Comparison of Citation Categories in Chinese, Japanese, and English News Discourses Based on Corpus

Cuicui Li

School of Foreign Languages, Hebei North University, Zhangjiakou, Hebei, China
442733980@qq.com

Abstract: Since the 19th century, research on reporting language has been flourishing in Western countries, represented by English, and has been extensively explored from a syntactic perspective. In recent years, with the rise of disciplines such as pragmatics and discourse analysis, more and more attention has been paid to the discourse structure and pragmatic function of citation. This article took news corpus as the topic, based on the relevant theories of English, Chinese, and Japanese reporting languages, and referring to the research results of Japanese scholars, conducted an in-depth analysis of the discourse pragmatic functions of the quoted language in the news report from two aspects: reporting methods and sources, and explored how the narrator used these methods to achieve their intended expression. The experimental results showed that in English newspaper news reports, indirect discourse appeared 402 times, accounting for 42.41%, nearly half. Compared to English and Chinese, direct citation in Japanese only accounted for 8.96%.

1. Introduction

Quotations are a common language phenomenon in Chinese news discourse and an important discourse function. In news discourse, the use of quotations plays a role in enhancing the authenticity of content, slowing down the narrative pace, increasing the fun of the story, and to a certain extent creating a sense of "presence", implying the author's true attitude towards news reporting. In news discourse, there are a large number of quotations, and their differences are mainly reflected in the quoted sentences and punctuation. Quotations generally include the source of the quote and the verb used in the quote. For example, "General Secretary emphasized", "General Secretary" is the source of quotation and "emphasized" is the verb of quotation. Different quotes have different pragmatic functions.

In Chapter 3, this article introduces the selection of direct quotes and quotes in news discourse. In Chapter 4, an experimental analysis of the frequency of various forms of news reporting is presented, and a summary of the entire article is made.

2. Related Works

Experts have long conducted specialized research on news discourse citation. "News

terminology" is a method proposed by Orwell in 1984 to manipulate the political views of the public. Shesterina E A focused on analyzing the discourse expressions of contemporary German politicians. Through the analysis of German political discourse, it was pointed out that news grammar has two effects on political discourse: one is the confusion of existing concepts, and the other is the reinterpretation of existing concepts [1]. Konopelkina O analyzed the semantic system from multiple aspects such as linguistics, cognition, and sociology, and analyzed the conditions under which speakers use discourse in discourse. The practical significance is to define the language and cognitive characteristics in English political discourse [2]. Kula A M's survey found that a large number of citations were cited in social media (profiles of organizational anchors). There are many ways to cite, and it can have various impacts. Twitter and Facebook had 73% and 61% of direct citations, respectively [3]. Zarinabonu K specialized in discussing paraphrasing in newspapers. Since ancient times, paraphrasing, along with other rhetorical devices, has been regarded as a means of decorating language, endowing language expression, and enhancing language effectiveness. Zarinabonu K believed that paraphrasing is considered a genre [4].

Pavone L pointed out that impure direct/mixed quotes, which were translated (or modified, improved) direct and mixed quotes, had certain enlightening significance for understanding the general operation of quotes [5]. Brennan N M established a set of standards for evaluating the quality of access citations, and conducted research on the processing of interview records, selection of citations, quantity and length of citations, location and presentation of citations, ratio of cited respondents, translation of citations, and disclosure of editing protocols [6]. Warsidi W used genre analysis and descriptive reporting methods to compare and analyze citations in English and Indonesian as examples. Research has found that English and Indonesian authors often use descriptive methods when citing data sources, rather than other methods [7]. Wenno E C used magazine content as the data source, focusing on listening and analysis, and strictly identified the parts of the magazine that include indirect citations, direct citations, and explanatory citations of theories. From the commentary results, it can be seen that some citations do not include relevant information and reference materials, and the sources of citations are incorrect [8].

Lihawa K aimed to explore how to prevent academic plagiarism from the perspectives of academic literature, academic journals, etc., and apply it to the writing of academic papers, providing theoretical basis for future academic research. Under this qualitative analysis method, the score of a paper can be divided into three levels: the first is the argument selected before the citation; the second is the correct rhetorical device; the third is the key points and associations in good sentences after citation [9]. Tripp J used verifiable internal citations to test Baukham's theory of "internal citations," where a separate article discusses the initial speech act and is then cited by the speaker or other characters. It was found that most of the citations were modifications to the original narrative, and the number of times witnesses repeated them was similar to that of others [10]. Auer M conducted an investigation through interviews with 22 newly appointed local government officials from Australia and Canada. He focused on two leadership traits that influenced the Prime Minister's news coverage in the first week: strong/weak and capable/incompetent. He found that journalists evaluate the Prime Minister's abilities through anonymous evaluations, rather than from sources. However, when they are cited, they are often elites [11]. Wang Y studied the correlation between language and reporting opinions, as well as the differences in the way English media interact. Although the geographical distribution of bilingual news reports in Chinese and English is similar, there are significant differences in the content of the reports, interviewees, management strategies used, and the causes of conflicts [12]. Wright A V studied various factors in the translation of Indian newspaper authors: guidance methods and constraint mechanisms. He aimed to explore how Indian journalists perceive their reporting in multiple languages and how much importance they place on news translation when transitioning from Hindi interviews to

English reporting. He conducted in-depth research on the use of Hindi, English, or other Hindi languages from the context of news in India [13].

AL Harrasi K T S described an action study conducted by researchers at the University of Oman to address the issue of plagiarism, with the aim of helping university students make full use of data resources for second language writing. Through a survey questionnaire, this study explored the level of mastery of citation skills among university students [14]. Zhang H conducted research on citation classification and compared the classification methods of full-text citations and incomplete citations, and concluded that format based classification is necessary. He established a new "incomplete citation type" based on the formal characteristics of citations. The citations in the article, such as choir citations and solo citations, were arranged in descending order of clarity [15]. The existing research on citation categories in Chinese, Japanese, and English news discourse is not in-depth enough.

3. Methods

3.1 Direct Quotes in News Discourse

In Western media, both academia and journalism attach great importance to the use of direct speech. In the West, especially in the United States, journalism schools have made direct speech the main teaching content of a course. In a book called "News Reporting and Writing" by the University of Missouri, it is stated that "direct citation can add color to news and enhance credibility." Through direct citation, the other party can know that they have established a direct connection with the speaker. When viewing quotation marks, it can be immediately understood that it is an unusual thing. Moreover, direct citation can also change the rhythm of news, instantly turning the originally serious expression from dark to clear. American journalists are always interested in the use of direct speech. It may be thought that in the United States, news is not considered news without a direct quote [16-17].

Direct speech and indirect speech are the two most commonly used application methods, the former is to quote someone else's original words without leaving them. Indirect citation refers to explaining the propositional content in the original text, rather than limiting the language used. Compared with indirect speech, direct speech often cannot fully maintain the oral characteristics of the original text, such as colloquialism, idioms, and prosody, due to its own characteristics, while maintaining the original style [18]. In literary works, the fundamental goal of translators using paraphrasing is to enable readers to directly or indirectly hear the discourse of different characters. In the text, no matter what kind of quoted language is used, it helps to form what Bakhtin called "polyphony". On the other hand, novelists and journalists also intervene to varying degrees between the reported and the reader according to their communication goals, in order to influence the reader with the language being reported and achieve the expected results [19]. Therefore, when studying the pragmatic functions of various quotations, it is necessary to first examine their degree of intervention in the quoted discourse and the communicative goals they achieve [20].

3.2 Selection of Quotations

Before writing a press release, it is important for journalists to be clear that not all articles have good citations. Choosing the right meaningful and valuable citations can add color to the news; on the contrary, if the citation is inappropriate, it not only affects the quality of the news, but also fails to convey the intended message. Below are some guidelines to follow when selecting quotes.

(1) Distinctive statements

Distinctive comments can often leave a deep impression. In interviews, such language is often

the most memorable for journalists. These words may be wisdom, true emotions, or specific numbers. On the contrary, bad quotes are empty, flashy, and without substance, which clearly goes against the purpose of news dissemination. Therefore, when choosing quotations, journalists should try to use visually impactful words to leave a deep impression on the audience.

(2) Statement with rich connotations

This type of discourse contains a lot of information, not just expressing one's own meaning in one sentence, but conveying one's own meaning politely to others. This may be due to the limitations of the interviewee themselves, or it may be due to factors such as excessive allergies in the interview content. As a journalist, one should be good at grasping news clues and recording everything the interviewee says.

4. Results and Discussion

4.1 Corpus and Methods

This study selected 50 articles in Chinese, English, and Japanese as the research subjects. Among them, the Chinese discourse was selected from the People's Daily and Xinhua News Agency, while the English discourse was selected from the Washington Post and Los Angeles Times. The total number of Chinese words was 134268, and the total number of English words was 191014. The Japanese discourse was selected from Asahi Shimbun and Shogun Shimbun, with a total word count of 152312. After selecting the corpus, the author manually annotated the paraphrasing forms in each article using UAM Corpus Tool according to different classification modes. Then, the frequency of the encoded paraphrasing forms was compared and analyzed using SPSS software.

4.2 Statistical Results and Discussion

According to statistics, there were a total of 998 reported forms in 50 Chinese newspaper discourses, and 940 reported forms in 50 English newspaper discourses. Table 1 lists the frequency of the main forms of paraphrasing in Chinese English Japanese newspapers. Among them, the types of quotations include: DS (Direct Speech), IS (Indirect Speech), PDS (Pre-tuned Direct Speech), SQ (Safety quotation), NRSA (Narrative relaying of speech acts).

Table 1: Frequency statistics of various forms of news reporting

	DS	IS	PDS	SQ	NRSA	Total	Total number of words
English	336	402	32	109	69	948	191014
	35.44%	42.41%	3.38%	11.50%	7.27%		
Chinese	223	465	36	215	65	1004	134268
	22.22%	46.31%	3.59%	21.41%	6.47%		
Japanese	103	500	12	34	500	1149	152312
	8.96%	43.52%	1.04%	2.96%	43.52%		

From Table 1, it can be seen that the citations in English corpus were much fewer than those in Chinese, although there were more words in English than in Chinese. 336 direct quotes appeared in English newspaper news discourse, accounting for 35.44% of all citation methods; in Chinese newspaper news reports, there were a total of 223 direct introductions, accounting for 22.22%; from the comparison between the two languages, it can be seen that English tends to prefer the use of direct speech. In English newspaper news reports, indirect discourse appeared 402 times, accounting for 42.41%, nearly half of the total. Chinese indirect discourse accounted for 46.31%, indicating that there are more indirect discourse in Chinese newspaper news reports than in English

newspapers. Compared to English and Chinese, direct citation in Japanese only accounted for 8.96%.

Table 2: Statistical data on the average values of 5 types of paraphrasing forms appearing in newspapers

Reference type	Languages	Mean	Std.Deviation	
DS	English	6.74	4.92	t=2.21
	Chinese	5.41	5.06	df=92.251
	Japanese	2.15	3.87	p=0.031
IS	English	8.67	6.05	t=-0.47
	Chinese	10.06	8.45	df=81.644
	Japanese	5.34	6.84	p=0.467
PDS	English	0.92	1.37	t=0.26
	Chinese	1.57	1.79	df=87.493
	Japanese	3.4	3.49	p=0.632
SQ	English	2.5	2.66	t=-2.81
	Chinese	4.6	4.91	df=58.064
	Japanese	1.72	6.41	p=0.003
NRSA	English	1.46	2.16	t=0.53
	Chinese	2.18	2.13	df=92.541
	Japanese	3.01	5.71	p=0.886

From Table 2, it can be seen that there were many differences in the mean frequency and standard deviation of different citations in news reports in English, Chinese, and Japanese. The average frequency of direct citations in English news reports was three times that of Japanese reports. The directly referenced t-value was 2.21; the degree of freedom (df) was 92.251; the p-value was 0.031. These values were typically used for statistical hypothesis testing. In this case, a p-value of 0.031 indicated a significant difference in the frequency of direct citations in news reports in English, Chinese, and Japanese at a significance level of 0.05. The t-value of indirect citation was -0.47; the degree of freedom (df) was 81.644; the p-value was 0.467, indicating that there is no difference in the frequency of indirect citation in news reports in English, Chinese, and Japanese. The pre-tuned direct introduction had a t-value of 0.26, a degree of freedom (df) of 87.493, and a p-value of 0.632. In statistics, the p-value represents the probability that the observed data is consistent with the original hypothesis. In this case, a p-value of 0.632 indicated a significance level greater than 0.05, indicating no difference in the frequency of pre-tuned direct introductions in news reports in English, Chinese, and Japanese. The t-value of the warning introduction was -2.81; the degree of freedom (df) was 58.064; the p-value was 0.003. In statistics, the p-value represents the probability that the observed data is consistent with the original hypothesis. In this case, a p-value of 0.003 indicated a significant difference at a significance level of less than 0.05. The t-value of the narrative frequency of speech acts was 0.53; the degree of freedom (df) was 92.541; the p-value was 0.886. In statistics, there was no difference.

In news discourse, news reporters often use different forms of reporting to meet different reporting needs. The form of paraphrasing refers to the choice made by news reporters between direct and indirect quotes when quoting someone else's words. Paraphrasing quotes are divided into three categories: (1) direct discourse; (2) transformed direct discourse; (3) indirect discourse. Transformed direct discourse is a mixed form of direct discourse and indirect discourse, which contains indirect discourse. It does not use direct speech throughout the sentence, but instead adds double quotes to individual words or phrases in the indirect speech. Indirect discourse refers to the indirect retelling of the original content by news reporters, which can demonstrate their viewpoints
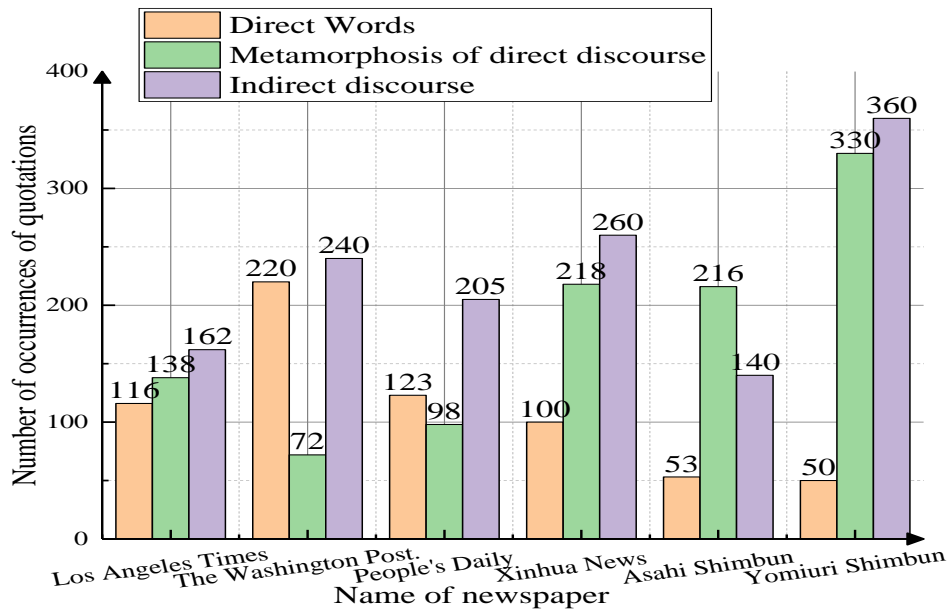
and attitudes.

Figure 1: The form and frequency of paraphrasing quotations

From Figure 1, it can be seen that the frequency of direct quotes in the Japanese news "Sales News" was 50, and the frequency of direct quotes in Japanese was relatively low. The direct quotes for Asahi News and Sales News were 53 and 50, respectively.

Academic journals or research papers provide corpora or related data on literary works and technological literature. This experiment obtained experimental corpus data from academic journals or research papers. In order to compare the citation frequency between news discourse and other field discourse, a comparative analysis was conducted on the frequency of citations in different languages, while ensuring that the text word count of the corpus is not significantly different.
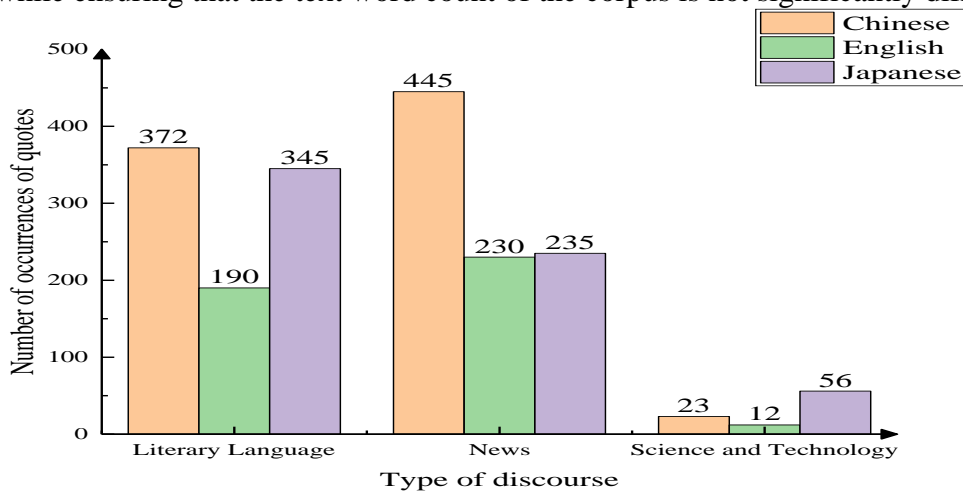

Figure 2: The frequency of citation mention and usage under three discourse types

From Figure 2, it can be seen that the frequency of citations appearing in Japanese literary, news, and technological discourse was 345, 235, and 56 times, respectively. The frequency of citations appearing in Chinese literary, news, and technological discourse was 372, 445, and 23 times, respectively. The frequency of citations appearing in English technology discourse was 12. Technology discourse usually focuses on conveying facts, data, and scientific viewpoints, rather than quoting other people's opinions or statements. Therefore, in texts in the field of technology, the

use of quotations may be relatively limited.

5. Conclusions

Language is an important tool for human communication and expression of ideas, and news discourse, as a special form of language, plays an important role in disseminating information, shaping public opinion, and reflecting social reality. Quotations are a common rhetorical device in news discourse of different languages, used to support reporting content, increase credibility, and enrich textual information. This study aims to conduct a comparative study of news discourse corpora in Chinese, Japanese, and English, explore their similarities and differences in citation categories, and gain a deeper understanding of the characteristics of news reporting in different language and cultural backgrounds. This study adopted corpus analysis method, collecting news corpora from Chinese, Japanese, and English languages, and selecting the cited discourse as the research object. Firstly, news corpora in each language were screened and organized to extract cited discourse, and corresponding databases were established. Secondly, by comparing and analyzing the differences in form, structure, and function of citation texts in Chinese, Japanese, and English, this study explored the impact of different language and culture on news reporting.

Acknowledgement

References

[1] Shesterina E A. "Newspeak" in Contemporary German Political Discourse. Professional Discourse & Communication, 2023, 5(3): 93-106.

[2] Konopelkina O, Yashkina V, Bezrodnykh I. Linguistic and Cognitive Features of English-Language Political Discourse. Studies in Media and Communication, 2022, 10(3): 109-116.

[3] Kula A M, Grzelka M. Quotation in Social Media: How Sharing Other People's Words Could Increase Misinformation. Studia Humanistyczne AGH, 2022, 21(2): 81-98.

[4] Zarinabonu K. The Stylistic Functions of Paraphrasing in the Newspaper. Current Research Journal of Philological Sciences, 2023, 4(09): 14-19.

[5] Pavone L. The semantics and pragmatics of impure direct/mixed quotation. Intercultural Pragmatics, 2023, 20(3): 239-250.

[6] Brennan N M. Methodological insights: interview quotations in accounting research. Accounting, Auditing & Accountability Journal, 2022, 35(9): 382-411.

[7] Warsidi W, Irawan A M, Adnan Z. Citation studies in English vs. Indonesian research article introductions (RAIs) in the history discipline. Studies in English Language and Education, 2023, 10(2): 598-613.

[8] Wenno E C, Karuna K, Serpara H. A critical review of whole theory: Stationenlernen learning technique and German language learning outcomes. Journal of Education and Learning (EduLearn), 2022, 16(3): 392-399.

[9] Lihawa K, Machmud K, Helingo A. Portraying Academic Writing Apptitude And The Plagiarism Issues Among EFL Students. English Review: Journal of English Education, 2023, 11(2): 515-524.

[10] Tripp J. The Eyewitnesses in their Own Words: Testing Richard Bauckham's Model Using Verifiable Quotations. Journal for the Study of the New Testament, 2022, 44(3): 411-434.

[11] Auer M, Trimble L, Curtin J. Getting straight to the source: who evaluates the leadership skills of premiers in Canada and Australia. Australian Journal of Political Science, 2023, 58(1): 1-18.

[12] Wang Y, Mumby H. Differences in reporting human-wild boar interactions in Chinese and English news media. Human Dimensions of Wildlife, 2023, 28(4): 295-310.

[13] Wright A V. On the [translated] record: Journalistic translation and creative agency in India's multilingual reporting. Journalism, 2022, 23(7): 1449-1468.

[14] AL Harrasi K T S. Developing a needs-based plagiarism management in second-language writing in a higher education institute: practice-oriented research. Instructional Science, 2023, 51(6): 1079-1115.

[15] Zhan H, Shi X. Explicitness of Attribution in Academic Discourse. *International Journal of Translation, Interpretation, and Applied Linguistics (IJTIAL)*, 2022, 4(1): 1-13.

[16] Putra A A, Gulo I. Identifying schizophrenia using indirect characterization of the main character in Love for A Deaf Rebel. *Linguistics and Literature Journal*, 2023, 4(1): 42-49.

[17] Kapidzic S, Neuberger C, Frey F, et al. How news websites refer to Twitter: a content analysis of Twitter sources in journalism. *Journalism Studies*, 2022, 23(10): 1247-1268.

[18] Khabib S. Introducing artificial intelligence (AI)-based digital writing assistants for teachers in writing scientific articles. *Teaching English as a Foreign Language Journal*, 2022, 1(2): 114-124.

[19] Sinaga J, Sinambela J. Dissemination of Citation Writing Principles and Techniques in Scientific Article Manuscripts. *Jurnal Pengabdian dan Pemberdayaan Masyarakat Indonesia*, 2023, 3(5): 225-232.

[20] Busri H, Badrih M. Representation of linguistic characteristics in mass media. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, dan Pengajarannya*, 2022, 8(1): 1-14.