



Analysis on Intertextuality Mining of News English Based on Computer Corpus

Yuanyuan Chai^(✉) and Hongding Bai

School of Foreign Language, Jingchu University of Technology, Jingmen 448000, Hubei, China
276708241@qq.com

Abstract. Nowadays, with the rapid development of information technology and increasingly frequent international exchanges, English news, as a medium for communicating the world and transmitting information, is playing an increasingly important role. For English learners and researchers, English news in newspapers, periodicals and magazines has become a convenient means to obtain information related to English language. Political English news often accounts for half of the world's major English publications, which is not only in newspapers and magazines published in English-speaking countries, but also in non-English-speaking countries. Considering the specific research practice, this paper selects the British National Corpus, which is the largest online corpus and widely used in linguistic research, as the corpus source of political English news published in the UK. Intertextuality is an important feature of discourse, and any discourse is the absorption and transformation of other discourses. The reported discourse in news discourse reflects the intertextuality of discourse, which is often used by news reporters to express their ideology. The main research object of CDA is news and other non-literary texts. Critical discourse analysis of reported discourse in news discourse can help readers understand the true intention of news reporters behind news reports. To sum up, this paper selects the research category of English news, based on the corpus research method, takes the political English news as the text, compares and analyzes the stylistic characteristics of Chinese and English political English news, and tries to find out the stylistic differences of political English News published in China and Britain. The stylistic research with news entities as the text and corpus based research method is highly credible. Therefore, this paper can effectively clarify the similarities and differences between Chinese and English political English news, detail the stylistic characteristics of political English news to the public, and provide a perspective to clearly distinguish the content and purpose of news writing, as well as the attitude and position of the media, Help readers read political English news more effectively, understand its profound meaning and criticize the advantages and disadvantages of current politics.

Keyword: Corpus · English news · Intertextuality analysis

1 Introduction

With the development of the information age, people urgently need to obtain information from all over the world, so news English has attracted more and more attention [1]. Corpus based research on News English has begun to rise in recent years. This study is based on the research of self-built radio news English corpus and newspaper news English corpus, using the combination of quantitative analysis and qualitative analysis to reveal the language characteristics of publishing news English at home and abroad. Intertextuality is a prerequisite for all text comprehension [2]. Through the analysis of intertextuality in English news discourse, we find that intertextuality plays an important textual function in English news discourse, which is mainly reflected in three aspects: confirming the discourse theme, realizing the discourse intention and establishing the discourse coherence [3]. Any discourse has a theme, which refers to the basic object of a discourse information fragment, that is, the theme of the whole discourse; For any text, theme is the representation of the main content of the text actively established by the text recipient on the basis of the content expressed in the text [4]. Understanding and using a text depends on the users' knowledge of other texts, that is, intertextuality, which is also an indispensable part of textuality [5]. The intertextuality analysis of a text can help to analyze how the ideas to be expressed in a text can be realized through the use of language, and how the ideas it expresses are related to other texts with related ideas, that is, to determine the theme of the text [6]. Discourse intention is the subjective attitude of discourse producers, which means that discourse producers purposefully create a cohesive and coherent discourse to achieve a planned goal. Only when the discourse receiver understands the intention of discourse producers can the purpose of communication be achieved, because in discourse communication, discourse producers always want their discourse to achieve his planned goal, and successful communication clearly requires the discourse receiver to be able to perceive or infer the purpose of discourse producers [7]. Similarly, the text producer must be able to predict the response of the text receiver [8].

Intertextuality analysis of discourse can help to realize discourse intention, and analyze how to express intention and whether intention is expressed [9]. Coherence refers to the way in which the components of the discourse world influence and relate to each other. As the main carrier of information and comments, news is not only a tool for transmitting information, but also an ideological battlefield for various social power groups to strive for the right of meaning interpretation, reflecting the ideology of a certain social superstructure and serving certain class interests [10]. Therefore, English news discourse will inevitably imply the author's ideology [11]. In this study, the author takes quotation, thematic intertextuality and genre intertextuality as examples to explore the relationship between the use of intertextual means in English news discourse and the embodiment of ideology. Ideology is expressed through intertextual means in news discourse. On the other hand, intertextual means in English news discourse also help to construct the existing ideology. The role of speech is very important in news discourse [12]. Generally speaking, quotes are widely used in news reports. The author often quotes the participants, witnesses, experts or some authoritative institutions of the event. The main purpose is to add a sense of authenticity to the report, show the objectivity and impartiality of the report, and do not mix the personal views of the reporter [13].

However, in writing, the author has a screening process for citations. He will only quote what he thinks credible, that is, what people in the same ideological camp say [14]. The choice of quotation is also the choice of discourse right. According to the theory of intertextuality, each discourse is the absorption and transformation of other discourses, which refer to each other and implicate each other. No English news discourse can exist in isolation from other discourses, and always has more or less intertextual connection with the surrounding discourses. This paper discusses the characteristics of intertextuality in news discourse, makes a detailed analysis of intertextuality in English news discourse from two aspects: explicit intertextuality and constructive intertextuality, and then discusses the textual function of intertextuality in English news discourse and the relationship between intertextuality and ideology in English news discourse. Analyzing English news discourse from the perspective of intertextuality provides a new method for the study of English news discourse and expands the scope and space of news discourse research. Intertextuality analysis of news discourse can also provide reference for journalists, and can make more conscious use of written means to report news more effectively. In addition, readers should learn to analyze and read news reports from a critical perspective, so as to improve their critical language awareness. This method is of great significance to the study of the generation and understanding of English news discourse, and also plays a reference role in the study of other discourses [15].

2 An Analysis of English News Discourse from the Perspective of Intertextuality

2.1 Concept of Intertextuality

Intertextuality is a concept of structuralist literary theory, which first appeared in post structuralist literary criticism. Reporting discourse, that is, quotation, is one of the important contents of intertextuality research. It refers not only to the form of dialogue between the author, the reader and the characters in the works, but also to the interaction between different texts. In the past few decades, intertextuality research has made great progress in the fields of literary criticism and cultural research. In recent years, intertextuality research has been more and more used in discourse analysis. Intertextuality is an important feature of discourse. Some linguists begin to use critical discourse analysis to interpret intertextuality in non literary texts. Intertextuality analysis realizes the production, reproduction and transformation of the existing social structure and power relations through “the production, reproduction and transformation of the existing discourse order”. In the analysis of news discourse, discourse reporting is one of the main contents of intertextuality. News reporters can not only reflect the position and views of the speaker, but also convey the reporter’s own ideology by relaying the words of others. By analyzing the reported discourse in news discourse, we can clearly see the intertextuality of reported discourse.

Reporting discourse, that is, quotation, is one of the important contents of intertextuality research. There are two kinds of reporting forms of speech, namely direct speech and indirect speech. In news reporting, reporters will use quotations to convey their ideology and have an impact on readers’ understanding of news reports. Critical discourse

analysis of the reported discourse can present the ideology of the news reporter implied behind the news report and the intervention of language in social events, so as to help readers better interpret the news report.

2.2 English News Discourse Analysis

There are many reported discourses in news discourse, that is, quotations, and reported discourses are one of the important contents of intertextuality research. In news reporting, reporters will use quotations to convey their thoughts and ideology, and have an impact on readers' interpretation of the content of news reports. Using critical discourse analysis to analyze the reported discourse in news discourse includes not only the analysis of language level, but also the interpretation and interpretation of social level and ideology. "Critical discourse analysis reveals the influence of ideology on the process of discourse generation, the reaction of discourse to ideology, and how they originate from social structure and power relationship and serve it". It is often believed that news reporters use the words of the interviewees to strengthen the authenticity of news reports. However, in fact, different treatment methods of reported words can show the ideological tendency of news reporters. The most frequently used discourse reporting methods in news discourse are direct speech and indirect speech. Direct speech refers to directly quoting the original words of others, while indirect speech refers to indirectly reporting the speaker's words. Whether direct speech or indirect speech, the discourse reporter can borrow other people's words to express his thoughts. Compared with direct speech, indirect speech provides a greater space for the speaker to "intervene" in other people's discourse. This paper will make a critical analysis of the reported discourse in news discourse from three aspects: the source of the reported discourse, the reported mode of the discourse and the reported verbs of the discourse, in order to present the real ideology of the reporter behind the news report. The source of the utterance refers to who is the originator of the original utterance of the utterance. The corresponding relationship between different semantic granularity of bilingual parallel corpus is shown in Fig. 1.

Quoting others' words directly or indirectly in news reports can improve the authenticity and objectivity of reports. However, in the use of reported discourse, "each speaker will intervene between the respondent and the reader in some way according to their respective communicative purposes, and exert influence on the reader through the reported discourse". In the report, the author can quote the words of experts and scholars to interpret news events, so as to express his own views and opinions through the words of others, and integrate his own views, others' views and objective reports. In the report, the source of some quotations is clearly pointed out, and some sources are not clearly pointed out, but can be judged by the context. We can divide the sources of reported discourse into three categories: direct and clear discourse sources, unspecified discourse sources, and discourse sources that cannot be pointed out directly. It should be pointed out that when reporting discourse, the reporter will choose different forms of quotation sources and discourse according to his own interests and ideology. The reporter has the right to choose who to use and who not to use. For the discourse of the same source, he also has the right to decide which part to use and which part not to use, which will affect the readers' understanding of news reports.

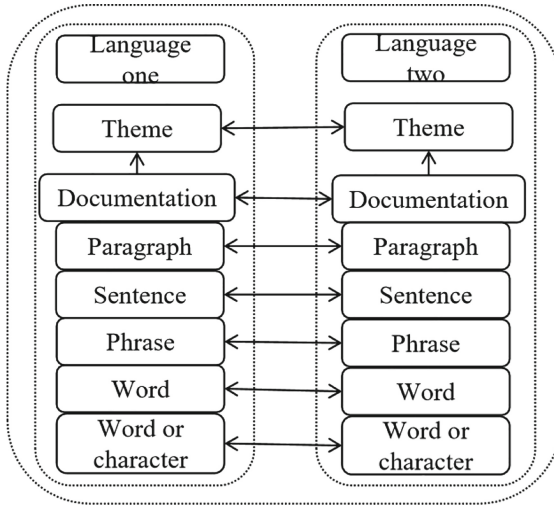


Fig. 1. Correspondence diagram of bilingual parallel corpus with different semantic granularity

3 Result Analysis and Discussion

3.1 Corpus-Based Syntactic Structure Analysis

Corpus linguistics is based on real language data, from the investigation of the distribution frequency of language information to study the laws and patterns of language in actual use. This discourse analysis is mainly based on a self-built computer corpus, which is available for computer retrieval, and the corpus is authentic. Labeling is to attach an appropriate part-of-speech tag to each word, and parsing is to syntactically analyze the sentences in the corpus, and also to semantically label the corpus to explain the semantic properties of the word or phrase. These work are mainly done through Qtag software. The establishment of a corpus is only to complete the collection, arrangement and processing of the corpus. The most important part of using a corpus is the use of retrieval tools. The analysis and statistical software such as Antwordprofiler, Antconcordance, and Spss are used for discourse research and analysis. This paper analyzes ideology in discourse from the aspects of classification and so on. The classification here is mainly realized through the choice of words. On the other hand, according to the functions of corpus and retrieval tools, combined with the characteristics of critical linguistics, it mainly monitors the frequency of information, such as the frequency of words, phrases and sentences, the collocation of new words, the change of collocation range, the frequency of tense, etc., to analyze these seemingly ordinary and unusual hidden ideological meanings. The text classification system mainly refers to the naming and description of characters and events in the text, which is mainly realized through vocabulary. It is the most basic cognitive way of human beings. Therefore, the choice of words can often reflect people's inner world and inner feelings. Of course, it can reflect people's views and attitudes towards something or a phenomenon. The following analyzes the text from the aspects of word selection, word difficulty, part of speech, collocation and co-occurrence through the self-built corpus and retrieval tools.

Let T be a tree, and its left and right subtrees are T_1, T_2 respectively, then the binding force of T is defined as:

$$D(T) = D(T_1) \times D(T_2) \times Q(R(T_1), R(T_2)) \tag{1}$$

Among them, $R(T)$ represents the syntactic attribute corresponding to T . In order to avoid underflow, take logarithms on both sides:

$$\text{Log}(D(T)) = \text{Log}(D(T_1)) + \text{Log}(D(T_2)) + \text{Log}(Q(R(T_1), R(T_2))) \tag{2}$$

Since the subtree does not include individual leaf nodes, the number of leaf nodes of each subtree must be greater than 1. If we can calculate different subtrees ($i = 2, \dots, n - 1$) $R(i)$ with $i(> 1)$ leaf nodes, we can get the total number as follows:

$$P(n) = \sum_{i=2}^{n-1} R(i) \tag{3}$$

For $R(i)$, because the leaf nodes are orderly, the i node numbers it includes can be taken. The general process of cross-language text clustering under the overall framework is shown in Fig. 2.

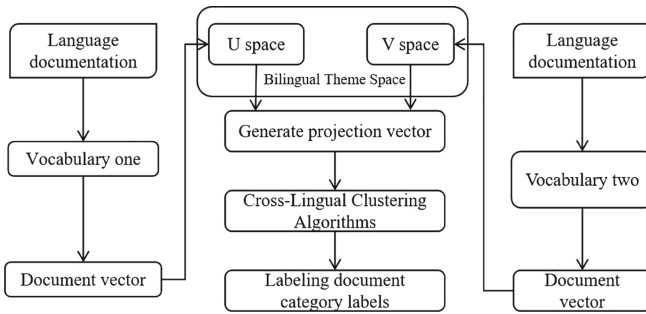


Fig. 2. General process of cross language text clustering under the overall framework

Optimize the English news corpus. Theoretically, as long as we get all the corpus of a certain language, we can show that the corpus is completely representative. However, from the actual development, corpus is only a collection of representative language samples. In the actual development process, the English news corpus should be optimized and perfected to improve the coverage of the corpus. English news has a large vocabulary and information, so in the process of translation, the language should be more concise. At the same time, English news should be made clear in translation, and the least number of words should be used to express the most comprehensive information. Improve the scale and collocation of thesaurus. The large-scale general corpus will not only cover the basic statistical results of word frequency, but also make statistical analysis on the word frequency distribution of different genres. In the actual translation work, it is necessary to upgrade the sub-database scale in the corpus, improve the collocation between different words and optimize the usage environment of words. By this method, English news can be translated with the simplest and most capital morphemes, the quality of translation can be improved, and the acceptance difficulty of the audience can be reduced.

3.2 Translation of News English Headlines from the Perspective of Intertextuality

Translation studies from the perspective of intertextuality have rich connotations. It not only refers to the simple transformation between languages or texts, but also involves a broader intertextuality movement “field”, including language, text, culture, translation subject, external social and historical context, and even those who are involved in it. This makes translation a polyphonic dialogue activity, in which the translator plays a vital role. Then proceed to the original text. Second writing “,that is, the reconstruction process of the translated text; In this process of intertextuality translation or “rewriting”, translators are required to have rich knowledge of intertextuality not only in the original text language but also in the target text. If the knowledge of intertextuality in the original text is lacking or not fully understood, it will be difficult to make enough interpretation. Obviously, the translation process from the perspective of intertextuality is the process of deconstructing the original text and reconstructing the translated text. Next, through this intertextuality, the translation of news English headlines is tentatively explained. The general process of cross-language retrieval under the overall framework is shown in Fig. 3.

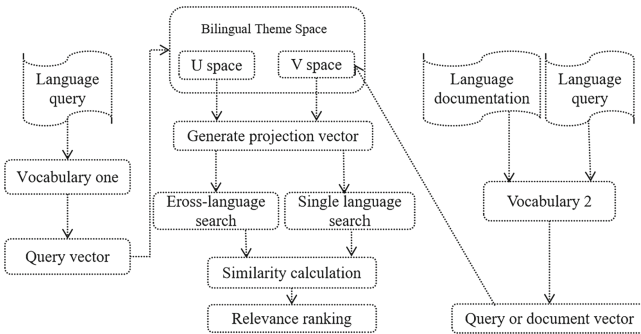


Fig. 3. The general process of cross-language retrieval under the overall framework

News English headlines have their own characteristics not only in rhetoric, but also in tense, voice, vocabulary, punctuation and even arrangement. “Appropriate” also needs the help of intertextual manipulation. First of all, we must understand and identify the characteristics of news English headlines in tense, voice, vocabulary and punctuation, and make appropriate reference in translation operation: in tense, news English has obvious signs, uses more general present tense, continuous tense and future tense, and uses less general past tense. This tense is called news present tense in journalism. Punctuation, ellipsis, interval and parentheses are rarely used in English titles, mostly colons, slashes and possessive signs; Omitting function words is another feature of the title. Omitting all meaningless words can greatly shorten the length of the title, be concise and comprehensive, and highlight the core information. The words that can be omitted in the title include articles, prepositions, conjunctions, possessive pronouns, Copula be, etc. After having an understanding of the intertextuality of news English headlines, the translator should refer to the application of similar Chinese news headlines in tense, voice, words

and punctuation, which is an intertextual process between the translator and the target language. Because of the change of translation service objects, the intertextual reference of news Chinese headlines is particularly important. There is no obvious feature in the tense of Chinese English titles, which can be used in general present tense, general future tense, continuous tense, past tense, etc. However, due to the features of English titles, there is often a change in the tense in translation. Colons, slash marks and possessive case marks are less used in Chinese translation of punctuation, and ellipses, dashes and quotation marks are more used. Only when the translator has an intertextual understanding of English and Chinese news headlines can he ensure their accuracy and efficiency in translation.

4 Conclusions

Based on the research theory of systemic functional linguistics on genre, this paper focuses on the contextual configuration of English economic news, and then summarizes the potential of its genre structure. Deliver information in the best way to help readers correctly understand the substance of the news and better grasp the way to read the news. Corpus is of great significance for English translation work, especially news English translation work. Staff in related fields should pay attention to imparting scientific translation methods to students in the teaching and management of actual news English, and at the same time pay attention to the updating and upgrading of news English corpus. Through this method, students can be supplemented with more news English vocabulary, improve the accuracy and fluency of translation, and optimize the connection and conversion effect between the source language and the target news English language.

Acknowledgements. Jingchu University of Technology Fund Projects: Comparative Study of Sino-US Leaders' Political Discourses Under the Framework of the United Nations (No: YB202112) AND Philosophy and Social Science Research Project of Education Department of Hubei Province: Comparative Study of Sino-US Leaders' Political Discourses Under the Framework of the United Nations (No: 21Y261).

References

1. Walter, S., Christopher, F., Neil, C.: The sense of a connection: Automatic tracing of intertextuality by meaning. *Digital Scholarship in the Humanities*, no. 1, p. 1 (2016)
2. Lombardi, A.: More is more: explicit intertextuality in university writing placement exam essays. *J. Engl. Acad. Purp.* **50**(3), 100955 (2020)
3. Lu, W.: Intertextuality and palimpsestuousness between different texts: marilyn chin's poem "turtle soup" as a case study. *Foreign Literat. Stud.* **40**(6), 31–41 (2018)
4. Luka, M.: Grassroots prescriptivism an analysis of individual speakers' efforts at maintaining the standard language ideology. *English Today* **34**(4), 5–12 (2018)
5. Khan, S.I., Bilal, A.R., AhMaD, B.: Who will land and stay? Page-specific antecedents of news engagement on social media. *Online Inf. Rev.* **44**(5), 1013–1025 (2020)

6. Quan, A., Wilson, L.A., Mithani, S.S. et al.: Reporting on the opioid crisis (2000–2018): role of The Globe and Mail, a Canadian English-language newspaper in influencing public opinion. *Harm Reduction Journal*, no. 17, p. 1 (2020)
7. Boholm, M.: Risk and quantification: a linguistic study. *Risk Anal.* **39**(6), 1243–1261 (2019)
8. Ev?En, K., Michala, M.: Intraday effect of news on emerging European forex markets: An event study analysis. *Econ. Syst.* **42**(4), 597–615 (2018)
9. Janke, K., Propper, C., Shields, M.A.: Assaults, murders and walkers: the impact of violent crime on physical activity. *J. Health Econ.* **47**, 34–49 (2016)
10. Ortega, J.L.: How do media mention research papers? Structural analysis of blogs and news networks using citation coupling. *J. Informet.* **15**(3), 101175 (2021)
11. Ahmed, W., Lugovic, S.: Social media analytics: analysis and visualisation of news diffusion using NodeXL. *Online Inf. Rev.* **43**(1), 149–160 (2019)
12. Ortega, J.L.: Blogs and news sources coverage in altmetrics data providers: a comparative analysis by country, language, and subject. *Scientometrics* **122**(1), 555–572 (2020)
13. Rahnemoon, S.N., Ahangar, A.A., Nourmohammadi, E.: Thematic structure and translation: a case study of the translation of English news into Persian. *Lingua* **194**, 26–50 (2017)
14. Sobel, M.R.: Confronting sex trafficking: gender depictions in newspaper coverage from the Former Soviet Republics and the Baltic states. *Eur. J. Commun.* **31**(2), 52–168 (2016)
15. Thompson, C.G., Gillern, S.V.: Video-game based instruction for vocabulary acquisition with English language learners: a Bayesian meta-analysis. *Educ. Res. Rev.* **30**, 100332 (2020)