Pacific International Journal, Vol. 8(4); 2025
ISSN (Print) 2663-8991, ISSN (Online) 2616-4825[1]
DOI:10.55014/pij.v8i4.823
https://rclss.com/index.php/pij

Revisiting the Role and Potential of Corpus Linguistics in Critical Discourse Analysis

Yingfang Chen, Xiaolei Ma, Xuchuan Chen

National University of Defense Technology, China
E-mail: chenyingfang_cs@163.com

Abstract: Persistent debates on methodological compatibility between corpus linguistics (CL) and Critical Discourse Analysis (CDA) necessitate reassessing CL's untapped potential. Addressing criticisms of contextual depth and objectivity, this study leverages recent Natural Language Processing (NLP) advancements to demonstrate CL's enhanced capacity for critical discourse research. A comparative analysis of 398 New York Times and 1,201 China Daily articles on a contested international issue reveals—through integrated corpus techniques—systematic differences in agent representation, verb usage, semantic framing, and network centrality. These findings expose structural biases in global discourses. The synergistic integration of corpus methods with CDA frameworks strengthens analytical scope and methodological pluralism while maintaining core commitments to power-relation analysis. This underscores CL's vital role in advancing CDA's rigor and impact.

Keywords: Corpus Linguistics, Critical Discourse Analysis (CDA), Discursive Agents, Semantic Analysis, Social Network Analysis

1. Introduction

Since 1990s, Critical Discourse Analysis (CDA) offered linguistics, applied linguistics and language study, as well as communication a framework and a means of exploring the imbrications between language and social-institutional practices, and beyond these, the intimate links between language as discourse and broader social and political structures.[1] Then, the widespread adoption of corpus methodologies represents a significant developmental trend in the field of CDA. As a quantitative approach to analyzing large-scale textual data, corpus methods are regarded as a potential solution to long-standing criticisms of CDA, including concerns about insufficient textual representativeness and a lack of objectivity. When scholars such as Stubbs[2] and Hardt-Mautner[3] pioneered the integration of corpus linguistics with CDA, the synergy between these disciplines has grown increasingly robust[4][5]. However, academic reception of corpus methods remains polarized. Within CDA—a field traditionally rooted in qualitative analysis—corpus approaches are seen as inherently limited by unresolved methodological challenges. This raises a critical question whether advancements in corpus methodologies can finally address the persistent skepticism among CDA scholars.

Based on the above considerations, this study demonstrates possibilities through a corpus of 398 *New York Times* articles and 1,201 *China Daily* articles on a contentious international issue, illustrating how innovative methods including named entity recognition, syntactic parsing, semantic analysis, cluster analysis, social network analysis, etc. can advance critical discourse research.

2. Reflections on the Limitations of Corpus Methodologies

The widespread adoption of corpus methods in Critical Discourse Analysis (CDA) stems primarily from the traditional focus of CDA on single-text analyses. Such an approach allows for in-depth, nuanced interpretations of individual texts, leveraging the researcher's expertise in synthesizing social, cultural, and discursive knowledge. Exemplary CDA studies often reflect profound engagement with social realities, meticulous observation of cultural phenomena, and mastery of textual analysis[6]. Undoubtedly, such work holds significant value. However, analyses centered on single or limited texts face two major limitations: insufficient representativeness and subjective biases inherent to qualitative methods. Corpus linguistics offers a potential remedy for these issues. First, corpora—as large-scale electronic text collections—prioritize sampling representativeness and balance, enabling comprehensive insights into a given issue. Second, corpus methodologies, grounded in quantitative analysis, provide both abundant linguistic examples and statistical tools. Addressing subjectivity requires robust evidentiary support, which corpus-derived data can effectively supply. Moreover, while single-text analyses were once commendable in an era of limited information access, the digital age has exponentially expanded the volume of texts available for analysis, naturally pushing CDA toward larger text collections. Notably, both CDA and corpus linguistics have been influenced by functionalist traditions, facilitating methodological cross-pollination—though this is not the sole driver of their convergence, as corpus-based research is not exclusively tied to functional linguistics.

Despite these synergies, corpus methods have not been universally embraced by CDA scholars. Criticisms include claims that corpora lack contextual information, represent only what is said rather than what is unsaid, and cannot replace human critical thinking. While these critiques hold partial validity, they are not entirely fair. First, all methodologies have

[Received 05 March 2025; Accepted 06 May 2025; Published (online) 20, August, 2025]

strengths and limitations; corpora are no exception, offering data that illuminates specific dimensions of a research question. Second, while corpus data serves as vital evidence in CDA, its interpretive value depends on how analysts contextualize it with social realities. Corpus-derived patterns and manually identified textual features both serve as interpretive tools, each with inherent biases. Third, corpora, while limited to observable textual data, can still reveal unspoken discourses through statistical probabilities—especially when comparative perspectives highlight what is omitted or underrepresented by specific discursive agents over time.

Ultimately, discourse analysts prioritize intellectual rigor and sensitivity to dynamic contexts—precisely the factors that have driven CDA's turn to corpus methods. Even as CDA emphasizes contextual knowledge, it remains anchored in linguistic forms as analytical entry points—a practice shared with corpus linguistics. Criticisms of corpus methods for neglecting context overlook the fact that CDA itself begins with linguistic forms rather than abstract meanings. Furthermore, the lack of robust theoretical frameworks for contextual analysis may be addressed through future corpus-driven knowledge mining. Advances in artificial intelligence, particularly in encoding contextual knowledge as computational rules, could further bridge this gap.

To clarify: defending corpus methods does not imply viewing them as a panacea. Rather, they represent one data collection strategy among many. Their value lies in maximizing their strengths while mitigating limitations during interpretation.

3. Corpus-based studies of CDA

Existing critiques of corpus methodologies often reduce corpora to mere "databases," overlooking the analytical innovations central to corpus linguistics. Traditional corpus analyses—reliant on tools like WordSmith and AntConc for concordances, keywords, collocations, and frequency lists—have popularized corpus techniques but also constrained methodological creativity. This is particularly true for researchers with limited computational skills, who often depend on user-friendly software features.

Yet corpus analysis extends far beyond these basic functions. A core task of corpus linguistics and natural language processing is extracting meaningful linguistic patterns from corpora. Advances in artificial intelligence are revolutionizing the depth and breadth of natural language understanding, bringing us closer to automated identification of significant information in large-scale texts.

Collaboration between corpus linguistics and CDA could follow two paths: Methodological Inspiration: Cutting-edge corpus techniques could inspire novel analytical frameworks in CDA; Theoretical Guidance: CDA's research questions, theories, and frameworks could steer corpus-based textual analysis. The integration of these approaches may reveal new synergies, though researchers must remain open to alternative methodologies better suited to specific problems.

What emerging corpus techniques hold potential for CDA? This study illustrates some methods.

3.1 Named Entity Recognition

Named Entities refer to noun phrases within a corpus that denote specific entities such as person names, locations, and organizations. The significance of Named Entity Recognition (NER) lies in its ability to identify agents of social practice—key actors in critical discourse analysis (CDA). A foundational premise of CDA is that texts reflect social practices, and these practices are shaped by the agents (actors) who initiate or receive actions. Most CDA studies focus on specific groups, organizations, or nations as analytical subjects. Once named entities are identified, researchers can conduct targeted analyses of these agents. As Shi Xu argues, determining who speaks and who is silenced constitutes a critical dimension of CDA[7]. By applying NER to corpora, researchers can systematically identify which agents are privileged in discourse and which are marginalized.

The table below lists high-frequency discursive agents that appear in China Daily but are absent from *New York Times* coverage of the same issue.

Table 1: High-Frequency Discourse Agents in *China Daily* Not recontextualized in *New York Times*

Discursive Agents	Frequency	Discursive Agents	Frequency
Zhou Yongsheng	74	Qu Xing	18
Yang Bojiang	62	Wang Xinsheng	17
Lu Yaodong	53	Li Wei	17
Gao Hong	36	Liang Yunxiang	15
Wang Ping	32	China's Ministry of Commerce	12
the State Oceanic Administration	30	Jia Xiudong	10
Shen Shishun	23	Yuan Peng	9
Huo Jiangang	20	Yao Haitian	9

The data show that, apart from the State Oceanic Administration and China's Ministry of Commerce, the majority of discursive agents not recontextualized by *New York Times* are scholars and experts. For instance, Zhou Yongsheng, a professor at the Institute of International Relations of the China Foreign Affairs University and Deputy Director of the

Japanese Studies Center, ranks prominently in terms of frequency within *China Daily*, yet his perspectives remain absent from *New York Times* coverage. A small number of officials were similarly omitted. Overall, these un-recontextualized discursive agents can be categorized into three groups: Chinese scholars, Chinese government institutions, and Chinese officials. Conversely, certain experts cited by *New York Times*—including Ren Xiao, Ni Fang, and Zhan Qixiong—do not appear in China Daily.

In news reporting, scholarly discourse often conveys an impression of impartiality and objectivity. Through *China Daily*, a significant volume of perspectives from Chinese scholars are disseminated in English. However, such discourses have yet to gain substantial traction within Western mainstream media outlets. Whether these findings warrant further critical discourse analysis—particularly regarding power asymmetries in cross-media representation—requires researchers to contextualize them within specific theoretical frameworks aligned with their analytical objectives.

Tools capable of conducting Named Entity Recognition (NER) include the Stanford Named Entity Recognizer, among others. Using Stanford Named Entity Recognizer (NER), this study automatically identifies entities in *China Daily* and *New York Times*, focusing on person names, locations, and organizations. This step primarily aims to extract and quantify quoted discursive agents within the news.

3.2 Agency Analysis (Syntactic Parsing)

Traditionally, agency analysis, which examines how language assigns responsibility to agents, has relied on manual annotation. However, syntactic parsing, a computational linguistics process that analyzes sentence structure to uncover grammatical relationships between words, enables more efficient implementation of this process. While computational syntactic analysis is not entirely error-free, its results generally provide valuable auxiliary insights. Building upon Named Entity Recognition (NER), extracting agent-action relationships (i.e., identifying actions performed by specific agents) can generate critical data for Critical Discourse Analysis (CDA). In English sentences, the structural relationships between subjects and predicates are often complex, making simple part-of-speech (POS) tagging or regular expressions inadequate for extraction. Syntactic parsing offers a more robust approach to identify sentence subjects and their corresponding predicate verbs or phrases.

The table below lists predicate verbs associated with sentences where "China" functions as the grammatical subject in *New York Times*.

Table 2: High-Frequency Predicate Verbs in *New York Times* Sentences with "China" as Subject

Frequency	Predicate Verbs	Frequency	Predicate Verbs
26	say	8	maintain
24	send	8	show
16	becom	7	react
16	take	7	express
15	claim	7	consid
13	assert	7	go
12	use	6	stop
12	announc	6	put
12	make	6	establish
11	engag	5	want
10	tri	5	sought
10	demand	5	sign
10	call	5	set
9	declar	5	regard
9	accus	5	ratifi

Categorizing the verbs in the table above reveals the following semantic clusters:

(1) Discursive verbs that recontextualize China's stances, primarily say, claim, assert, announce, declare, accuse, express. These verbs constitute *New York Times*' predominant framing strategies for representing China's perspectives. Notably, drawing on Speech Act Theory, the articulation of positions in international politics often functions as a performative act. The discursive verbs employed by *New York Times* predominantly depict China's negative attitudes toward Japan, constructing a discursive image of China as grievance-prone, accusatory, and uncompromising.

(2) Action verbs denoting China's concrete measures, such as send, use, show, react. These verbs frequently carry pejorative connotations in context, portraying China as overreactive and inclined to resort to force—a narrative aligned with the "China threat" discourse perpetuated by *New York Times*. Unlike *China Daily*, which employs subjunctive mood (e.g., "would take" or "might adopt") to frame China's countermeasures as efforts to "prevent escalation," *New York Times* predominantly uses indicative mood to report actions as either completed or ongoing.

(3) Stative verbs reflecting China's situational positioning, including become, engage, try. Contrasting with *China Daily*'s usage of become to highlight achieved milestones in economic development and international cooperation, *New York Times* predominantly deploys this verb to speculate on China's potential or impending status. The verb engage almost

exclusively appears in contexts where China is framed as embroiled in disputes or conflicts with neighboring states, reinforcing a belligerent image. Similarly, the verb "try" functions as a preemptive evaluative marker, emphasizing China's purported eagerness to alter the status quo—a rhetorical move amplified by the incomplete action denoted by the infinitive verb phrases following try.

In summary, compared to *China Daily*, *New York Times* disproportionately emphasizes China's potential actions. It adopts a strategy of conflation and ambiguity in reporting China's implemented versus proposed measures. In contrast, *China Daily* maintains sharper discursive boundaries between support/opposition and completed/forthcoming actions. These divergent linguistic strategies in representing China's agency constitute a key factor underlying the contrasting affective stances in their coverage of the same events.

The table below lists the top 30 predicate verbs occurring in sentences where "Japan" functions as the grammatical subject within *New York Times* corpus.

Table 3. High-Frequency Predicate Verbs in *New York Times* Sentences with "Japan" as Subject

Frequency	Predicate Verbs	Frequency	Predicate Verbs
39	say	6	claim
13	control	6	hold
11	refus	6	give
11	call	5	maintain
11	administ	5	face
10	scrambl	5	send
9	make	4	win
9	take	4	return
8	releas	4	recogn
8	acknowledg	4	propos
7	respond	4	join
7	ask	4	continu
6	seiz	4	conclud
6	see	4	come
6	consid	4	back

The verbs in the table above can be broadly categorized as follows:

(1) Discursive verbs that recontextualize Japan's positions, such as say, call, respond, ask, propose. In terms of content framing, *New York Times* conspicuously portrays Japan as a calm, passive, and objective party in the dispute—a narrative starkly incongruent with factual representations in *China Daily*.

(2) Action verbs denoting Japan's concrete measures, including control, administer, scramble, release, respond, seize. Compared to action verbs used in *China Daily* to describe Japan's behaviors, *New York Times* systematically understates Japan's military provocations. For instance, verbs like scramble trivialize Japan's militaristic maneuvers, while control and administer implicitly frame Japan's territorial claims as legitimate administrative acts, reinforcing the discursive construction of Japan as a passive stakeholder.

(3) Light verbs like make and take. Unlike *China Daily*, *New York Times* predominantly employs light verb constructions to depict Japan's actions as measured and non-confrontational. This lexical strategy further exposes the newspaper's pro-Japan bias in its coverage.

In summary, *New York Times* consistently represents Japan as a reactive party countering China's assertiveness—a portrayal diametrically opposed to its construction of China as an aggressive actor. Conversely, *China Daily* foregrounds Japan's militaristic and provocative agency in the dispute. These divergent discursive strategies in verb selection and framing constitute a key mechanism through which the two media outlets construct conflicting ideological narratives about the same geopolitical event.

3.3 Semantic Analysis

Critical discourse analysis (CDA) typically prioritizes the semantic connotations underlying linguistic forms. Within corpus linguistics, semantic prosody analysis also constitutes a significant research strand, which investigates the consistent evaluative associations conveyed by words through their habitual co-occurrence patterns. However, semantic prosody studies traditionally rely on manual summarization by analysts through extensive concordance line examination. In practice, computational tools for semantic annotation have begun emerging in corpus research. While these tools remain underdeveloped, they nevertheless signify a methodological shift toward automation. Particularly in domain-specific studies, resolving ambiguities in linguistic forms becomes relatively more tractable.

The table below lists semantic categories with statistically significant differences in sentences where China versus Japan function as grammatical subjects, revealing contrasting discursive representations of the two nations. Due to space constraints, only the top 20 semantic categories by log-likelihood value are presented.

Table 4. Divergent Semantic Categories Associated with China and Japan as Subjects in *New York Times*

Code	Semantic Category	China Subject Freq.	Japan Subject Freq.	Log-Likelihood	Sig. (p<0.01)	Direction (C/J)
W3/M4	Geographical terms/Weather	212	63	27.94	0.000	+
A8	Seem/Appear	47	8	14.59	0.000	+
M7/G1.1	Places/Government etc.	14	0	13.30	0.000	+
X4.2	Mental object: Means, method	37	6	12.07	0.001	+
I2.1	Business: Generally	23	3	9.09	0.003	+
S1.1.2-	Reciprocity	9	0	8.55	0.003	+
K5.1/L2	Sports/Living creatures generally	25	4	8.26	0.004	+
M5fnc	Means of transport (Air)	39	9	8.23	0.004	+
X4.1	Mental object: Conceptual object	54	15	8.22	0.004	+
N3.2+/A2.1	Measurement: Size/Affect: Modify, change	51	14	7.95	0.005	+
G2.2+	General ethics	14	1	7.90	0.005	+
L2	Living creatures generally	14	1	7.90	0.005	+
S1.2.5+	Toughness; strong/weak	30	6	7.73	0.005	+
P1/G2.1	Education in general/Crime, law and order: Law & order	8	0	7.60	0.006	+
X3.2/E4.2-	Sensory: Sound/Happy/sad: Contentment	8	0	7.60	0.006	+
A1.1.1	General actions, making etc.	252	113	7.49	0.006	+
S4	Kin	13	1	7.09	0.008	+
Q2.2	Speech acts	412	200	6.99	0.008	+
Q1.2/W3	Paper documents and writing/Geographical terms	7	0	6.65	0.010	+
S8-	Helping/hindering	48	14	6.59	0.010	+

Table 4 reveals significant differences in the semantic resources employed by *New York Times* when reporting on China versus Japan, diverging markedly from *China Daily*'s strategies. Specifically, *New York Times* disproportionately utilizes the following semantic categories for China compared to Japan:

(1) Speculative modality (A8: Seem/Appear) and evaluative degree (A13: Degree), e.g., appeared, seemed, more, increasingly, nearly.

(2) Means/methods (X4.2: Mental object: Means, method) and general actions (A1.1.1: General actions, making), e.g., way, tactics, system, action, enforce.

(3) Economic discourse (I2.1: Business: Generally), e.g., economy, infrastructure, commercial.

(4) Unilateralism (S1.1.2-: Reciprocity) and obstruction (S8-: Helping/hindering), e.g., unilateral, fight, prevent, impede.

(5) Expansionist framing: Scale/growth (N3.2+/A2.1: Measurement: Size/Affect: Modify, change), e.g., growing, escalating, expanded; Power projection (S1.2.5+: Toughness; strong/weak), e.g., strong, strengthen;Transformative change (A2.1+: Affect: Modify, change), e.g., become, change.

(6) Negative affect (X3.2/E4.2-: Sensory: Sound/Happy/sad: Contentment), e.g., uproar.

(7) Speech act prominence (Q2.2: Speech acts), with *New York Times* disproportionately recontextualizing China's discourse through verbs such as *claim*, *declared*, *accused*, *announced*, *demanded*, *asserted*, *refused*, *denied*, and *criticized*. These lexical choices collectively construct China as hyperactive and interventionist. In stark contrast, Japan's action verbs are conspicuously underrepresented in *New York Times* coverage. This discursive asymmetry reinforces the newspaper's implicit narrative of constructing the subtext of a "China threat."

Conversely, *New York Times* foregrounds these semantic categories for Japan:

(1) Conflict avoidance (A1.9: Avoiding) and status quo maintenance (A2.1-: Affect: Modify, change), e.g., avoid, refrain, stabilize, stable.

(2) Measured responses (M1: Moving, coming and going), e.g., scrambled, move, steps.

(3) Modalized uncertainty (A7+: Definite (+ modals)), e.g., would, could, can, may.

(4) Military normalization (G3: Warfare, defence and the army; Weapons), e.g., military, wartime, weapons.

(5) Alliance-centric framing: Cooperation (S3.1/S2mf: Relationship: General/People), e.g., ally, partner; Membership (S5+/S2mf: Groups and affiliation/People), e.g., member.

This semantic asymmetry constructs Japan as a peacekeeping, conflict-averse nation while systematically positioning China as its antithesis. By accentuating Japan's alliance networks and institutional legitimacy, *New York Times* implicitly reinforces ideological binaries that shape readers' affective stances toward China.

Platforms enabling semantic analysis include UCREL and Wmatrix, which have been employed in prior studies. Using the UCREL Semantic Analysis System (USAS), this study performs semantic annotation on sentences from *China Daily* and *New York Times* that featured China or Japan as grammatical subjects. It then analyzes differences in these semantic resources between *China Daily* and *New York Times* with the Log-likelihood algorithm.

3.4 Cluster Analysis

Cluster analysis, a widely used technique in natural language processing, operates on the core principle of grouping semantically similar texts while distinguishing divergent ones. This method enables researchers to efficiently identify latent thematic patterns within large corpora. Compared to keyword analysis, cluster analysis offers three key advantages: (1) automated categorization without reliance on manual summarization; (2) self-contained operation requiring no reference corpora; and (3) scalability for processing extensive textual datasets.

The table below juxtaposes discursive agents and their associated discourse content across *China Daily* and *New York Times*, revealing systemic divergences in narrative framing.

Table 5. Comparative Analysis of Discursive Agents and Discourse Content in *China Daily* and *New York Times*

Cluster Properties	*China Daily*	*New York Times*
Discursive Agents	Chinese Academy of Social Sciences (CASS) experts	political science expertise
	Japanese studies scholars (Academic Institutions)	international affairs experts from universities (especially Fudan University)
	State Agencies & Spokespersons (e.g., Ministry of Foreign Affairs, State Council Information Office)	Chinese Ministry of Foreign Affairs and its spokespersons
	international affairs experts	China's Xinhua News Agency
		newspapers and media outlets
		news media Discourse
Discourse Content	Diaoyu Islands patrol	maritime patrols (e.g., in disputed waters)
	regional peace & stability	regional and global stability (especially regarding the Korean Peninsula issue)
	Diaoyu Islands Sovereignty	the Diaoyu/Senkaku Islands dispute
	East China Sea Air Defense Identification Zone (ADIZ)	East China Sea Air Defense Identification Zone (ADIZ)
	Japanese investment in China	military exercises
	China-Japan economic cooperation	"purchase of the Diaoyu Islands" controversy
	China-Japan military capabilities	military expenditure growth
	China-Japan bilateral relations	the South China Sea issue
	China-to-Japan tourism	
	China-Japan Leadership Summit	reciprocal visits between Chinese and Japanese leaders
	Liaoning Aircraft Carrier	

In terms of discourse agents, both *New York Times* and *China Daily* incorporate voices from academic experts, state institutions and their spokespersons, as well as international affairs specialists. The distinction lies in *New York Times*'s extensive reproduction of discourse from Chinese media outlets, particularly Xinhua News Agency. Furthermore, clustering analysis reveals divergent emphases in discourse representation between the two newspapers. For instance, *China Daily* demonstrates a stronger tendency to cite experts from the Chinese Academy of Social Sciences, whereas *New York Times* appears to quote Fudan University scholars more frequently. This indicates that although both publications reference scholarly perspectives, their selection of cited authorities differs. These findings corroborate earlier analytical conclusions: certain experts prominently featured in *China Daily* fail to achieve comparable visibility in *New York Times*.

Regarding discourse content, *China Daily* covers a broader spectrum of topics compared to the relatively constrained scope of Chinese discourse reproduced in *New York Times*. Both outlets address themes such as Diaoyu Islands patrols, the East China Sea Air Defense Identification Zone (ADIZ), regional peace and stability, sovereignty disputes over the Diaoyu Islands, military capacity expansion, and leadership summits. However, *China Daily*'s reproduced discourse additionally incorporates economically oriented topics like Japanese investment in China, Sino-Japanese economic cooperation, and tourism to Japan—subjects absent from *New York Times*. Notably, *New York Times* includes content related to the South China Sea issue, reflecting Western mainstream media's propensity to link the Diaoyu Islands dispute with South China Sea tensions. The South China Sea issue has manifested pronounced internationalization trends, whereas the Diaoyu Islands dispute remains comparatively contained within bilateral Sino-Japanese management

frameworks. Furthermore, *China Daily* addresses the Liaoning aircraft carrier—a topic entirely omitted in *New York Times*.

This divergence in thematic focus underscores the enduring discursive hegemony exercised by Western nations, which dominate the international agenda-setting process and maintain a highly monopolized international discourse landscape. Such structural imbalances constitute significant barriers to the effective dissemination of Chinese perspectives to Western audiences.

Multiple methodologies are available for clustering analysis, with CLUTO representing one of the more commonly employed tools. With CLUTO software, this study conducts cluster analysis on the speaker discourses in *China Daily* and *New York Times*. The clustering principle aims to maximize similarity within each category while minimizing similarity between categories.

3.5 Social Network Analysis

Social Network Analysis (SNA) is a methodological framework that maps and measures relationships between entities (e.g., people, organizations, words) to reveal hidden structures of power, influence, and information flow. In discourse studies, SNA uncovers how language constructs social realities by analyzing connections within textual data.

In a social network, there are some nodes, with each node representing a discourse agent. The centrality of a node and its size are determined by its frequency of association. In terms of network density, the *China Daily* network exhibits tighter connections between nodes. This is primarily due to *China Daily*'s significantly higher volume of articles on the Diaoyu Islands dispute compared to *New York Times*, resulting in greater co-occurrence probabilities in statistical data. Regarding the identities of discourse agents, both networks predominantly include voices from China, Japan, and the United States. A key distinction lies in the heightened prominence of U.S. agents within *New York Times* network.

In the two networks, notable differences emerge in the central discourse agents. Central agents in *China Daily*'s network—such as Zhou Yongsheng, Hong Lei, Liu Jianyong, Yang Bojiang, Gao Hong, and Lu Yaodong—occupy marginal positions in *New York Times* network. Conversely, Chinese agents are largely peripheral in *New York Times* network, while U.S. agents similarly occupy the edges in *China Daily*'s network. This indicates that both nations prioritize their own domestic agents in news reporting. However, Japanese agents—particularly Shinzo Abe and Shintaro Ishihara—occupy near-central positions in both networks, reflecting their heightened visibility and coverage in the Diaoyu Islands dispute.

Next, this study calculates the Betweenness Centrality of each node in the two social networks to identify core discourse agents and contrast the differences between the two newspapers.

Betweenness Centrality is a critical metric for assessing a node's importance within a network. Nodes with high betweenness centrality typically occupy connective positions between clusters. These nodes act as bridges for establishing relationships among other nodes. In discourse networks, agents with high betweenness centrality occupy core positions, and their viewpoints are more likely to elicit responses—whether supportive or oppositional—from other agents.

Table 6: Betweenness Centrality of Nodes in the *China Daily* Discourse Network

	Betweenness	nBetweenness
Kyodo_news_agency	21.204	2.049
Hong_Lei	19.086	1.844
Zhou_Yongsheng	18.803	1.817
Observers	18.256	1.764
Analysts	17.369	1.678
experts	13.825	1.336
Liu_Jiangyong	13.107	1.266
Japanese_media	12.8	1.237
Shinzo_Abe	11.585	1.119
Xinhua_News_Agency	11.488	1.11
Yang_Bojiang	9.463	0.914
Ruan_Zongze	8.481	0.819
Yomiuri_Shimbun	8.112	0.784
Japan	7.353	0.71
Experts	7.109	0.687
Gao_Hong	6.72	0.649
Reuters	6.086	0.588
Shen_Shishun	5.764	0.557
Yoshihiko_Noda	5.703	0.551

Lu_Yaodong	5.571	0.538
Qu_Xing	5.259	0.508
Wang_Ping	5.059	0.489
Koichiro_Gemba	4.844	0.468
Washington	4.447	0.43
Osamu_Fujimura	4.225	0.408
Yoshihide_Suga	3.824	0.369
Hua_Chunying	3.682	0.356
Shintaro_Ishihara	3.613	0.349
Huo_Jiangang	3.048	0.294
Yang_Jiechi	2.654	0.256
Wang_Xinsheng	2.641	0.255
Barack_Obama	2.631	0.254
Feng_Zhaokui	2.568	0.248
Xi_Jinping	1.803	0.174
Wang_Yi	1.545	0.149
the_State_Oceanic_Administration	1.445	0.14
Yang_Yujun	1.177	0.114
Li_Wei	0.891	0.086
Hillary_Clinton	0.61	0.059
Uichiro_Niwa	0.407	0.039
Chuck_Hagel	0.188	0.018
Toyota_Motor_Corp	0.033	0.003

Table 7: Betweenness Centrality of Nodes in *New York Times* Discourse Network

	Betweenness	nBetweenness
Xinhua_News_Agency	64.325	5.703
Japanese_officials	55.382	4.91
Shinzo_Abe	51.017	4.523
analysts	37.365	3.313
American_officials	35.488	3.146
Japanese_Coast_Guard	29.59	2.623
Hillary_Rodham_Clinton	26.114	2.315
Japanese_governments	22.231	1.971
experts	21.029	1.864
Chuck_Hagel	17.307	1.534
the_Pentagon	17.078	1.514
Chinese_government	13.097	1.161
Barack_Obama	12.478	1.106
the_Chinese_news_media	11.863	1.052
Shintaro_Ishihara	11.437	1.014
Hong_Lei	11.154	0.989
Shi_Yinhong	10.493	0.93
the_Obama_administration	10.106	0.896
Chinese_officials	9.818	0.87
Yang_Yujun	8.945	0.793
Yoshihiko_Noda	8.444	0.749
The_United_States	7.195	0.638
Global_Times	7.107	0.63
Xi_Jinping	5.798	0.514

The_Chinese	5.604	0.497
Yoshihide_Suga	4.913	0.436
Chinese_analysts	4.673	0.414
Itsunori_Onodera	4.058	0.36
Jiang_Yu	3.643	0.323
Narushige_Michishita	3.614	0.32
Japanese_news_media	3.534	0.313
Hua_Chunying	2.952	0.262
administration_officials	2.254	0.2
Fumio_Kishida	2.053	0.182
Defense_officials	1.827	0.162
M_Taylor_Fravel	1.676	0.149
Leon_Panetta	1.521	0.135
Osamu_Fujimura	1.313	0.116
Seiji_Maehara	0.904	0.08
Reuters	0.9	0.08
Beijing	0.862	0.076
a_senior_administration_official	0.82	0.073
Western_diplomats	0.665	0.059
Donald_J_Trump	0.474	0.042
Kyodo_News_agency	0.32	0.028
Naoto_Kan	0.31	0.027

A comparison of Tables 6 and 7 corroborates the analysis of Figures 1 and 2. The core discourse agents differ significantly between the two networks. In *China Daily*, the central agents include Kyodo News Agency, Hong Lei, Zhou Yongsheng, Liu Jiangyong, Shinzo Abe, Xinhua_News_Agency, Yang Bojiang, Ruan Zongze, and Shintaro Ishihara. In *New York Times*, core agents are Xinhua News Agency, Shinzo_Abe, Japanese Coast Guard, Hillary Rodham Clinton, Chuck Hagel, Barack Obama, Shintaro Ishihara, Hong Lei, and Shi Yinhong.

Notably, several central agents in China Daily—such as Zhou Yongsheng, Liu Jiangyong, Yang Bojiang, and Ruan Zongze—are primarily scholars. While these figures enjoy high visibility in Chinese media, they remain marginal in Western mainstream outlets like *New York Times*. Conversely, Shi Yinhong—a scholar less prominent in China Daily—receives greater attention in *New York Time*s. This highlights divergent discursive priorities in Sino-U.S. reporting on the Diaoyu Islands. From the U.S. perspective, coverage selectively amplifies voices aligned with its ideological narratives. How to ensure Chinese scholars' voices resonate not only domestically but also within Western mainstream media remains a critical question.

Common tools for social network analysis include UCINET and Pajek. Co-occurrence matrices are imported and analyzed for centrality metrics in this study and networks from UCINET were visualized using Pajek.

4. Conclusion

The above analysis introduces only a subset of natural language processing (NLP) techniques. The integration of semantic prosody, social network analysis, and agents mapping in this study extends Fairclough's dialectical-relational model by quantifying "power in discourse" as operationalized through network centrality disparities and evaluative language patterns. In reality, NLP encompasses a far broader array of methodologies, such as mutation analysis and topic modeling. This paper serves merely as a preliminary exploration, aiming to demonstrate how linguistic research might increasingly integrate insights from computer science.

It must be noted that the accuracy of the analytical methods discussed here remains subject to improvement. A central challenge in NLP lies in continuously refining the precision of these tools. While linguists often prioritize flawless data accuracy, current NLP technologies inevitably exhibit varying degrees of bias or error. However, it is equally important to acknowledge the significance of probability in such analyses. When processing large text corpora, the identification of prevailing trends retains substantial interpretive value, even amidst localized inaccuracies.

Although this study emphasizes the utility of corpus-based techniques, it does not diminish the indispensable role of the critical discourse analyst. Beyond data acquisition, the crux lies in determining which forms of data to prioritize and how to interpret results rigorously. Here, theoretical frameworks within critical discourse analysis remain paramount. If future corpus-driven critical discourse analysis is to advance meaningfully, it will hinge on groundbreaking theoretical developments. After all, tools exist to serve human inquiry—not to supplant it.

References

[1] Fairclough, N. (1995). *Critical Discourse Analysis: The Social Sciences in Question*. Routledge.

[2] Stubbs, M. (1993). *Text and Corpus Analysis*. Cambridge University Press.
[3] Hardt-Mautner, A. N. (1993). *Introducing Corpora: Non-technical Guides for Social Science Research*. Multilingual Matters.

[4] Van Dijk, T. A. (1998). *Discourse and Power*. Cambridge University Press.

[5] Van Leeuwen, T. *Discourse and Practice: New Tools for Critical Discourse Analysis*[M]. New York: Oxford University Press, 2008.

[6] Zhu, G, & Lin, L. (2025). The Image Construction of China-Europe Freight Trains by American and British Think Tanks from the Perspective of Critical Discourse Analysis. *Think Tank:Theory & Practice,* 10(01),138-148.
[7] Shi, X. (2010). *Cultural Discourse Studies: Exploring Chinese Theories, Methods, and Issues*. Peking University Press.