ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

# Data-Driven Mining of Discourse Patterns in French News and Quantitative Analysis of Their Audience Influence

*Yan Sun*[1,*]

[1] Xi'an International University, Xi'an, Shaanxi, 710077, China

**SUMMARY:** *This article takes French news from 2023 to 2025 as the object, constructs a cross media news corpus and audience interaction database, uses French Transformer representation, topic clustering, supervised classification, and econometric modeling methods to identify the main discourse patterns in news texts, and examines their differential impact on audience participation at different levels. Research has found that French news mainly forms five stable modes: institutional/governance type, conflict/opposition type, testimony/character story type, data/expert interpretation type, and crisis/risk warning type. Among them, institutional/governance type accounts for the highest proportion, conflict/opposition type is more likely to bring clicks, browsing, and sharing, testimony/character story type has the strongest promoting effect on comments and long discussion chains, and data/expert explanation type, although not achieving the highest superficial popularity, can steadily enhance deep participation. Further analysis shows that the performance of different modes is not consistent between mainstream media and local media, serious news and soft news, official websites and social platforms, and crisis and normal periods. Starting from the context of French news, this article combines discourse analysis, computational text research, and audience quantification evaluation to provide empirical evidence for understanding news expression adjustment and audience relationship reconstruction in a platform based communication environment.*

***KEYWORDS:*** *French news media; discourse pattern mining; audience engagement; news framing; computational journalism*

## 1 Introduction

In the reality of news dissemination in France, news production is no longer limited to a linear process of "writing publishing", but rather forms a continuous linkage between platform distribution, audience feedback, and editorial decision-making. Faced with the complex communication environment composed of social media platforms, mobile homepages, push notifications, and comment areas, news organizations not only need to report on the events themselves, but also constantly adjust the way titles are condensed, the information density of introductions, the order of quotes, and the pace of narrative perspective switching to enhance the visibility, readability, and interactivity of content. Research on local media in French speaking regions of Europe shows that many news organizations have actively introduced audience engagement mechanisms, incorporating comments, feedback, collaborative topic selection, and even community interaction into the content production process. This means that the organization of news texts is being continuously guided by audience oriented logic.

*yansun0111@163.com
https://doi.org/10.65102/is2026732 Published: 30 April, 2026

Furthermore, this change is not a scattered editorial fine-tuning, but a strategic shift at the organizational level: the relationship between French media and audiences is gradually shifting from traditional one-way transmission to a dynamic structure that includes connection, negotiation, and adaptation. Audiences are not only information receivers, but also indirectly shaping the form of news expression. One direct consequence of this change is that the discourse arrangement within news texts has become more functional. Whether the title highlights conflicts, whether the main text prioritizes expert quotes, and whether the narrative revolves around institutional explanations or character experiences all affect the dissemination performance of news on different platforms. Related studies have pointed out that news content faces significant "promotion pressure", and whether a report can gain cross platform visibility and subsequent interaction is often closely related to its content packaging and presentation, rather than just the importance of the event itself. At the same time, research on data news and audience engagement also suggests that news organizations are increasingly relying on quantitative indicators to determine whether content is "effective," making the correspondence between text features and audience reactions more worthy of analysis than ever before. From this, it can be seen that the language style, structural selection, and narrative organization in French news are not neutral textual surfaces, but a set of observable, comparable, and measurable discourse practices gradually formed in a platform based communication environment.

However, although existing research has extensively discussed audience engagement, social media editing strategies, and news data-driven production, there are still two significant shortcomings. Firstly, many studies focus on English news, social platform operations, or general interaction mechanisms, with insufficient attention paid to the internal structure of French news texts, particularly the lack of systematic research that integrates titles, quotes, emotional expressions, source configuration, and narrative frameworks into the same analytical framework. Secondly, many studies still focus on coarse-grained indicators such as total likes and comments when discussing "interaction", and there is insufficient recognition of differences in audience behavior at different levels. In fact, audience data practice itself is not just a technical tool, it will in turn shape news organizations' judgments on what is worth writing and how it should be written, and may exacerbate the imbalance of different types of content in dissemination. Even on specific platforms, different interactive spaces can trigger different ways of participation, such as message and group environments with strong privacy, which are not consistent in terms of news discussion intensity and interactive nature [6]. Therefore, discussing the audience influence of French news solely based on macro platform performance is not sufficient to explain how the discourse organization within news texts affects audience behavior. Furthermore, there is a conceptual ambiguity in the current understanding of audience metrics. Superficially similar interactive data may not necessarily have the same meaning in editing practice and user experience. Whether 'Like' represents agreement, attention, polite response, or light exposure is not a single and stable signal. At the same time, participatory journalism practices have begun to change the way political news and other types of content are written. News is no longer just about events and facts, but also emphasizes responding to audience concerns, presenting life experiences, and raising the threshold for entry through lower forms of expression. This indicates that audience influence does not occur as an "afterthought" outside of the text, but may have been predetermined and absorbed during the selection, writing, and structural design stages. For French news, it is difficult to truly demonstrate how "platform based communication has changed news expression" without breaking down the news text into several computable discourse features and further examining their relationship with audience indicators at different levels.

Based on this, this article will clearly define the research objectives as two levels. Firstly, identify the main discourse patterns with stability in French news reporting. Secondly, quantify the differential impact of these patterns on different types of audience indicators. The "discourse pattern" here is not equivalent to a simple topic classification, but is composed of multiple dimensions such as title packaging, emotional intensity, quote structure, number of sources, narrative focus, and framework orientation. Previous studies have found that an increase in the number of sources in news can have a significant impact on audience response, indicating that the organizational characteristics within the text can be used as quantitative explanatory variables in the model [9]. Therefore, this article does not view news content as a complete black box, but attempts to break it down into identifiable, clustered, and comparable expression units, in order to reveal in more detail which narrative styles are more likely to attract clicks and which discourse configurations are more likely to trigger comments, sharing, or ongoing discussions.

In terms of research positioning, this article aims to make three advances in the study of news audience relations in the digital age. Firstly, this article focuses on French news corpus, addressing the issue of excessive emphasis on English samples in current research, and testing specific expression patterns in the French news field through empirical data. Secondly, this article operationalizes the "discourse pattern" into a set of computable variables, rather than staying at the qualitative impression level, and combines news discourse analysis with text mining, cluster recognition, and audience behavior modeling in terms of methodology. Again, this article divides "audience influence" into different levels, not only focusing on overall interaction volume, but also distinguishing shallow attention, deep discussion, and diffuse participation to avoid compressing heterogeneous audience responses into a single indicator. As emphasized in recent reviews of news audience relations in the digital age, the connection between news organizations and audiences has taken on more complex forms of interaction, including contact and feedback, as well as normative expectations, participation boundaries, and power redistribution. The core significance of this article is to provide quantifiable and interpretable textual evidence for this complex relationship in the specific context of French news, and to provide a more engineering and empirical analysis path for subsequent research on news dissemination mechanisms.

# 2 Methods

## 2.1 Corpus Construction and Audience Data Collection

This study sets the time window of the corpus as January 1, 2023 to December 31, 2025, and constructs a multi-source news corpus based on French news in France. Considering the need for research to cover both national mainstream media, local media, and digital oriented media, a preliminary sample of 8 French news organizations was selected, including Le Monde, Le Figaro, Lib é ration, Franceinfo, 20 Minutes, Ouest France, La D é p ê che du Midi, and Slate.fr. These media have significant differences in organizational attributes, reporting styles, topic distribution, and level of platformization, which helps to improve the horizontal comparability of the sample and avoid excessive influence of single media writing norms on research results. The data capture design of news organizations is shown in Table 1.

*Table 1: Data Capture Design for News Agencies*

| Media name | Media type | Main off-site platform entrance | Plan to include sample size (articles) |
|---|---|---|---|
| Le Monde | National mainstream media | Facebook and X | 12,000–18,000 |
| Le Figaro | National mainstream media | Facebook and X | 12,000–18,000 |
| Libération | National mainstream media | Facebook and X | 8,000–14,000 |
| Franceinfo | Public Service News Media | Facebook and X | 10,000–16,000 |
| 20 Minutes | Urban/Digital oriented Media | Facebook, X and Instagram | 10,000–15,000 |
| Ouest-France | Local media | Facebook | 8,000–14,000 |
| La Dépêche du Midi | Local media | Facebook | 8,000–12,000 |
| Slate.fr | Digital native media | Facebook and X | 6,000–10,000 |

In Table 1, at the implementation level of the research, the corpus size is controlled between 80000 and 150000 articles, with an actual target size of about 100000 articles to ensure sufficient statistical stability for subsequent discourse pattern clustering, supervised classification, and audience influence modeling. Each article retains core fields such as title, introduction, body, publication time, column, author, visibility of comment area, inclusion of direct quotes, media type, and article link [11]. The data source adopts a "three-layer parallel" collection structure. The first layer is Media Cloud Online News Archive, which provides publicly searchable news story level metadata with fields covering titles URL、 The publication date, language, media name, etc., and the source directory clearly indicate that the news source is mainly continuously accessed through RSS and Google News Sitemap, making it suitable as an upstream entry for candidate article discovery, source verification, and deduplication comparison. However, although its story text can be retrieved, it does not support batch downloading. Therefore, this article does not use it as the sole platform for the full-text corpus, but rather as an intermediate layer for "discovery verification supplement chain". The second layer is GDELT open data, which publicly provides downloadable news metadata and derived fields such as topic and tone, which can be used to supplement cross media timestamps, domain names, topic tags, and event clues. The third layer is the public archive and subscription entrance of the official media website, for example, Le Monde provides archive pages for browsing by date, and explains that column pages can form RSS entrances by adding rss_full.xml. La D é p ê che du Midi also provides a public archives page organized by year. These three types of sources work together to complete corpus discovery, link tracing, and field completion without relying on closed platform interfaces.

The crawling rules follow the order of "index first, then return to source, and then clean". Firstly, extract candidate URLs from Media Cloud and GDELT based on the media domain name, publication time, and French language tag. Subsequently, return to the official website of the media to retrieve the title, introduction, main text, and page structure information, and standardize the canonical URL by removing UTM parameters, mobile jump suffixes, and social tracking parameters [12]. For situations where the same news appears repeatedly on multiple pages, AMP pages, or reposted pages within the website, merge them using "standardized links+title similarity+publishing time proximity". To avoid mistakenly writing page navigation, recommendation modules, or copyright notices into the main text, French text cleaning must also be performed after extracting the main text. This includes removing HTML tags, scripts, and ad blocks, unifying French apostrophe and quotation mark encoding, retaining French book titles, dashes, and citation marks, clearing contiguous blanks, repeated paragraphs, and short message fragments of less than 250 words. If the article lacks a stable main text, contains only videos, image collections, or live streaming scrollbars, it will not be included in the formal sample.

In terms of sample selection, this article does not pursue a "full site" approach, but rather focuses on two highly discussed topics. One is immigration and social governance, and the other is international conflicts. The former corresponds to French search terms such as immigration, migrant, asile, sans papiers, impulse, fronti è re, policy, justice, é meute, and declaration. The latter corresponds to keywords such as guerre, conflict, Ukraine, Gaza, Proche Orient, cessez le feu, arm é e, frappe, etc. When filtering, first recall keywords, and then manually review them based on the column, title, and context of the main text to eliminate noisy texts with the same form but different meanings in finance, sports, and entertainment [13]. At the same time, pure comments, editorials, reader contributions, photo collections, real-time live streaming pages, and extremely short news alerts with only a few hundred words are generally excluded. If the research requires retaining commentary text, it should be separately labeled as an opinion class and not mixed with straight news for modeling.

The collection of audience data is synchronized with the main text corpus. Considering that French news websites rarely disclose stable page views, this article takes the number of comments, shares, social media reactions, comment depth, and secondary dissemination rate as the main audience indicators, and lists page views as supplementary variables when available. At the internal level, record whether the comment section is open, the total number of comments, and the thread level. At the off-site level, prioritize matching public posts on Facebook or X (formerly Twitter) with news official website links, and capture visible reactions, comments, and shares/reposts. If the same article is repeatedly published on multiple official or secondary accounts, the first post shall prevail, and the number of secondary distributions shall be recorded to calculate the secondary dissemination rate. The purpose of this approach is not to compress all interactions into a single heat value, but to measure the "seen", "discussed", and "forwarded" separately, in order to provide finer grained dependent variables for modeling the differences between subsequent discourse patterns and audience responses.

## 2.2 Discourse Pattern Mining and Feature Engineering

To avoid simplifying the "discourse pattern" into a single topic label, this article divides French news texts into four types of features: content layer, expression layer, narrative layer, and structural layer, and integrates them into a unified vector space for modeling. The flowchart of discourse pattern mining and feature engineering is shown in Figure 1.

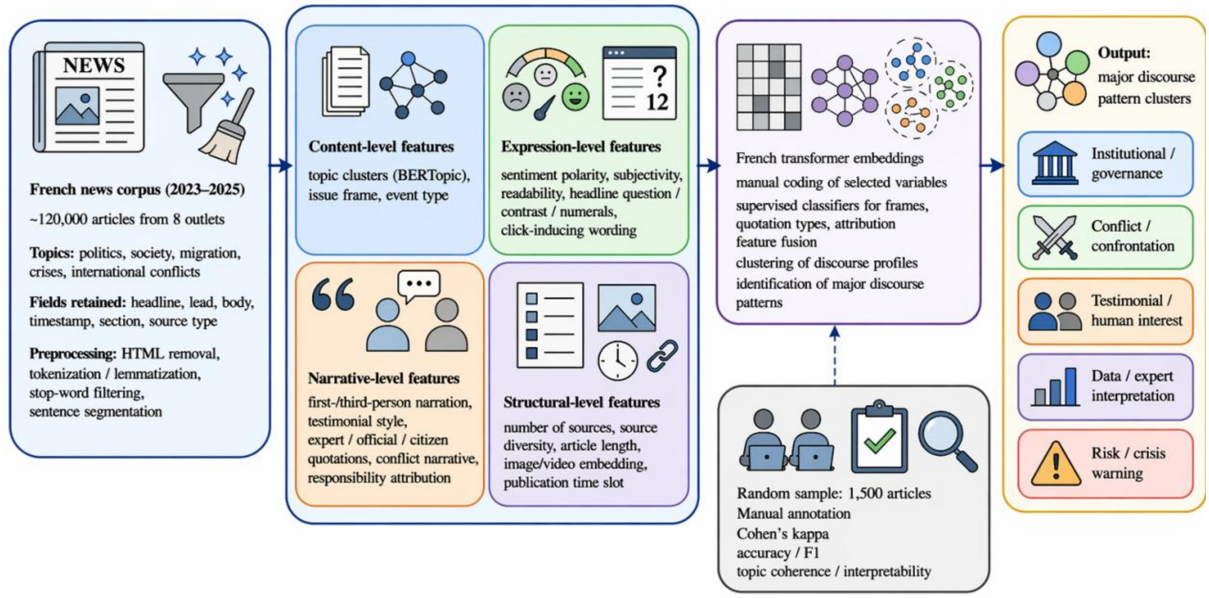



*Figure 1: Flow Chart of Discourse Pattern Mining and Feature Engineering.*

In Figure 1, the research process does not directly perform coarse-grained clustering on the entire text. Instead, a French pre trained transformer is used to vectorize the title, introduction, and main text. Then, BERTopic and Top2Vec are used for initial topic clustering to identify high-frequency topic distributions and potential topic boundaries in the news corpus [14]. On this basis, further introduction of manually defined multidimensional features enables clustering results to not only reflect 'what was written', but also characterize 'how it was written'. Specifically, the content layer mainly includes themes, news frameworks, and event types. Among them, the topic is generated by unsupervised clustering, and the framework and event types are identified by training a supervised classifier through small sample manual annotation [15]. The expression layer mainly characterizes the text style, including emotional polarity, subjectivity, readability, interrogative structure in the title, oppositional structure, digital expression, and click induction degree. Emotions and subjectivity are extracted jointly by a French emotion dictionary and a pre trained classification model. To quantify the emotional intensity of an article, the emotional score of the third article can be calculated as shown in formula (1).

$$s_i = \frac{1}{N_i}\sum_{j=1}^{N_i} \omega(w_{ij}) \tag{1}$$

In formula (1), $s_i$ represents the emotional score of the i-th article. $N_i$ represents the number of terms involved in emotion calculation in the i-th article. $w_{ij}$represents the j-th term in the i-th article. $\omega(\cdot)$ represents the emotional weight corresponding to the term. The narrative layer focuses on the way of speaking and the allocation of responsibility in the report, including first person or third person narration, testimonial narration, expert quotes, official quotes, ordinary public quotes, conflict narratives, and attribution of responsibility. The structural layer records variables such as the number of information sources, source diversity, article length, embedding of images or videos, and publishing time period. In the structural layer variables, source diversity is quantified through information source distribution entropy, as shown in formula (2).

$$H_i = -\sum_{m=1}^{M_i} p_{im} \times \ln(p_{im}) \tag{2}$$

In formula (2), $H_i$ represents the information source diversity index of the i-th article. $M_i$ represents the total number of information source categories in the article. $p_{im}$ represents the proportion of the m-th type of information source. In the French news text vectorization stage, this article first uses a French pre trained Transformer model to represent the context of the title, introduction, and body text, and aggregates the token level representations into article level vectors. The semantic representation definition of article i is shown in formula (3).

$$z_i = \frac{1}{T_i}\sum_{t=1}^{T_i} h_{it} \tag{3}$$

In formula (3), $\boldsymbol{z}_i$ represents the document vector of the i-th article. $T_i$ represents the number of tokens in this article. $h_{it}$ represents the contextual latent vector of the t-th token in the i-th article. This semantic representation is used for subsequent topic clustering and multidimensional feature fusion. After the initial clustering is completed, this article further calculates the weights of representative words within the topic cluster to improve the interpretability of topic boundaries. For the weight of term v in topic cluster k, it is represented by class TF-IDF, as shown in formula (4).

$$\mathrm{TFIDF}_{\mathrm{k,v}} = \frac{\mathrm{f_{k,v}}}{\sum_{\mathrm{u=1}}^{\mathrm{V}} \mathrm{f_{k,u}}} \log \frac{K}{d_v} \tag{4}$$

In formula (4), $TFIDF_{k,v}$ represents the weight of term v in topic cluster k, which is used to extract high probability words and assist in determining the interpretability of the topic cluster. $f_{k,v}$ represents the word frequency of term v in topic cluster $k$. $V$ represents the size of the vocabulary. $K$ represents the total number of topic clusters. $d_v$ represents the number of topic clusters containing the word term $v$. Subsequently, 1500 news articles were randomly selected from the total sample for manual annotation. The framework, citation type, and attribution of responsibility were each established with annotation standards, which were independently completed by two trained coders [16]. After completing consistency verification, use this annotation set to train a supervised classifier. For the task of identifying frameworks, speech types, and attribution of responsibility, the objective function of the classifier is shown in formula (5).

$$\mathcal{L}_{cls} = -\sum_{\mathrm{i=1}}^{\mathrm{N}} \sum_{\mathrm{c=1}}^{\mathrm{C}} \mathrm{y_{ic}} \log \widehat{\mathrm{p_{ic}}} \tag{5}$$

In formula (5), $\mathcal{L}_{cls}$ represents classification loss. $N$ represents the number of training samples. $C$ represents the number of categories. $y_{ic}$ represents the true label of sample i on category $c$. $\widehat{p_{ic}}$ represents the probability of the model predicting that the sample belongs to category c. To verify the reliability of manual annotation, this article will report on Cohen's kappa. When label items with a kappa value below 0.75 appear, revise the rules again and re label them. After completing the consistency check, train the supervised classifier using the annotation set and evaluate the model performance based on accuracy, macro average F1, and weighted F1. In addition to the classification results, this article also conducts interpretability checks on topic clusters by extracting high weight words, representative articles, and title samples from each cluster, checking their internal semantic consistency and readability in journalism, and avoiding cluster structures that are only similar in mathematical space but have vague propagation meanings [17].

In the end, this article does not directly consider a single feature as a "discourse pattern", but combines four layers of features and performs secondary clustering to obtain several stable discourse pattern clusters. For example, reports on immigration issues may be categorized into different modes based on whether they use conflicting headlines, highlight testimonies from ordinary people, and introduce multiple sources of interpretation. The purpose of this approach is to move news texts from "topic classification" to "expression structure recognition", providing a more detailed feature basis for subsequent testing of the differential impact of different discourse patterns on comments, sharing, and secondary dissemination rates. In other words, the process of discourse pattern mining and feature engineering emphasizes not only the technical sequence, but also a methodological judgment. Only by putting content, expression, narrative, and structure in the same framework can the audience influence of news discourse be more reliably quantified.

## 2.3 Quantitative Modeling of Audience Influence

In the stage of audience influence modeling, this article divides response variables into two groups: shallow participation and deep participation. The former corresponds to clicks, browsing, likes, or reactions, mainly reflecting attention and light contact. The latter corresponds to the depth of comments, sharing, comment chains, and cross platform re dissemination, which is closer to the extension of discussion and public dissemination. This hierarchical processing is consistent with the expectations of digital news interaction, social media editing practices, and mobile contact logic, and is also more suitable for explaining the differences between news being "seen" and "discussed" on different platforms [18]. Meanwhile, testimonial storytelling, platform packaging, and storytelling often have additional impacts on comments and dissemination, so deep engagement cannot be replaced by shallow indicators [19]. To clearly demonstrate the correspondence between discourse pattern characteristics, audience indicators, control variables, and econometric models, and to illustrate the overall modeling path of shallow participation, deep participation, and robustness testing, this paper constructs a quantitative modeling mechanism diagram of audience influence, as shown in Figure 2.

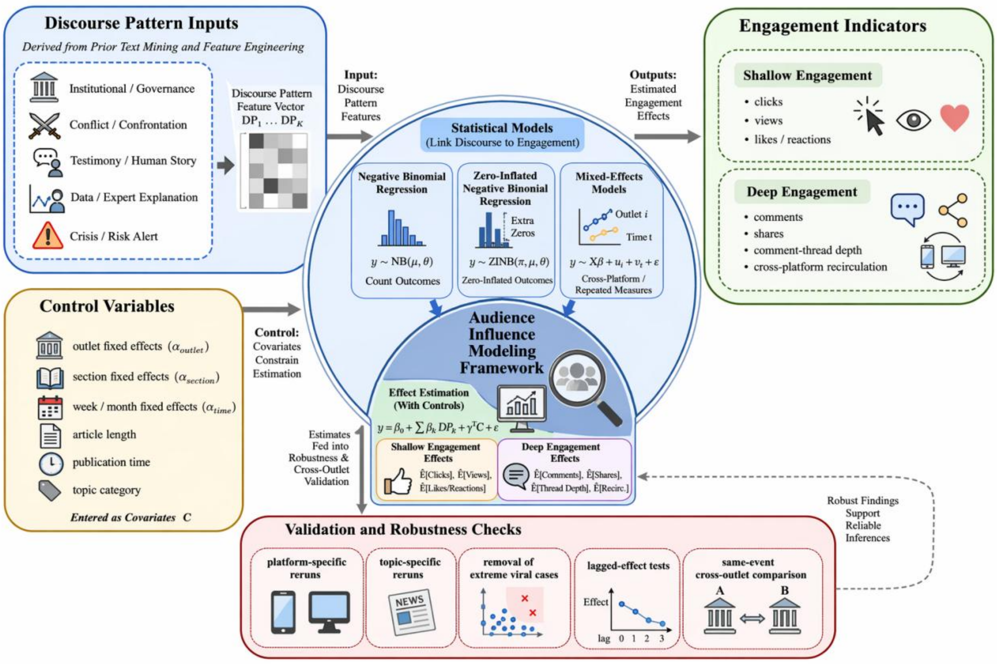



*Figure 2: Analytical framework for quantitative modeling of audience influence in French news discourse.*

As shown in Figure 2, the discourse pattern cluster and its four layer features are first introduced into the modeling framework as core explanatory variables, and then mapped to two types of audience indicators: shallow participation and deep participation. Factors such as media, program, time, length, and topic category are included as control items in the estimation process. Different types of indicators are then matched with negative binomial regression, zero inflation negative binomial regression, and mixed effects models [20]. The robustness test is ultimately completed through platform segmentation, topic segmentation, extreme value elimination, and comparison with the same event. For count based indicators such as comments, shares, reactions, and cross platform re dissemination times, this article uses negative binomial regression. If the proportion of zero values is too high, use zero inflation negative binomial regression instead. For repeated observations across platforms, multimedia, and time, a mixed effects model is further used to control for the correlation of internal observations within the same media, platform, and time unit, as shown in formula (6).

$$log(\mu_i) = \beta_0 + d_i^{T\beta} + c_i^{T\gamma} + u_j + v_p + w_t \tag{6}$$

In formula (6), $\beta_0$ represents a constant term, and $\beta$ represents the coefficient vector corresponding to the discourse pattern variable. $\gamma$ represents the coefficient vector corresponding to the control variable. $\mu_i$ represents the conditional mean of the i-th observed sample. $d_i$ represents the feature vector of the $i$-th sample. $c_i$ represents the control variable vector of the i-th sample. T represents transposition. $u_j$ represents the random intercept of the media j to which the $i$-th observation sample belongs. $v_p$ represents the random intercept of platform $p$. $w_t$ represents the random effect of the time unit t to which it belongs. The core independent variable is the discourse pattern cluster, and four levels of feature scores are included as auxiliary explanatory items. The control variables include media fixed effects, column fixed effects, weekly or monthly fixed effects, article length, publication time period, and topic category. This setting can more finely depict the marginal impact of title packaging, narrative perspective, and structural configuration on different audience indicators [21]. The robustness test includes four items: re running the model by platform, re running the model by topic, removing extreme hot news with interaction values in the top 1%, and adding lagged items to test the stability of the results. In addition, this article adds a "comparison of the same event" design. Identify cross media reports of the same public event based on named entities, keyword overlap, and proximity of release dates, control the event window within 3 days before and after, and compare the interaction differences under different discourse patterns. This approach helps to compress the external disturbances caused by the significance of events, making the model closer to identifying the impact of expression rather than just describing the overall heat differences [22, 23]. Considering the language boundary between opinion and straight news in French news, as well as the extension conditions of framework analysis in multilingual scenarios, this paper synchronously refers to the method constraints of French news genre recognition and multilingual framework analysis in the event comparison and model interpretation stages to ensure consistency between estimation results and text classification results [24, 25].

# 3 Results and Discussion

## 3.1 Distribution of Major Discourse Patterns in French News

After completing theme clustering, framework recognition, and multidimensional feature integration, the French news corpus presents five relatively stable discourse patterns, namely institutional/governance type, conflict/opposition type, testimony/character story type, data/expert interpretation type, and crisis/risk warning type. The cross channel and cross topic distribution of the main discourse patterns in French news is shown in Figure 3.

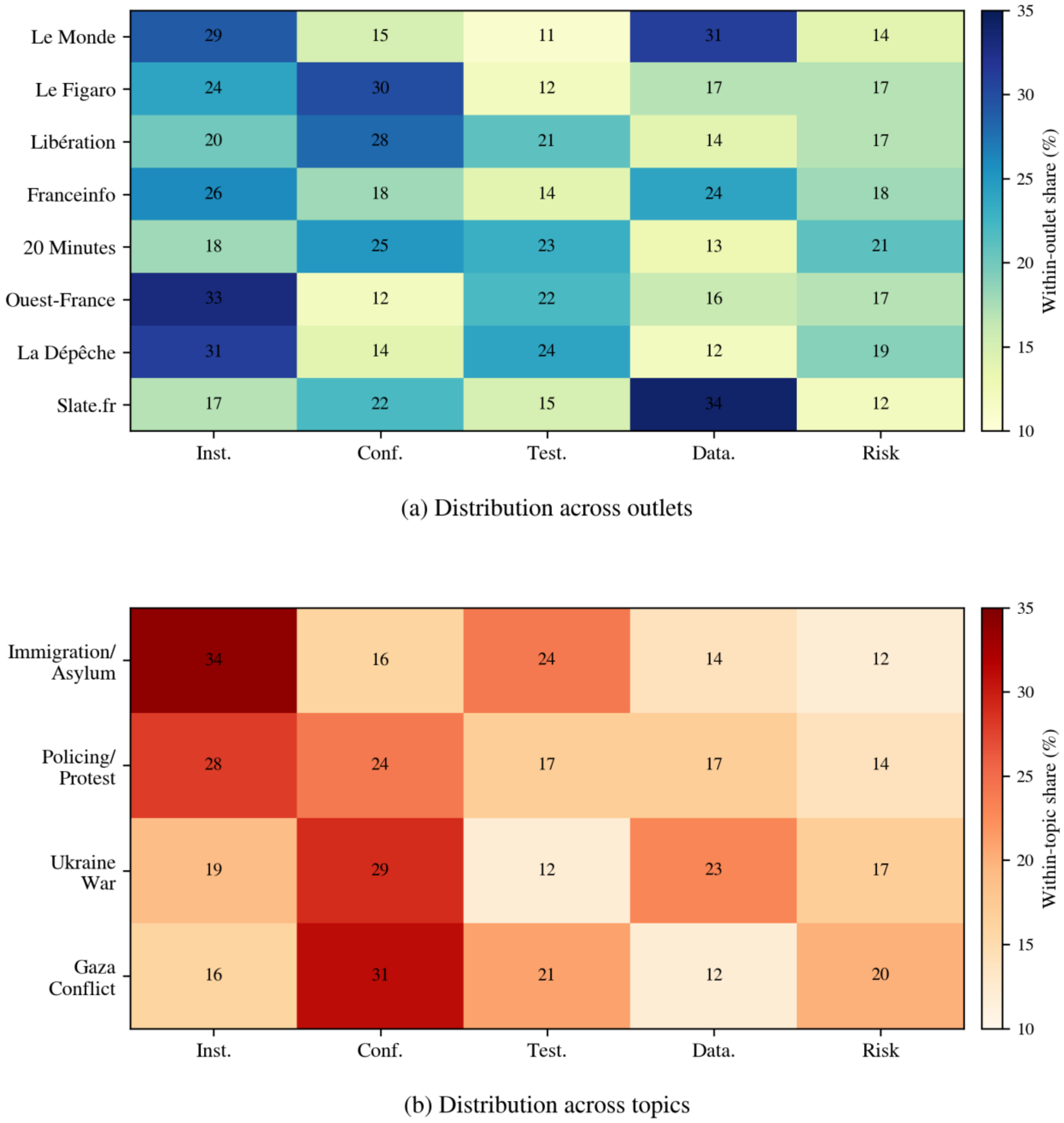



*Figure 3: Cross-outlet and cross-topic distribution of major discourse patterns in French news.*

As shown in Figure 3 (a), Le Monde, Franceinfo, and Slate.fr have a higher proportion of institutional/governance and data/expert interpretation types, with Le Monde's institutional/ governance type accounting for 29% and data/expert interpretation type accounting for 31%, while Slate.fr's data/expert interpretation type accounts for 34%. Relatively speaking, the conflict/opposition types of Le Figaro and Lib é ration are 30% and 28%, respectively, significantly higher than the sample mean of 23.1%. The institutional/governance type of local media Ouest France and La D é p ê che reached 33% and 31% respectively, while the testimony/character story type was 22% and 24% respectively, indicating that local reporting often juxtaposes policy treatment with individual experiences. In Figure 3 (b), the institutional/governance type of immigration and asylum issues has the highest proportion at 34%, higher than the 16% of Gaza conflict issues. The conflict/opposition type in the Gaza conflict has risen to 31%, and the crisis/risk warning type has also reached 20%. The report on the Ukrainian war presents a bimodal structure, with conflict/opposition type accounting for 29% and data/expert interpretation type accounting for 23%. These differences indicate that there is no single narrative focus within the same language news field, and media positioning and topic attributes can rewrite the way discourse is organized. To further illustrate that these patterns are not loosely labeled, this article summarizes the overall proportion, direct speech rate, dominant speech source, and stability indicators of each pattern, as shown in Table 2.

*Table 2: Overall prevalence, quotation structure, and stability indicators of the five discourse patterns*

| Pattern | Proportion (%) | Direct speech rate (%) | Dominant source of speech | Quarterly standard deviation | Bootstrap Jaccard |
|---|---|---|---|---|---|
| Institutional/governance oriented | 27.4 | 38.2 | Official/Institutional | 1.1 | 0.81 |
| Conflict/opposition type | 23.1 | 44.7 | Bilateral actors | 1.4 | 0.78 |
| Testimony/Character Story Type | 18.7 | 62.5 | Ordinary people | 0.9 | 0.80 |
| Data/expert explanatory type | 16.2 | 35.9 | Expert/Researcher | 0.8 | 0.77 |
| Crisis/risk warning type | 14.6 | 29.4 | Warning agencies/authoritative departments | 0.7 | 0.79 |

In Table 2, institutional/governance type accounts for 27.4% of the entire sample, with a direct citation rate of 38.2%, mainly from official and institutional sources. High probability words are concentrated in government, law, reform, procedure, and jurisdiction, and the representative title can be summarized as "Immigration: the government tightening asymptotes procedures". Conflict/opposition type accounts for 23.1%, with a direct speech rate of 44.7%. Common bilateral speech and attribution expressions include high probability words such as conflict, tension, accommodation, protein, and policy. Testimony/character story type accounts for 18.7%, but the highest rate of direct speech is 62.5%, dominated by ordinary people's speech, with typical words being family, victim, resident, recovered, and experience. Data/expert explanatory type accounts for 16.2%, with expert quotes being the most concentrated; Crisis/risk warning type accounts for 14.6%, characterized by warning type wording and timely sources. The verification results are also relatively stable, with quarterly standard deviations of the five types of patterns ranging from 0.7 to 1.4, and Bootstrap Jaccard coefficients ranging from 0.77 to 0.81, indicating that their boundaries are reproducible and not temporarily pulled by individual hot topic reports.

Overall, the discourse distribution of French news in France has a clear hierarchy: institutional/governance type forms the foundation, conflict/opposition type and testimony/character story type bear the increase of topics, and data/expert interpretation type and crisis/risk warning type form complementary advantages in specific media and specific topics. From this, it can be seen that the subsequent analysis of audience influence cannot regard "news content" as a single entity, but should compare the differential effects of different discourse patterns at the interactive level.

### 3.2 Differential Effects on Shallow and Deep Audience Engagement

After distinguishing audience response into shallow participation and deep participation, there is a clear differentiation in the direction of different discourse patterns. To compare the direction and intensity of the effects of different discourse patterns on shallow and deep engagement, this paper reports the estimated incidence ratio of each pattern on two types of audience indicators, as shown in Figure 4.

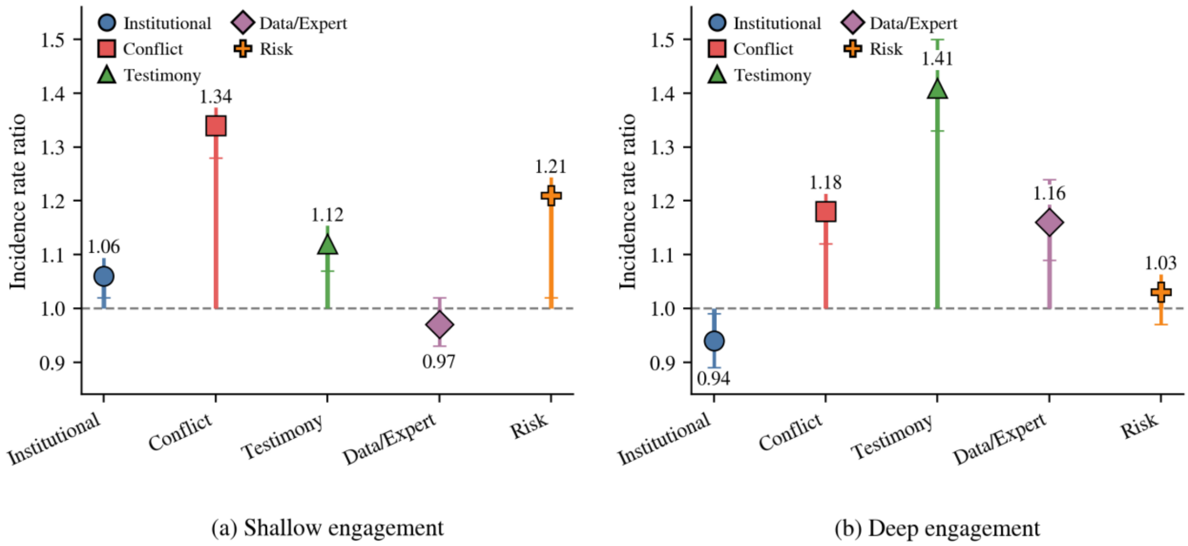
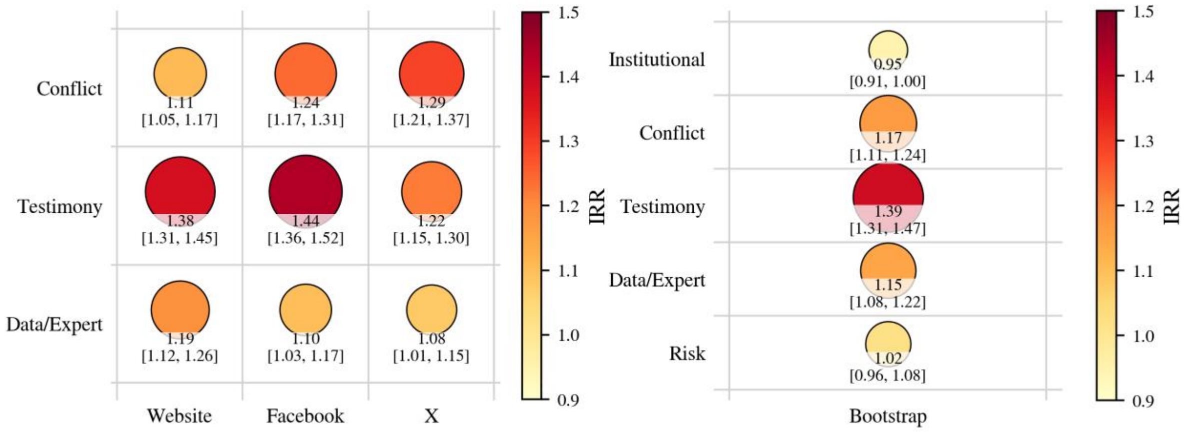



*Figure 4: Estimated effects of discourse patterns on shallow and deep audience engagement.*

In Figure 4 (a), the conflict/opposition type has the strongest pull on shallow participation, with an IRR of 1.34 and intervals of [1.28, 1.40]. The crisis/risk warning type is 1.21 [1.02, 1.27], the testimony/character story type is 1.12 [1.07, 1.18], the institutional/governance type is only 1.06 [1.02, 1.10], and the data/expert interpretation type is 0.97 [0.93, 1.02]. This indicates that title tension, risk warnings, and emotional packaging are more likely to generate clicks, browsing, and light reactions. Figure 4 (b) shows that the ranking of deep involvement has changed. The testimony/character story type has the highest IRR of 1.41 [1.33, 1.50]. The conflict/opposition type is 1.18 [1.12, 1.24], the data/expert interpretation type is 1.16 [1.09, 1.24], the crisis/risk warning type is close to the benchmark, the IRR is 1.03 [0.97, 1.09], and the institutional/governance type is 0.94 [0.89, 0.99]. From this, it can be seen that conflict based expression is more conducive to obtaining reach, testimony based narrative is more likely to lead to comments and lengthy discussions, multi-source citation and explanatory text clicks are not dominant, but they have a stable gain in deep interaction, which is consistent with the research conclusions of title narrative, testimony news, and information packaging in recent years. To further examine whether the impact of different discourse patterns on deep engagement varies with the platform environment and to examine the stability of the estimation results, this paper compared the effect values of each pattern on different platforms and reported the bootstrap resampling results, as shown in Figure 5.




(a) Platform heterogeneity of deep engagement (b) Bootstrap stability of deep engagement

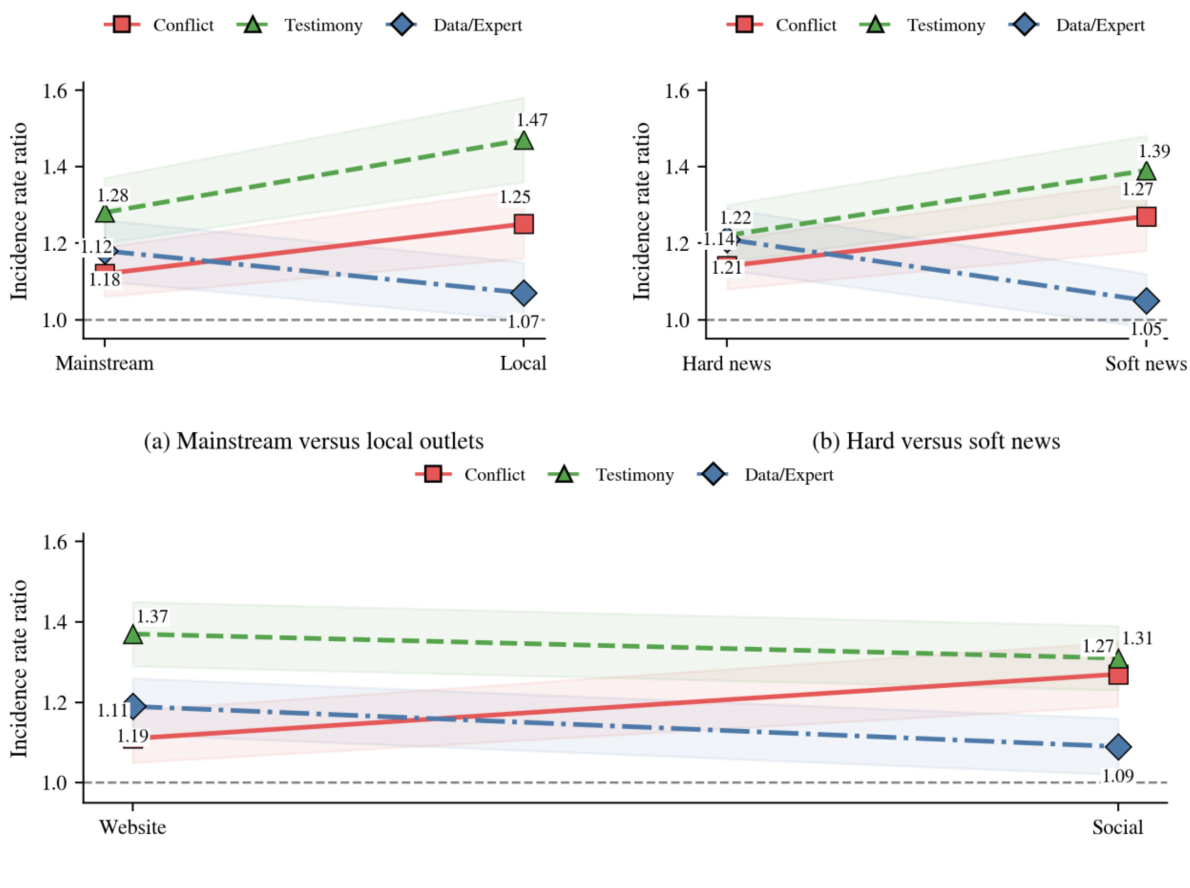
*Figure 5: Cross-platform differences and bootstrap stability of deep-engagement effects across discourse patterns.*

Figure 5 (a) shows the sub platform estimation for deep involvement. The conflict/opposition type has an IRR of 1.11 on the website, rising to 1.24 on Facebook and further increasing to 1.29 on X. The testimony/character story type on the website and Facebook were 1.38 and 1.44, respectively, and fell back to 1.22 on X. The data/expert interpretation type is 1.19 on the website, 1.10 on Facebook, and 1.08 on X. This indicates that X is more likely to amplify adversarial expressions, Facebook is more conducive to extending character narratives into comment chains, and the website is more friendly to explanatory texts. The verification results are shown in Figure 5 (b). In 100 bootstrap resamples, the median IRR for testimony/character story type was 1.39, with a 90% interval of [1.31, 1.47]. The conflict/opposition type is 1.17 [1.11, 1.24], the data/expert interpretation type is 1.15 [1.08, 1.22], the institutional/governance type is 0.95 [0.91, 1.00], and the crisis/risk warning type is 1.02 [0.96, 1.08]. The narrow intervals of each mode and the stable direction of the main effect indicate that the results are not driven by a few extreme hot articles.

Overall, shallow participation is more likely to be driven by a sense of conflict, risk, and title stimulation, while deep participation relies more on the full development of character experience, narrative input, and explanatory resources. The platform mechanism further amplifies this difference, so when discussing the audience impact of French news in the future, it is necessary to separate the understanding of "being seen" and "being continuously discussed".

## 3.3 Cross-Outlet Heterogeneity and Theoretical Implications

The heterogeneity results further indicate that the dissemination effect of discourse patterns does not maintain the same direction in all media environments. To compare the differences between mainstream media and local media, serious news and soft news, as well as between official websites and social platforms, this paper reports the deep participation estimation of a grouped mixed effects model, as shown in Figure 6.



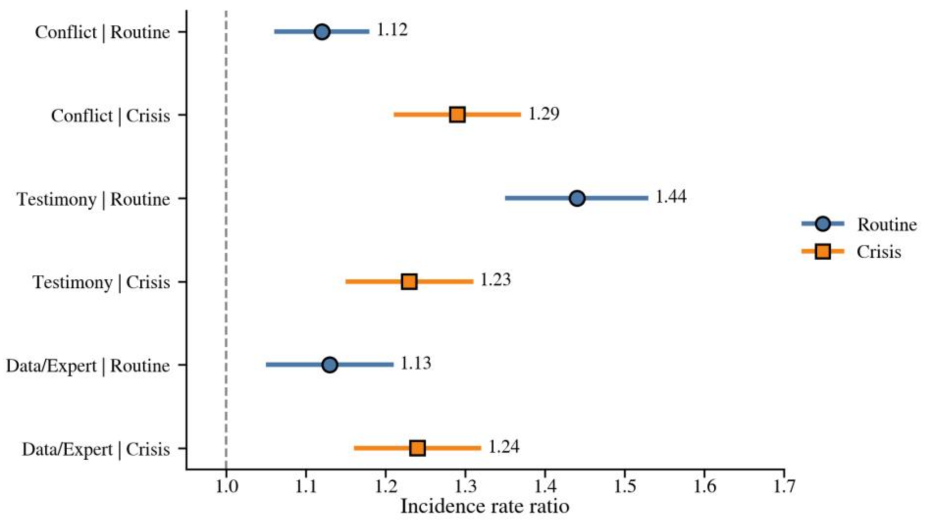

(a) Mainstream versus local outlets

(b) Hard versus soft news

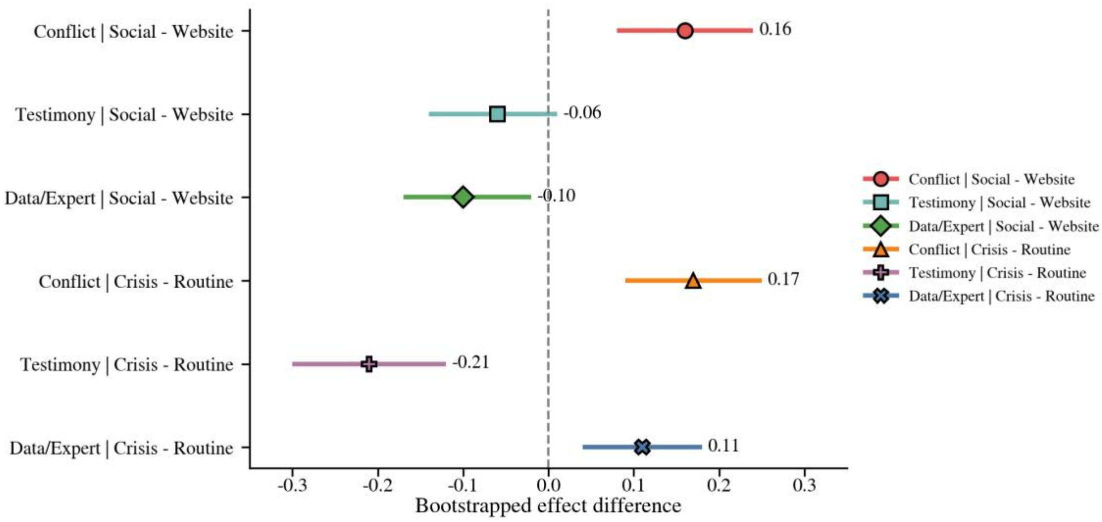
(c) Website versus social platforms

*Figure 6: Heterogeneous deep-engagement effects across outlet type, news style, and publication channel.*

In Figure 6 (a), the testimony/character story effect of local media is 1.47 [1.36, 1.58], which is higher than the 1.28 [1.20, 1.37] of mainstream media. Conflict/opposition type is also higher in local media, at 1.25 [1.16, 1.34], compared to 1.12 [1.06, 1.19] in mainstream media. Relatively speaking, the IRR of data/expert interpretation type in mainstream media is 1.18 [1.10, 1.26], higher than that of local media at 1.07 [1.00, 1.15]. This indicates that local media rely more on proximity and community experience to stimulate discussions, while mainstream media retains a stronger advantage in knowledge interpretation. Figure 6 (b) shows that the conflict/opposition type and testimony/character story type in soft news reach 1.27 and 1.39, respectively, both higher than the 1.14 and 1.22 in serious news. However, the effect of data/expert interpretation in serious news is 1.21 [1.13, 1.29], significantly higher than that of soft news at 1.05 [0.98, 1.12]. Figure 6 (c) further indicates that the effect of conflict/opposition on social platforms is 1.27 [1.19, 1.35], which is higher than the 1.11 [1.05, 1.18] on official websites. The testimony/character story type on the official website is 1.37 [1.29, 1.45], slightly higher than the social platform's 1.31 [1.23, 1.39]. The data/expert interpretation type is 1.19 [1.12, 1.26] on the official website and 1.09 [1.02, 1.16] on social media platforms. This set of results indicates that platform distribution amplifies the sense of conflict, but longer explanatory chains and character narratives are more likely to settle into ongoing discussions in the official website environment. The period differences and stability are shown in Figure 7.




(a) Routine versus crisis periods




(b) Bootstrap stability of heterogeneous gaps

*Figure 7: Period sensitivity and bootstrap validation of heterogeneous deep-engagement effects.*

Figure 7 (a) shows that the conflict/opposition type during the crisis period has increased from 1.12 [1.06, 1.18] in the normal period to 1.29 [1.21, 1.37], and the data/expert interpretation type has also increased from 1.13 [1.05, 1.21] to 1.24 [1.16, 1.32]. The testimony/character story type has decreased from 1.44 [1.35, 1.53] to 1.23 [1.15, 1.31]. Crisis situations compress the extension space of individual narratives, but enhance the demand for adversarial and explanatory information. Figure 6 (b) shows the resampling results. The conflict type gain difference between social platforms and official websites is+0.16 [0.08, 0.24], the conflict type gain difference between crisis and normal situations is+0.17 [0.09, 0.25], and the difference between data/expert interpretation on social platforms is -0.10 [-0.17, -0.02]. The difference between testimony/character story types during crisis periods is -0.21 [-0.30, -0.12]. Only the difference between the testimony type and the official website on social media platforms is close to the zero range of -0.06 [-0.14, 0.01], while the rest of the differences are stable in direction.

These results are closer to the explanation of the combined effect of platform logic and editorial adjustments. Platformized news enhances visibility competition, while news dataization continuously feeds back clicks, comments, and shares to the editorial department, enabling media to continuously adapt in terms of headlines, quotes, and narrative organization. At the same time, participatory journalism practices have also changed editors' judgments on who is speaking out and writing for whom, especially in local media and soft news. Therefore, the reconstruction of audience relations does not occur outside of the text. The discourse pattern itself has become an operational resource for media response platform indicators, organizing public connections, and reallocating news attention.

# 4 Conclusion

This article is based on the French news corpus from 2023 to 2025, systematically identifying the main discourse patterns and examining their differential impact on audience participation at different levels. The overall results indicate that the text organization of French news is not homogeneous, and different discourse configurations present measurable and distinguishable dissemination effects in platform communication and audience interaction.

(1) The discourse structure of French news in France has a clear and stable stratification, with institutional/governance type accounting for the highest proportion at 27.4%, followed by conflict/opposition type at 23.1%, testimony/character story type at 18.7%, data/expert interpretation type at 16.2%, and crisis/risk warning type at 14.6%.

(2) The impact of different modes on audience participation is not consistent. The conflict/opposition type has the strongest pull on shallow participation, with an IRR of 1.34. The testimony/character story type is most prominent for deep involvement, with an IRR of 1.41; The data/expert interpretation type is only 0.97 in shallow interaction, but reaches 1.16 in deep participation.

(3) The theoretical value of this article lies in linking French news research, discourse analysis, and audience quantification research, and replacing small sample case observations with large-scale French news computational analysis, thus presenting more specifically the interactive mechanism between platform based communication, editorial adaptation, and audience feedback.

(4) There are still limitations to this article, including platform interface limitations, difficulty in fully representing true attitudes through interactive indicators, and uneven availability of data from different media. Subsequent research can combine comment stance recognition, causal inference design, and cross lingual comparison of French and English news.

# About the Author

Yan Sun, born in Chengdu, China, in 1986. She earned a Master’s degree in Computational Linguistics from Université Paris Diderot (now known as Université Paris Cité), France and is currently affiliated with Xi’an International University, China. Her research primarily focuses on Natural Language Processing (NLP), French Language Pedagogy, and Computer-Assisted Language Learning (CALL).

# References


[1] Pignard-Cheynel, N., & Amigo, L. (2023). (Re)connecting with audiences. An overview of audience-inclusion initiatives in European French-speaking local news media. Journalism, 24(12), 2612–2631. DOI: 10.1177/14648849231173299.

[2] Standaert, O., Pignard-Cheynel, N., & Amigo, L. (2023). European French-Speaking Local Media's Relationship with Audiences: A Strategic Challenge between Diluted and Integrated Organizational Modalities. Digital Journalism, 13(6), 1070–1088. DOI: 10.1080/21670811.2023.2234410.

[3] Lamot, K., & Paulussen, S. (2025). News content promotion pressure: A content analysis of the cross-media visibility of and engagement with news stories. Journalism, 26(12), 2563–2580. DOI: 10.1177/14648849241279012.

[4] Tong, J., et al. (2024). Data journalism and audience engagement: Introduction to special issue. Journalism, 25(7), 1449–1459. DOI: 10.1177/14648849241256457.

[5] Schaetz, N. (2024). Journalism & audience datafication: How audience data practices shape inequity. Digital Journalism, 12(8), 1184–1204. DOI: 10.1080/21670811.2023.2171894.

[6] Murray, C., Riedl, M. J., & Stroud, N. J. (2024). Using Facebook Messenger versus Groups for News Engagement. Digital Journalism, 12(5), 637–655. DOI: 10.1080/21670811.2023.2208630.

[7] Dvir-Gvirsman, S. (2023/2025). The Meaning of Like: How Social-Media Editors and Users Make Sense of News Engagement Metrics. Journalism Practice. Advance online. DOI: 10.1080/17512786.2023.2228782.

[8] Robinson, S., Orozco, M., & Darr, J. P. (2025). How the Engagement Journalism Movement Is Changing Political News Content: An Applied-Research Study. Journalism & Mass Communication Quarterly. OnlineFirst. DOI: 10.1177/10776990251318649.

[9] Ju, Y., & Lee, P. (2025). More Sources Create Greater Audience Engagement: An Investigation into the Relationship Between the Number of News Sources and Audience Responses. Communication Research. OnlineFirst. DOI: 10.1177/00936502251324549.

[10] Uth, B., et al. (2025). The journalism-audience relationship in the digital age: A theoretical literature review. Journalism, 26(1), 45–64. DOI: 10.1177/14648849231221611.

[11] Uth, B. (2025). What Does the Audience Expect When Interacting with Journalists? A Q-sort Study. Journalism Practice. Advance online. DOI: 10.1080/17512786.2025.2551986.

[12] Opgenhaffen, M., & Hendrickx, J. (2024). Social media news editors as journalists or marketeers: Who are they and how do they identify themselves? Journalism, 25(6), 1383–1400. DOI: 10.1177/14648849231167282.

[13] Barnes, R., Mulcahy, R., & Riedel, A. (2025). Push notifications and news snacking: The impact of mobile news alert framing on reader engagement. New Media & Society, 27(3), 1486–1506. DOI: 10.1177/14614448231196580.

[14] Knauf, C., Reeves, H., & Mays, B. (2025). You're Being Kinda Pushy: Exploring How News Outlets Frame Push Notifications as Credible Clickbait to Engage with Their Audiences. Journalism and Media, 6(3), 96. DOI: 10.3390/journalmedia6030096.

[15] Hendrickx, J. (2025). "Normal news is boring": How young adults encounter and experience news on Instagram and TikTok. New Media & Society, 27(10), 5736–5754. DOI: 10.1177/14614448241255955.

[16] Or, S., et al. (2023). The Impact of Testimony Journalism on Audience Engagement: An Experimental Investigation of the Effects of Point of View. Journalism Studies, 24(5), 573–593. DOI: 10.1080/1461670X.2023.2173957.

[17] Hendrickx, J., Van Remoortere, A., & Opgenhaffen, M. (2023). Love, Like or Angry in Times of COVID-19? Analysing News Brands' Audience Engagement on Facebook Amidst a Pandemic. Journalism and Media, 4(3), 931–945. DOI: 10.3390/journalmedia4030060.

[18] Hendrickx, J., Van Remoortere, A., & Opgenhaffen, M. (2023). News packaging during a pandemic: A computational analysis of news diffusion via Facebook. Discourse & Communication, 17(6), 701–720. DOI: 10.1177/17504813231177280.

[19] Mukerjee, S., Yang, T., & Peng, Y. (2023). Metrics in action: how social media metrics shape news production on Facebook. Journal of Communication, 73(3), 260–272. DOI: 10.1093/joc/jqad012.

[20] Verstappen, M., & Opgenhaffen, M. (2024). Making it Fit: How Science News Gets Remediated for Facebook and Instagram. Journalism Studies, 25(9), 1010–1028. DOI: 10.1080/1461670X.2023.2263799.

[21] Feng, G. C. (2024). Effects of narratives and information valence in digital headlines on user responses. Asian Journal of Communication, 34(2), 156–177. DOI: 10.1080/01292986.2024.2317308.

[22] Jiang, X., Liu, L., Wu, B., Chen, L., & Lin, H. (2024). Which storytelling people prefer? Mapping news topic and news engagement in social media. Computers in Human Behavior, 158, 108248. DOI: 10.1016/j.chb.2024.108248.

[23] Gligorić, K., Lifchits, G., West, R., & Anderson, A. (2023). Linguistic effects on news headline success: Evidence from thousands of online field experiments (Registered Report). PLOS ONE, 18(3), e0281682. DOI: 10.1371/journal.pone.0281682.

[24] Escouflaire, L., Descampe, A., & Fairon, C. (2024). Automated text classification of opinion vs. news French press articles: A comparison of transformer and feature-based approaches. Language & Communication, 99, 129–140. DOI: 10.1016/j.langcom.2024.09.004.

[25] Akter, S. S., & Anastasopoulos, A. (2024). A Study on Scaling Up Multilingual News

Framing Analysis. Findings of ACL: NAACL 2024, 4156–4173. DOI: 10.18653/v1/2024.findings-naacl.260.