However, most existing studies focus on explicit sentiment or stance detection rather than on the subtler semantic shifts that arise when politically divergent sources employ ostensibly similar language.

This paper proposes a computational framework for identifying words and expressions whose interpretation varies according to political orientation. We analyze large-scale corpora of newspapers representing left and right-leaning positions, investigating how ideological differences manifest in both lexical semantics and thematic framing. Our focus extends beyond isolated keywords to encompass contextual variation across different topics such as politics, the environment, technology, and culture. We aim to detect where framing choices and word associations are especially salient—whether in specific topics or sources.

We employ contextual embedding models to capture semantic nuances and meaning shifts across political subcorpora. By comparing the embeddings of terms whose usage spans sources with different political leanings, we quantify how their semantics diverge. This allows us to identify *polarized words* — terms whose meaning or connotation systematically depends on political stance — as well as sources that exhibit minority opinions within particular topics.

The remainder of this paper is organized as follows. Section 2 presents related research. Section 3 details our methodology, and Section 4 describes the press corpus we compiled. In Section 5, we provide a representation of the main French newspapers relative to one another based on their lexical differences. Section 6 offers a deeper analysis of word usage variation across newspapers. We summarize our findings and outline future perspectives in Section 7.

2. Related works

Artificial intelligence tools have already been tested in the context of democratic debate. In their paper, (Tessler et al., 2024) present a large language model called the Habermas machine, designed to

Proceedings of The 3rd Workshop on Natural Language Processing for Political Sciences (PoliticalNLP 2026) @ LREC 2026, pages 111–119
16 May 2026, Palma, Mallorca (Spain). © ELRA Language Resources Association (ELRA), 2026

help discussion groups find common ground and reach agreements. This AI mediator was able to generate more consensual statements than human moderators, thereby reducing divisions among participants. AI has also been used to improve the quality of online political discussions. (Argyle et al., 2023) developed a chat assistant that provided suggestions to participants engaged in conversations on polarizing political topics, with the goal of fostering mutual understanding. These interventions improved the tone of the conversations and participants' feelings about the interaction, without necessarily changing their political positions. Large language models (LLMs) are capable of detecting subtle aspects of text, such as politeness (Yeomans et al., 2018), making them well-suited for tasks like automated moderation. However, as highlighted in the report *Algorithmic Gatekeepers: Impacts of LLM Content Moderation on Civic Space and Human Rights* (Wisniak, 2025), while these tools can be highly effective, their use raises concerns, particularly around biases in LLMs such as disadvantages for minorities or non-dominant language speakers.

The political biases of AI tools are a growing concern. (Potter et al., 2024), for instance, show that LLMs tend to exhibit a preference for Biden over Trump, and more importantly, that Trump supporters tend to reduce their support after interacting with an LLM even if the model was not explicitly instructed to persuade them. As a result, several studies propose methods for measuring political bias and preferences in LLMs, whether through prompt-based techniques (Rottger et al., 2025), or by focusing on users' perceptions of the model's behavior (Sean J. Westwood, 2025). These biases are often difficult to avoid or anticipate; for example, training a model to prioritize factual accuracy may unintentionally amplify a left-leaning orientation (Fulay et al., 2024).

A possible approach to analyzing texts is not only to consider what is being said, but also how it is said. Stylistic analysis can complement content analysis to help identify political biases in large language models (LLMs) (Bang et al., 2024). The choice of terms used in a text plays a significant role, as it can influence readers by framing the presented information in particular ways. However, such framing can also be detected by LLMs, as demonstrated by (Baumer et al., 2015) in their work on political content. The usage and interpretation of certain terms depend not only on the context in which they appear but also on the time period and the ideology of the author (Azarbonyad et al., 2017). As a result, the same terms may be understood or employed differently by different groups. Additionally, the inherent polysemy and ambiguity of some words make it valuable to determine their intended meaning in a given occurrence (Daille et al., 2016). It is therefore unsurprising that interlocutors may sometimes need to pause a conversation to negotiate and clarify the meaning of specific terms (Aina Garí Soler, 2025). Providing participants with a tool to help navigate these potential misunderstandings could thus prove highly beneficial.

In order to clarify the ambiguous meaning of certain terms used in a debate, it is first necessary to detect them. Several methods exist for this purpose, particularly for identifying diachronic shifts in meaning (i.e., changes over time) (Hamilton et al., 2018; Montariol et al., 2021). More broadly, there are also approaches for detecting words whose usage varies across different documents (Monroe et al., 2017). These shifts in meaning and usage are especially relevant for us when they occur between individuals holding opposing viewpoints (Garí Soler et al., 2022; KhudaBukhsh et al., 2020).

3. Methodology

Our methodology is inspired by a recent approach that leverages language models to identify political biases in news corpora (Martinc et al., 2024). The first step is to create subcorpora relevant to our study based on both topics and political orientations.

We then use the `spaCy` library for lemmatization and part-of-speech tagging, retaining only nouns (including proper nouns). In future work, we may also include verbs and adjectives to enrich the analysis.

Next, we filter out stopwords, as they can sometimes be misidentified as nouns by the language model and introduce noise into the results. We retain only the terms that appear frequently enough to allow for meaningful comparison (with a threshold of 10 occurrences). This threshold ensures that we focus on words that occur sufficiently often across subcorpora to enable reliable semantic comparison.

For each retained term, we generate a set of contextual embeddings specific to each subcorpus using the French language model `CamemBERT`[1].

The embedding sets from all subcorpora are then concatenated and clustered using the k-means algorithm (Giulianelli et al., 2020) (with $k = 5$), grouping occurrences of a word that share similar contexts or meanings. An example of this clustering process is shown for the term *état* (*state*) in Figure 1.

We then apply a series of filtering steps. A cluster is retained only if it contains more than a minimum number of occurrences (10 in our case). Small clusters (with fewer than 10 occurrences) are merged

[1] `https://huggingface.co/docs/transformers/model_doc/camembert`

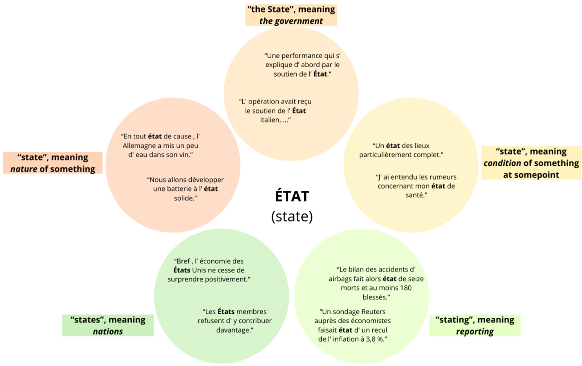

Figure 1: Clusters (with examples) for the term *état* (*state*).

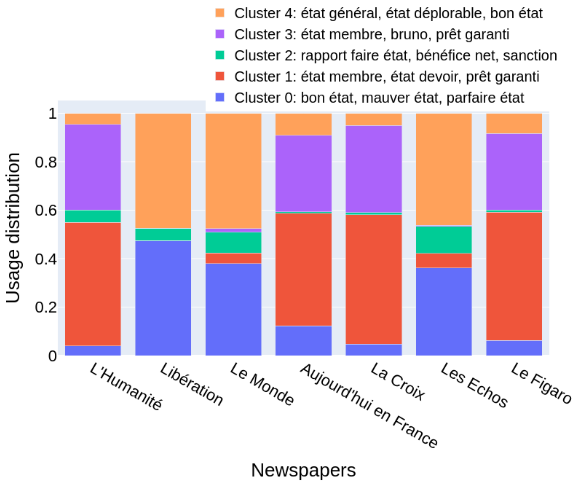

Figure 2: Distribution of clusters across 7 of the newspapers for the term *état* (*state*).

with the nearest larger cluster or with another small cluster if their distance is below a defined threshold (set as the average cosine distance between all large enough clusters minus twice the standard deviation). If no suitable cluster for merging is found, the small cluster is discarded.

Once the final clusters are obtained for each noun, we analyze their distribution across subcorpora. The number of embeddings per cluster is counted and normalized, producing a cluster distribution that reflects how each noun is used across the different subcorpora. Figure 2 illustrates the resulting distribution for the term *état* (*state*).

We then use the Jensen–Shannon Divergence (JSD) to measure the similarity between these distributions. The resulting *dissimilarity score* quantifies how differently a word is used across subcorpora: a score of 0 indicates identical distributions, while higher scores reflect greater divergence in usage.

Each cluster contains multiple sentences in which the noun appears. To facilitate interpretation, we compute term frequency–inverse document frequency (TF–IDF) scores for all terms in the

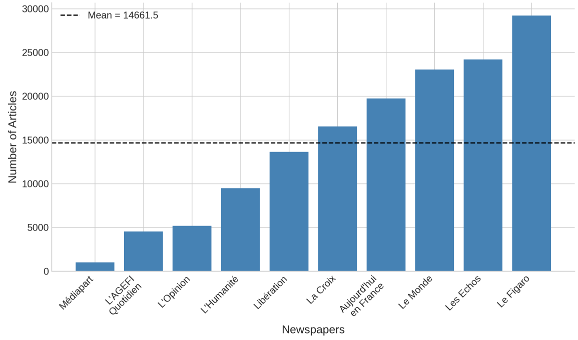

Figure 3: Number of articles per newspaper.

sentences, treating each cluster as a single document. Stopwords and terms appearing in more than 70% of all clusters are excluded to maintain keyword diversity. This threshold of 70% represents a compromise between representativeness and specificity. The three terms with the highest TF–IDF scores are selected as representative of each cluster's content.

4. Corpus Presentation

Our corpus is composed of articles collected from the Europresse portal[2]. It includes all articles published in 2024 from nine French newspapers belonging to the "national press" category of this platform: *Le Figaro*, *La Croix*, *Les Echos*, *L'Opinion*, *L'AGEFI Quotidien*, *L'Humanité*, *Libération*, *Le Monde*, and *Aujourd'hui en France*. We also included all 2024 articles from *Mediapart*, as left-leaning newspapers were underrepresented in the initial selection.

The dataset consists of 146,615 articles, whose distribution across newspapers is shown in Figure 3.

The corpus is not balanced in terms of the number of articles per theme (see 4.1) or per newspaper. However, this imbalance is inevitable when considering the entirety of the French press over a full year. We also note that the articles are evenly distributed throughout 2024, except for a drop in July and August (the summer holidays period), as shown in Figure 4.

To further improve the consistency of the corpus, we excluded *L'Opinion*, *L'AGEFI Quotidien*, and *Mediapart* from the subsequent analyses. These three newspapers contain the fewest articles overall, which limits their contribution to a balanced comparison across outlets. In addition, *L'AGEFI Quotidien* is highly specialized in economic news and exhibits a strong concentration of articles in the **Economy** category, with very limited coverage of other themes (see Figure 6). *L'Opinion* focuses

[2]https://www.europresse.com/

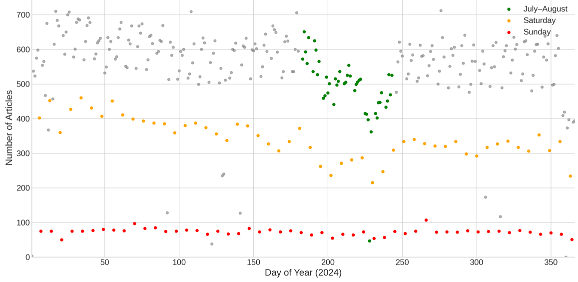

Figure 4: Number of articles per day (Sundays in red, Saturdays in orange, July and August in green).

Voc size	Before (7 newspapers)	After (all newspapers)
Culture	550	21
Economy	556	678
Education	134	1
Environment	540	372
Opinion	558	105
Politics	565	259
Society	559	41
Sport	567	5
Technology	490	113

Table 1: Vocabulary size before and after adding *L'Opinion*, *L'AGEFI Quotidien*, and *Mediapart*.

primarily on "political, economic, and international news"[3] and, given both its editorial scope and its relatively small total number of articles, provides only sparse coverage in several themes. *Mediapart*, as an online investigative outlet without a daily print edition, also differs structurally from the other newspapers in the corpus. These imbalances affect the total number of words retained after preprocessing and ultimately reduce the size of the final vocabulary (see Table 1).

4.1. Articles Themes

To ensure greater consistency in our lexical analyses, we automatically categorized the articles into nine thematic domains: **Culture**, **Economy**, **Education**, **Environment**, **Opinion**, **Politics**, **Society**, **Sport**, and **Technology**. This classification was performed using a fine-tuned French large language model (Scialom et al., 2020) based on the `FlauBERT` model[4]. Figure 5 shows the number of articles per topic, and Figure 6 illustrates the distribution of these topics across sources.

[3] [https://www.lopinion.fr/utiles/qui-sommes-nous

[4] https://huggingface.co/flaubert/flaubert_base_cased

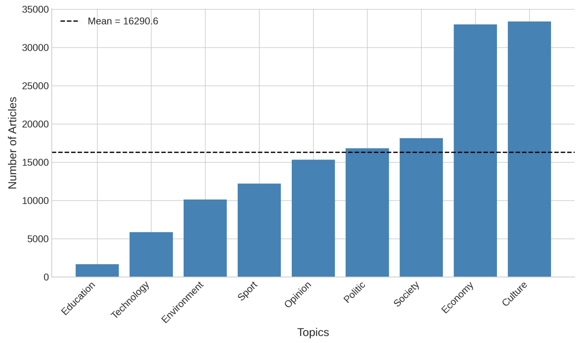

Figure 5: Number of articles per topic in the corpus.

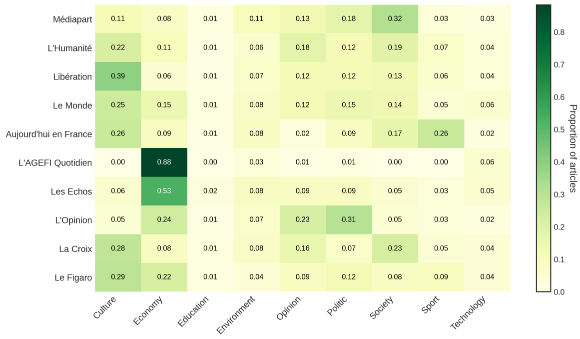

Figure 6: Proportion of articles in each topic for each newspaper.

4.2. Newspaper Political Orientation

To study how ideological differences influence vocabulary choices, we classified the newspapers into two political orientations: **right-leaning** (*Le Figaro*, *Les Echos*, *La Croix*, *L'Opinion*, *Aujourd'hui en France*) and **left-leaning** (*L'Humanité*, *Libération*, *Le Monde*).

This classification is based on the voting patterns of each newspaper's readership. We used the results of a 2022 Ifop study on voting behavior during the first round of the French presidential election[5], summarized in Table 2. To attribute political orientations to the candidates, we relied on the classification of their parties as defined by the *Conseil d'État*[6]. Although President Emmanuel Macron's party, *Renaissance*, is officially centrist, most of its voters self-identify as right-wing[7], and French citizens more broadly describe it as center-right[8]. For this reason, we chose to include *Renaissance* and Emmanuel Macron within the right-leaning category.

[5] Ifop: https://tinyurl.com/4xpr2zj7

[6] Légifrance: https://www.legifrance.gouv.fr/circulaire/id/45472?origin=list

[7] Ifop: https://tinyurl.com/434jhve3

[8] Ipsos: https://tinyurl.com/9wanx47m

Party (%)	Far-left	Left	Right	Far-right
L'Humanité	3	**72**	9	16
Libération	4	**56**	24	16
Le Monde	2	**46**	33	19
Aujourd'hui en France	0	31	**38**	31
La Croix	1	31	**41**	27
Le Figaro	2	20	**46**	32
Les Echos	3	35	**47**	15
All voters	*1.4*	*30.6*	*37.6*	*30.4*

Table 2: Percentage of votes for different political orientations by newspaper readership (abstention not included).

5. Respective Political Stance of the Newspapers

We measure differences in word usage across newspapers using the Jensen–Shannon divergence, which yields a score we refer to as *dissimilarity*. A dissimilarity value of 0 indicates that the corresponding distributions are identical. For each word, we compute pairwise dissimilarities between newspapers. Averaging these pairwise scores provides a general measure of the *distance* between newspapers.

5.1. Average Results

We first examine results without separating by topic, as shown in Figure 7. When averaging across all words, distances between newspapers are small, which is expected since most words exhibit low dissimilarity.

To evaluate the overall dispersion of newspapers, we compute the standard deviation of their coordinates in the distance space. The results are presented in Figure 8. The global dissimilarity increases by nearly a factor of eight when considering only the top 100 most divergent words, compared to using the entire vocabulary. We also observe that left-leaning newspapers appear closer to one another than right-leaning ones.

Then in the next sections, to prevent the numerous low-dissimilarity words from disproportionately lowering the overall averages and to ensure comparability across topics (the vocabulary size being different for each topic, as shown in Table 1), we compute the *distances* between newspapers using the 100 words with the highest dissimilarity values.

5.2. Topic by Topic Results

However, averaging across all topics obscures topic-specific variations. We therefore compute

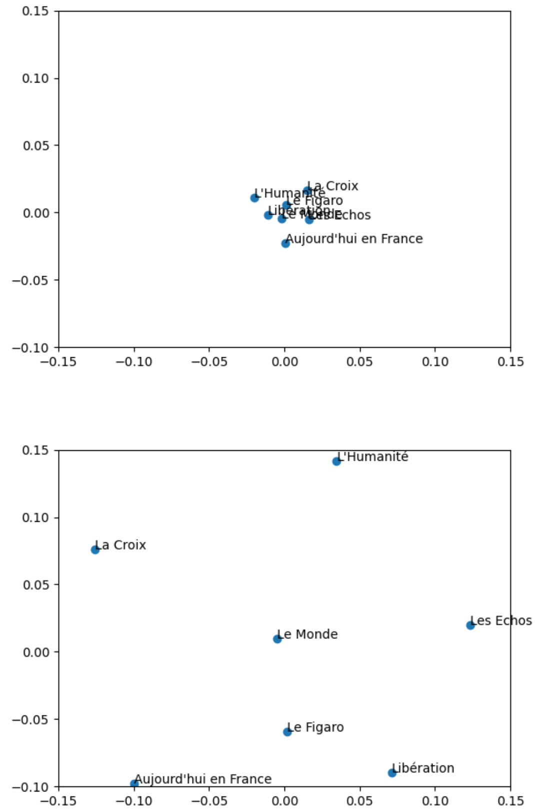

Figure 7: Distances between newspapers across all topics: (top) average over all words; (bottom) average over the 100 most dissimilar words.

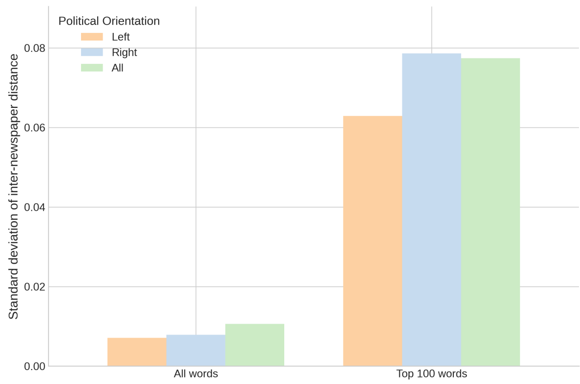

Figure 8: Standard deviation of newspaper positions using the two averaging methods.

distances between newspapers within each thematic category by averaging the 100 highest dissimilarities per theme. This method reduces the concentration of points seen in global averages while mitigating the influence of potential outliers. An example for the theme **Technology** is shown in Figure 9.

As a sanity check, we examine whether our results depend on the size of the final vocabulary

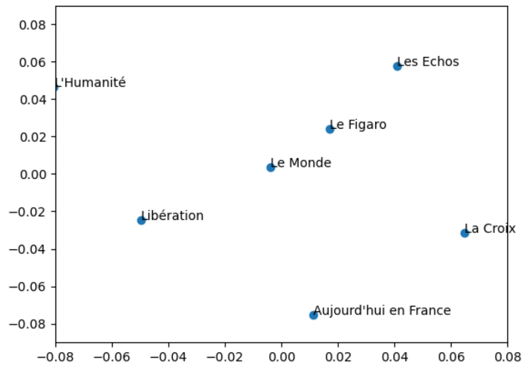

Figure 9: Distances between newspapers for the theme **Technology**.

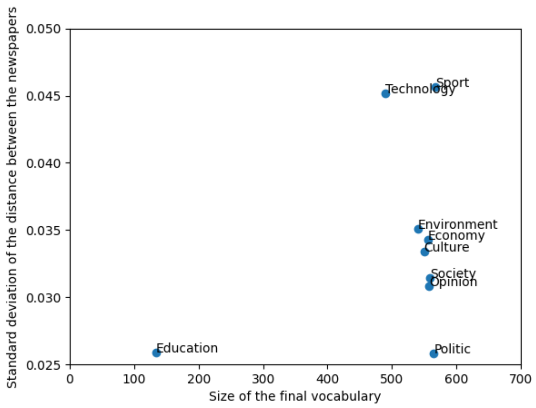

Figure 10: Dispersion of newspapers as a function of vocabulary size for each topic.

retained at the end of the analysis. Figure 10 compares the final dispersion of newspapers with the vocabulary size for each topic. No clear correlation is observed—topics with similar vocabulary sizes can exhibit very different dispersions.

We also compute standard deviations per topic while grouping newspapers by political orientation (Figure 11). For seven out of nine topics, the overall standard deviation (across all newspapers) is higher than the within-group deviations for left or right-leaning newspapers. This suggests that newspapers tend to be lexically closer to others sharing their political orientation than to those from the opposite side. However, two exceptions emerge: in **Culture**, right-leaning newspapers are considerably more dispersed than left-leaning ones, whereas the opposite holds true in **Economy**.

Finally, the thematic separation also allows us to visualize how newspapers shift relative to one another across topics. For instance, *Le Figaro* and *Les Echos* are close in the **Economy** theme but distant in **Sport**, as illustrated in Figure 12.

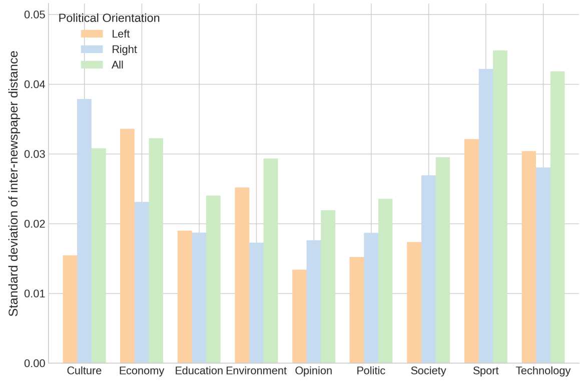

Figure 11: Standard deviation of newspapers by topic, separated by political orientation.

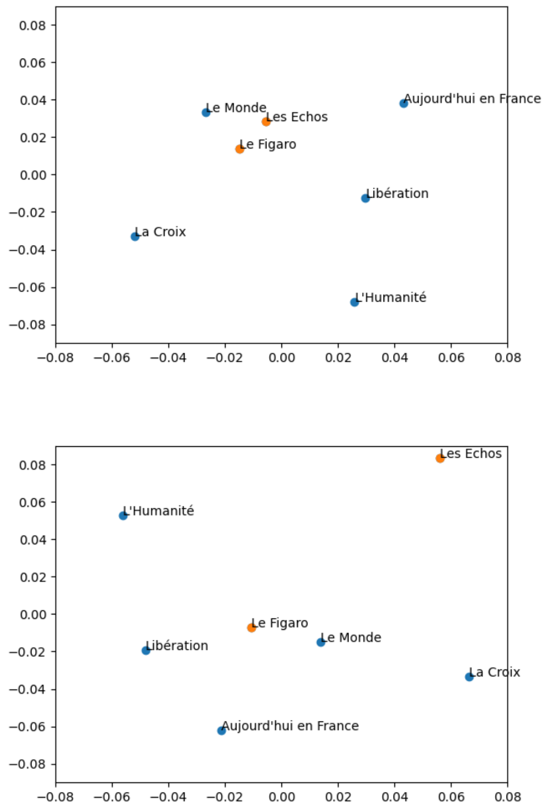

Figure 12: Relative positions of newspapers for the themes **Economy** (top) and **Sport** (bottom).

5.3. Addition of Other Newspapers

We also examine the results obtained when including *L'Opinion*, *L'AGEFI Quotidien*, and *Mediapart*. For this analysis, we focus exclusively on the theme **Economy**, as the final vocabulary is too limited for other themes.

We observe that *L'AGEFI Quotidien* and *Mediapart* appear relatively isolated from the other newspapers, which is consistent with their distinctive

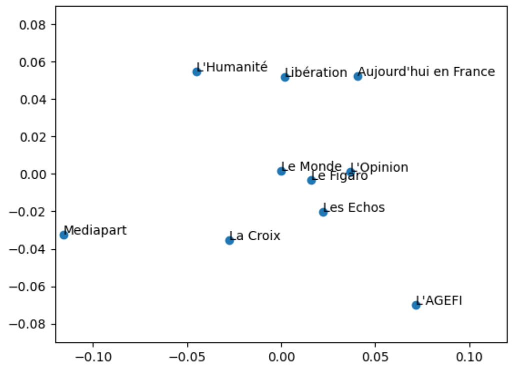

Figure 13: Relative positions of all newspapers for the theme **Economy**.

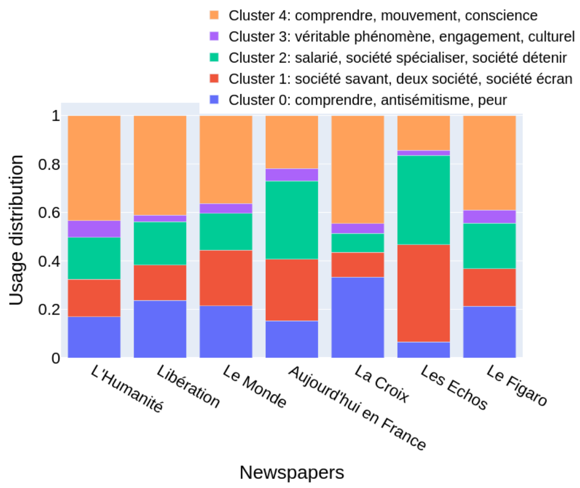

Figure 14: Distribution of clusters across newspapers for the term *société* (*society*).

editorial perspectives. Conversely, *L'Opinion* is positioned very close to *Le Figaro* on this theme.

These observations suggest that our method could be used to infer the political orientation or proximity of newspapers and to identify distinctive or marginal voices within the broader media landscape.

6. Overview of Highly Dissimilar Words

We now take a closer look at the words that exhibit the highest dissimilarity scores. Several distinct types of terms emerge.

First, we find polysemous words such as *état* ("state") or, more subtly, *sud* ("south"), which is also an acronym for *Solidaires Unitaires Démocrates*, a French labor union. We also identify very common terms that reflect each newspaper's editorial preferences, such as *page* (often associated with book reviews or cultural sections) or *euro* (whose usage depends on which items the newspaper chooses to report prices for).

Some terms capture major events from 2024, such as *covid* or *Jeux Olympiques* ("Olympic Games"). Other words reveal politically charged connotations, including *banlieue* ("suburb"), *hôpital* ("hospital"), and *privé* ("private").

These "types" can also overlap. An interesting example is *société* ("society"), whose cluster distribution is shown in Figure 14. Clusters 3 and 4 correspond to an understanding of "society" as a "social system" and are mainly associated with discussions of social movements. Cluster 1 corresponds to "society" in the sense of an association or organization, while Cluster 2 refers more specifically to the interpretation of "society" as a "company." This polysemy reflects the editorial orientations of the newspapers: *L'Humanité*, the newspaper of the French Communist Party, more often uses "société" in the sense of "social system," whereas *Les Échos*, a finance-focused outlet, tends to use it to mean "company". Finally, Cluster 0, characterized by representative terms such as "antisemitic" and "fear", appears to reveal yet another sense of "society", tied to the news of the time.

Overall, these results provide a snapshot of the political and social context of 2024 while also highlighting how certain words because of their polysemy or politically loaded meanings may be interpreted differently, or even misunderstood, by readers who rely on a single news source.

7. Conclusion and Perspectives

7.1. Conclusion

We introduced a framework for detecting semantic differences between newspapers in relation to their political orientation. Our approach combines topic segmentation, contextual embeddings, clustering, and divergence metrics to uncover variations that remain invisible to traditional frequency-based analyses.

Our findings indicate that left and right-leaning newspapers do not merely select different vocabularies; they activate distinct interpretations of the same words. This suggests that polarization is not a uniform linguistic phenomenon but rather the outcome of selective framing, agenda-setting, and the strategic use of polysemy. By operating at the topic level, our framework mitigates the smoothing effect of global averages and enables the identification of specific domains where semantic conflicts emerge.

One of the main limitations of our work is the difficulty of determining whether the observed differences stem from ideological variation or from differential topical coverage. The choice of topics to cover, as well as a newspaper's editorial decisions, also contributes significantly to media bias and may reflect underlying ideological positions. However, this constitutes a different type of bias, and it is important to distinguish between the two.

7.2. Perspectives

This study opens several promising avenues for further research.

First, the robustness of the framework could be strengthened by comparing alternative topic classifiers and thematic taxonomies, since topic segmentation directly influences the detection of semantic divergence. We could also experiment with an adaptive k selection in the k-means clustering. Similarly, assessing different similarity metrics (e.g., cosine, Jaccard), language models, and clustering strategies would help determine whether the observed ideological distances arise from methodological choices or reflect genuine linguistic variation.

Moreover, expanding the analysis to include verbs, adjectives, and multi-word expressions would provide a richer understanding of evaluative and ideological framing, as many biases are conveyed through actions, qualifiers, or recurring phrases.

Additionally, we would like to conduct a more qualitative analysis of the extracted words in order to better understand what explains their high dissimilarity scores.

Finally, applying this approach to other languages, media ecosystems (such as social networks, political speeches, or television), or historical periods would allow for cross-context comparisons: Are the same topics polarized across cultures? Are certain semantic tensions universal, or do they emerge from specific socio-political contexts?

Addressing these questions would contribute to a deeper understanding of how meaning, ideology, and discourse interact and evolve within democratic societies.

8. Bibliography

Chloé Clavel et al. Aina Garí Soler, Jenny Myrendal. 2025. The newme corpus: A gold standard corpus for the study of word meaning negotiation. In *PREPRINT (Version 1) available at Research Square*.

Lisa P. Argyle, Christopher A. Bail, Ethan C. Busby, Joshua R. Gubler, Thomas Howe, Christopher Rytting, Taylor Sorensen, and David Wingate. 2023. Leveraging ai for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences*, 120(41):e2311627120.

Hosein Azarbonyad, Mostafa Dehghani, Kaspar Beelen, Alexandra Arkut, Maarten Marx, and Jaap Kamps. 2017. Words are malleable: Computing semantic shifts in political and media discourse.

Yejin Bang, Delong Chen, Nayeon Lee, and Pascale Fung. 2024. Measuring political bias in large language models: What is said and how it is said. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11142–11159, Bangkok, Thailand. Association for Computational Linguistics.

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1472–1482, Denver, Colorado. Association for Computational Linguistics.

Ceren Budak, Sharad Goel, and Justin M Rao. 2016. Fair and balanced? quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1):250–271.

Dallas Card, Amber Boydstun, Justin H Gross, Philip Resnik, and Noah A Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444.

Béatrice Daille, Evelyne Jacquey, Gaël Lejeune, Luis Felipe Melo, and Yannick Toussaint. 2016. Ambiguity diagnosis for terms in digital humanities. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4353–4359, Portorož, Slovenia. European Language Resources Association (ELRA).

Dorottya Demszky, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky. 2019. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv preprint arXiv:1904.01596*.

Suyash Fulay, William Brannon, Shrestha Mohanty, Cassandra Overney, Elinor Poole-Dayan, Deb

Roy, and Jad Kabbara. 2024. On the relationship between truth and political bias in language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, page 9004–9018. Association for Computational Linguistics.

Aina Garí Soler, Matthieu Labeau, and Chloé Clavel. 2022. One word, two sides: Traces of stance in contextualized word representations. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3950–3959, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Matthew Gentzkow and Jesse M. Shapiro. 2010. What drives media slant? evidence from u.s. daily newspapers. *Econometrica*, 78(1):35–71.

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*, pages 3960–3973, Online. Association for Computational Linguistics.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2018. Diachronic word embeddings reveal statistical laws of semantic change.

Ashiqur R. KhudaBukhsh, Rupak Sarkar, Mark S. Kamlet, and Tom M. Mitchell. 2020. We don't speak the same language: Interpreting polarization through machine translation.

Vivek Kulkarni, Margaret L Kern, David Stillwell, Michal Kosinski, Sandra Matz, Lyle Ungar, Steven Skiena, and H Andrew Schwartz. 2018. Latent human traits in the language of social media: An open-vocabulary approach. *PloS one*, 13(11):e0201703.

Matej Martinc, Adélie Laruncet, Clara Bordier, Marceau Hernandez, Senja Pollak, and Gaël Lejeune. 2024. The Evolution of Bias in French News Media: How Does Political Orientation Affect Semantic Change? In *JADT*, Bruxelles, France.

Burt L. Monroe, Michael P. Colaresi, and Kevin M. Quinn. 2017. Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4):372–403.

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

Yujin Potter, Shiyang Lai, Junsol Kim, James Evans, and Dawn Song. 2024. Hidden persuaders: LLMs' political leaning and their influence on voters. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 4244–4275, Miami, Florida, USA. Association for Computational Linguistics.

Paul Rottger, Musashi Hinck, Valentin Hofmann, Kobi Hackenburg, Valentina Pyatkin, Faeze Brahman, and Dirk Hovy. 2025. Issuebench: Millions of realistic prompts for measuring issue bias in llm writing assistance.

Thomas Scialom, Paul-Alexis Dray, Sylvain Lamprier, Benjamin Piwowarski, and Jacopo Staiano. 2020. Mlsum: The multilingual summarization corpus.

Andrew B. Hall Sean J. Westwood, Justin Grimmer. 2025. Measuring perceived slant in large language models through user evaluations.

Michael Henry Tessler, Michiel A. Bakker, Daniel Jarrett, Hannah Sheahan, Martin J. Chadwick, Raphael Koster, Georgina Evans, Lucy Campbell-Gillingham, Tantum Collins, David C. Parkes, Matthew Botvinick, and Christopher Summerfield. 2024. Ai can help humans find common ground in democratic deliberation. *Science*, 386(6719):eadq2852.

Marlena Wisniak. 2025. Algorithmic gatekeepers: Impacts of llm content moderation on civic space and human rights. Technical report, European Center for Not-for-Profit Law (ECNL).

Michael Yeomans, Alejandro Kantor, and Dustin Tingley. 2018. The politeness Package: Detecting Politeness in Natural Language. *The R Journal*, 10(2):489–502.