

Reading Between the Lines: Evaluating GenAI's Detection of Attitudinal Meaning in News Discourse

Hong Wu

School of Foreign Languages, Guangxi University, Nanning, China
nnlilywu@gmail.com

Abstract. This study investigates the potential and limitations of ChatGPT in identifying attitudinal resources in news discourse. Fifteen news articles were analyzed by ChatGPT, and the results generated by it were compared with manual annotations obtained in a previous study. The findings show that ChatGPT is capable of identifying attitudinal resources and generally captures the overall evaluative orientation of the texts. Compared with human annotators, however, ChatGPT tends to identify fewer positive resources but more neutral ones. The analysis also reveals several limitations of ChatGPT in attitude identification. It tends to over mark a particular type of attitudinal resources within a single discourse; moreover, its responses are not always stable, as the identification and interpretation process may be continuously revised when it is challenged through subsequent prompts. These findings suggest that, although ChatGPT has considerable potential as an auxiliary tool for discourse analysis and may improve the efficiency of large-scale attitudinal analysis, its analytical results require careful human validation. At present, it should be regarded only as a complementary tool rather than a substitute for human in discourse research.

Keywords: News, Discourse analysis, Attitudinal resources, GenAI, ChatGPT

1. Introduction

Opinions play a vital role in people's decision-making process. Understanding people's opinion, that is, people's sentiment/stance/attitude toward a particular topic/object/event/entity, helps to explain how and why decisions are made. Therefore, the identification of attitudes in discourse has been an area of interest across disciplines such as linguistics, sociology and political science, etc. Appraisal Theory [1], which derived within the framework of Systemic Functional Linguistics, offers a powerful framework for examining evaluative meaning, helping researchers to discover how speakers and writers express emotions, make judgments, and appreciate the world around them. The detection of attitude, however, is a complex task due to the highly subjective nature of language. Even trained annotators may disagree on annotations. Besides, the annotation process is labor-intensive and time-consuming, limiting its applicability to large datasets.

With the advancement of computer science, sentiment analysis, which relies on lexicon-based or machine learning techniques, began to boom since early 2000 as more opinionated digital data available [2]. Automated sentiment analysis offers faster processing of large datasets, but the results

are restricted to the overall sentiment of a text. The attitude towards certain entity within a text, however, remained unknown [3].

The limitations of both manual annotation and automated sentiment analysis raise a critical question: is there a way to identify attitudes both quickly and accurately? Recent advances in generative AI (GenAI), particularly large language models (LLMs) have demonstrated the ability to recognize context-dependent meaning and to generate coherent, context-aware responses, which opened new possibilities for automated discourse analysis. Whether GenAI can be trusted to fulfill the task became a domain which has attracted much research interest. Its accuracy and reliability in identifying attitudes, nevertheless, especially when compared to human annotation and automated methods, remain underexplored. The present study seeks to address this gap by systematically evaluating GenAI, ChatGPT to be specific, in its performance in attitude identification. Specifically, it aims to compare the results obtained from manual annotation and ChatGPT, thereby shedding light on whether GenAI can complement or even challenge established methods in this field. It addresses the following research questions:

- (1) How accurately can ChatGPT identify attitudinal resources compared with human annotation?
- (2) What implications do these findings have for integrating ChatGPT into discourse analysis?

By examining these questions, the study aims to investigate how GenAI-assisted annotation can complement human analysis. The results are expected to shed light on the potential and limitations of using GenAI as a tool for evaluative language research, thereby contributing to discourse analysis.

This paper is organized as follows: Section 2 reviews relevant literature on Appraisal Theory and application of GenAI in discourse analysis. Section 3 details the methodology, including dataset selection and the prompt used. Section 4 presents the results, comparing ChatGPT's performance with human annotation, and discusses the findings, highlighting ChatGPT's strengths, limitations, and implications for automated discourse analysis. Finally, Section 5 concludes the study and outlines directions for future research.

2. Literature review

In this section, literature concerning Appraisal Theory, as well as previous research on applications of GenAI in discourse analysis are reviewed.

2.1. Appraisal Theory

Discourse analysis is the analysis of language-in-use, which focuses on how "language functions as a system of human communication" [4]. Appraisal Theory, which is developed from the interpersonal metafunction of Systemic Functional Linguistics (SFL) by a group of researchers led by J. R. Martin and P. R. R. White, is widely used as a guiding framework in identifying attitudes in discourse. The analytical framework of Appraisal Theory is to examine people's evaluation from three subsystems, namely, Attitude, Engagement and Graduation [1]. Attitude is concerned with feelings expressed in the texts [1]; Engagement is about the source of attitude and how the evaluation is made [1]; Graduation deals with the strength and clarity of the attitudes [1]. The study focuses on whether ChatGPT could identify attitudes in the text, while Engagement and Graduation are not the concern. Therefore, only the subsystem of Attitude will be examined.

According Martin and White [1], the subsystem of Attitude can be further divided into three sub-categories, Affect, Judgement and Appreciation. Affect is about emotions, whether the feeling is positive or negative; Judgement is about ethics, that is, judgement towards people's behaviour; and

Appreciation is about aesthetics, evaluating non-human entities or human-created products. Each of these sub-category can be further divided into several types.

The appraisal framework has been applied widely in discourses. Hasbi et al. [5] used the appraisal framework to analyze news reports on elections in Malaysia and found that Judgement was the most prevalent linguistic resource among all sub-categories. There are also research focusing on the subsystem of Attitude. Grundlingh [6] chose news articles on 4 murders and found that Judgement was the most frequent sub-category used, followed by Appreciation. Puspita and Pranoto [7] and Kazmi et al. [8] also discovered a similar pattern. These studies have manifested the applicability of Appraisal Theory in discourse analysis.

2.2. Applications of GenAI

The most important feature of GenAI is its generative nature. Leveraging deep learning and LLMs, GenAI can process contextual information in texts and produce new texts that closely resemble human language use. When OpenAI officially launched ChatGPT, the GenAI chatbox, to the public on November 30, 2022, people were marveled by what it could do, and research on and research using GenAI continued to emerge [9].

Fan et al. [10] tested ChatGPT's capabilities in topic segmentation and discourse parsing, and found that it could understand topic structure in general-domain conversations better than specific-domain conversations, but did poorly in understanding rhetorical structures. Curry et al. [11] examined the applicability of ChatGPT 4 in automated discourse analysis and found that ChatGPT could categorize keywords semantically, but could not gain satisfactory results in concordance analysis or function-to-form analysis. Gao and Feng [12] used LLMs to analyze media attitudes and found that the quantized LLM can capture 80% of attitudes correctly. Their research provided a research foundation for future work.

Previous studies have proven the feasibility of using GenAI in discourse analysis, but the capabilities of ChatGPT in identifying attitudinal resources has not been fully explored. This study aims to address this gap and see how well ChatGPT can identify attitudes in news discourse.

3. Methodology

The researcher selected 15 news articles from previous research [13], in which sentiment analysis and human annotation were both applied to 120 news texts and the results were compared. Among the 120 texts, only 59 of them were consistent in polarity results by both automated sentiment analysis and human annotation. The selected texts fell into the following three categories:

Category 1: 5 with the lowest negative sentiment from news texts with inconsistent results. The lowest sentiment scores meant that they were regarded as highly negative by automated sentiment analysis. However, these highly negative texts were found with more positive attitudinal resources by the human annotator.

Category 2: 5 with the most positive attitudinal resources from news texts with inconsistent results. The more positive attitudinal resources the human annotator uncovered, the more positive the news text should be.

Category 3: 5 with the most negative attitudinal resources from news texts with consistent results.

In short, Categories 1 and 2 showed a strong contrast of the differences in identifying positive attitude by AML's sentiment analysis and a human annotator. They were chosen to examine whether the human annotator was too subjective or AML's sentiment analysis did tend to highlight negativity. Category 3 was chosen to examine whether ChatGPT could also identify negative attitudes correctly.

All of these 15 selected news articles were given to ChatGPT for analysis with the following prompt:

You're a linguist researching under the theoretical guidance of Appraisal Theory. Please identify all the attitudinal resources about China in the article, and decide whether these resources are positive, negative or neutral towards China.

ChatGPT responded with answers in the following pattern:

Not only does he pump in additional funds to help Belt and Road nations, he has offered to share China's industrialisation experience with the world.

Type: Judgment (Propriety)

Sentiment: Positive

4. Findings and implications

With all the answers provided by ChatGPT, the researcher compared the results from ChatGPT with those of manual annotation, as is shown in Table 1. It could be observed that though the number of different attitudinal types identified by the human annotator and ChatGPT were not identical, the general judgement of whether the text was positive or negative was consistent. That is to say, ChatGPT did identify polarity in the news texts, even attitude towards China, which showed that ChatGPT is able to assist researchers in discourse analysis.

Table 1. Counts of attitudinal types: manual annotation vs. ChatGPT

Category	News text	Manual annotation			ChatGPT		
		Positive	Negative	Neutral	Positive	Negative	Neutral
1	201310	8	5	2	8	6	6
	201401	12	1	5	11	8	10
	201608	45	5	1	28	8	4
	201611	25	7	0	6	6	1
	201906	25	15	2	19	0	1
2	201512	59	1	1	29	1	1
	201604	46	1	1	63	2	17
	201705	36	2	2	16	1	2
	201711	38	0	0	38	0	1
	202105	39	4	3	21	3	0
3	201307	3	29	2	7	19	1
	201308	14	29	2	6	52	31
	201507	8	35	2	0	9	2
	201603	7	24	1	4	35	10
	201808	3	26	2	0	21	4
Total		368	184	26	256	171	91

4.1. Findings

Overall, the human annotator identified 578 attitudinal resources, while ChatGPT identified 518 ones, approximately 10.4% fewer resources. The results indicate that ChatGPT broadly captured the polarity identified by human annotators, but it differed greatly in the identification of positive and neutral attitude resources.

First, both methods identified positive attitude as the dominant sentiment. Manual annotation identified 368 positive instances, accounting for 63.67% of all attitude resources, while ChatGPT identified 256, taking up only 49.42%. Although ChatGPT detected fewer positive resources, positive attitudes overweighed the other two. This suggests that ChatGPT was generally able to recognize the overall positive orientation of the news texts, unlike sentiment analysis, but it appeared to be more conservative than human annotators in identifying positive evaluative meanings.

Second, ChatGPT showed the greatest consistency with human annotators in the identification of negative attitudes. The human annotator identified 184 negative resources (31.83%), whereas ChatGPT got 171 (33.01%), only 13 instances fewer. More importantly, the proportional distribution of negative resources is almost identical. This suggests that ChatGPT demonstrates relatively strong reliability in recognizing explicitly or semantically salient negative evaluations.

Third, the most noticeable discrepancy concerns neutral resources. The human annotator only identified 26 neutral resources, while ChatGPT identified 91. This suggests that ChatGPT tends to classify more neutral resources than human annotators. In other words, ChatGPT may have difficulties recognizing more complex or subtle attitudinal resources.

In summary, ChatGPT performed better in the identification of negative attitudes, whereas it under-identified positive resources and substantially over-identified neutral ones.

4.2. Implications

Though high consistency between the human annotator and ChatGPT could be found in the judgement of polarity of the 15 selected news texts. The use of ChatGPT should not be without any restraint. In examining the results generated by ChatGPT, the researcher found two disadvantages of ChatGPT that need to be addressed.

The first is that ChatGPT could not identify the attitudinal types correctly and tended to over mark a certain attitudinal type in one article. For example:

(1) *THE business community --- at least among Chinese businessmen --- still cannot get over the euphoria created by Premier Li Keqiang, who recently announced a string of measures that have improved sentiment for the financial market.* (Dec., 2015)

(2) *China is willing to help Malaysia.* (Dec., 2015)

Excerpts (1) and (2) were both taken from the same news article. In Excerpt (1), the phrase *cannot get over the euphoria* suggested that businessmen in Malaysia were happy to learned about the measures announced by the Chinese Premier, which had improved sentiment in finance; therefore, the attitudinal resource marked by the human annotator was happiness. Excerpt (2) indicated clearly China's intention of extending a hand to Malaysia to maintain the close collaborative relations; therefore, it was marked as inclination. ChatGPT marked both excerpts with attitudinal resources of capacity, which obviously were inappropriate. In the same news article, among the 31 attitudinal resources identified by ChatGPT, 24 of them were marked as capacity, while the human annotator only marked 2 out of the 61 as capacity. The same situation could also be found in other news articles.

The second is that the results generated by ChatGPT are always changing and should be used with care. ChatGPT is such a powerful tool as it keeps learning and improving from users' feedback; therefore, each time a prompt is given, it may adjust its response. Takes the news article from May 2021 as an example. The first time ChatGPT was given the instruction to analyse the article for its attitudinal resources towards China, it generated a purely positive analysis without any negative or neutral instances. During manual annotation, however, the human annotator did notice 4 negative attitudinal resources and 3 neutral ones. Therefore, the researcher asked ChatGPT whether there were any negative or neutral attitudinal resources in the article, and got the response that the majority of the attitudinal resources towards China were positive, but there were a few subtle expressions that could be interpreted as either neutral or negative. Then ChatGPT supplemented 3 sentences with neutral/slightly negative sentiment.

(3) *The US government, for instance, has put dozens of Chinese tech companies (including Huawei Technologies and Semiconductor Manufacturing International Corp) on an "entity list", which restricts their access to advance US technologies, such as crucial software and semiconductor equipment.* (May, 2021)

Excerpt (3) was among the three. In manual annotation, the sentence was marked as negative as the restrictions imposed by the US government showed its hostility against Chinese tech companies, which was antipathy under happiness. From the example, we could see that ChatGPT is able to identify such an implicit attitude, but it needs feedback from the user. Besides, ChatGPT again categorized the attitudinal type as capacity, which was inappropriate.

5. Conclusion

The present experiment demonstrates the potential of ChatGPT in assisting attitudinal analysis. Overall, ChatGPT was able to capture the broad evaluative orientation of news discourse. However, its performance was less satisfactory when it came to the precise categorization of specific attitudinal resources. This indicates that, despite its analytical capabilities, ChatGPT cannot yet be regarded as a substitute for human in linguistic research. Rather, it should be viewed as an auxiliary tool whose outputs require careful examination and validation. Therefore, at its current stage of development, ChatGPT is best positioned only as a complementary research assistant.

It should be noted that this paper only compares ChatGPT's results with human annotation in a simple polarity count, lacking fine-grained comparison of sub-categories. For future search on ChatGPT's performance in identification of attitudes, the comparison should be extended to the sub-categories.

Acknowledgement

This paper is funded by the Key Research Base of Humanities and Social Sciences of Universities in Guangxi Zhuang Autonomous Region, Grant/Award Number: 2026APTISYB01; Asia Pacific (Southeast Asia) Translation and Cross-Cultural Communication Center.

References

- [1] Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.
- [2] Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
<https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- [3] Wu, H., Bakar, K. A., Jaludin, A., & Mat Awal, N. (2022). Sentiment analysis of China-related news in The Star Online newspaper. *GEMA Online Journal of Language Studies*, 22(3), 155–175. <http://doi.org/10.17576/gema->

2022-2203-09

- [4] Thompson, G. (2004). *Introducing functional grammar*. <https://doi.org/10.4324/9780203431474>
- [5] Hasbi, N., Ting, S. H., & Jerome, C. (2024). Look beyond the facts: Evaluative language use in Malaysia's six-state election news coverage. *Asian Journal of Research in Education and Social Sciences*, 6(1), 764–778.
- [6] Grundlingh, L. (2017). Identifying markers of sensationalism in online news reports on crime. *Language Matters*, 48(2), 11–136.
- [7] Puspita, D., & Pranoto, B. E. (2021). The attitude of Japanese newspapers in narrating disaster events: Appraisal in critical discourse study. *Studies in English Language and Education*, 8(2), 796-817.
- [8] Kazmi, S., Mohd Noor, S. N. F., & Sanif, S. (2023). Social actors' representational role in political hard news reporting: An approach through attitude analysis. *European Chemical Bulletin*, 12, 271-285.
- [9] Sobiech-Buzala, M. (2023). Will Chat GPT take our jobs?: Discourse analysis on generative AI from the moral panic perspective. *Biuletyn Naukowy Wrocławskiej Wyższej Szkoły Informatyki Stosowanej. Informatyka*, 10, 46-59.
- [10] Fan, Y., Jiang, F., Li, P., & Li, H. (2024, May). Uncovering the potential of chatgpt for discourse analysis in dialogue: An empirical study. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 16998-17010.
- [11] Curry, N., Baker, P., & Brookes, G. (2024). Generative AI for corpus approaches to discourse studies: A critical evaluation of ChatGPT. *Applied Corpus Linguistics*, 4(1), 100082.
- [12] Gao, Q., & Feng, D. (2025). Deploying large language models for discourse studies: An exploration of automated analysis of media attitudes. *PloS One*, 20(1), e0313932.
- [13] Wu, H. (2026). China from the lenses of Malaysian news media: A corpus-driven analysis of sentiment and attitude. *PhD Thesis. Universiti Kebangsaan Malaysia*.