

A Study on the Linguistic Features and Brain Cognitive Response Differences of Human and AI Texts

Hanhui Sheng*

School of Languages, Shanghai University of International Business and Economics, Shanghai 201620, China

**Corresponding author: Hanhui Sheng.*

Abstract

After the rapid development of generative large language models, the boundary involved in human and AI texts has become increasingly blurred, and academic integrity and intellectual property issues therefore need to be reconsidered. At present, many studies focus on automatic identification methods from the perspective of surface linguistic features, while they give limited explanation of the deep cognitive and neural phenomena involved when humans process the two types of texts. The present study combines computational linguistics, behavioral experiments, and neurocognition for examination, and makes a detailed analysis with electroencephalography (EEG) and eye-movement synchronization as the basic approach. The paper first summarizes the stable differences between human and AI texts in syntax, semantics, pragmatics, and other aspects, then discusses behavioral or eye-movement performance according to processing depth. It then interprets the N400, P600, and late positive component observed in fixation-related potentials as neural markers, and finally considers the relationship between factors such as genre and current research. The paper analyzes the key problem of the absence of whole-chain evidence for “linguistic features - neural responses,” and also proposes future directions such as sample expansion and cognition-guided detection. The framework proposed here has theoretical significance for AI text identification and the naturalness of large models.

Keywords

AI-generated text, EEG, eye tracking, linguistic cognitive processing, fixation-related potentials

1. Introduction

With regard to generative artificial intelligence, the large language models currently involved have reached a state of high fluency and basically reasonable logic when generating texts. They are therefore suitable for various kinds of knowledge writing or information communication, and can clearly improve the efficiency of information production. However, the gradual disappearance of the boundary between human and AI texts also raises new problems for academic integrity, content control, and knowledge attribution. How to clearly distinguish the two and their nature, and how to define them, is a basic and difficult issue jointly concerned by language research, cognitive analysis, and information management.

The research now called human-AI text discrimination is basically a combination of two parallel approaches. The first type belongs to computational linguistic methods, which use surface information such as syntax, semantics, or style to construct automatic judgment models. There have already been relatively successful classification cases on relevant data, but the “textual features - classification results” discussed are limited to

external association [1] and cannot explain the actual cognitive causes behind the differences. The second type is a small number of neuroimaging studies, which have preliminarily pointed out that the human brain has identifiable electroencephalography (EEG) features when processing AI-generated content [2], but it still cannot correspond potential changes to specific texts and cognitive processes one by one. Overall, the evidence involved in the proposition of “textual linguistic features - reading behavioral performance - brain neural responses” is still incomplete, and there is still no comprehensive discussion of the source of cognitive differences between human and AI texts.

Following the above line of thought, the present paper organizes existing studies that combine relevant linguistic facts, behavioral and eye-movement phenomena, and EEG neural mechanisms, with a focus on the applicability of eye movement-EEG synchronization and fixation-related potentials (FRPs) to natural reading. It summarizes the current empirical materials and explains the main unresolved problems. The study intends to analyze the similarities and differences between humans and machines in text processing from a cognitive neuroscience perspective, and also applies this approach to future experiments and discussions on AI-related text identification, so as to provide corresponding reference conclusions.

2. Linguistic FEATURE Differences Between Human and AI Texts

The systematic difference in linguistic features is the fundamental basis of the cognitive processing differences between human and AI texts. Existing studies have identified stable feature differences between the two types of texts, from the surface lexical-syntactic level to the deeper semantic-pragmatic level.

2.1 Surface Lexical and Syntactic Features

Early studies on AI text discrimination basically took surface language as the starting point. The known corresponding conclusion is that, compared with human-written texts, AI-generated texts often use more regular sentence patterns, have more even sentence lengths, and generally show lower syntactic complexity. In terms of vocabulary, the lexical items in AI texts are mostly high-frequency words, and the distribution of word frequency is relatively concentrated; human texts contain more low-frequency words and non-fixed collocations. From the logic of generation, it can be seen that large language models decide the next word according to probabilistic prediction and prefer expressions with high predictability. As a result, they often finally form a templated and standardized linguistic appearance.

2.2 Deep Semantic and Pragmatic Features

Current research has already led feature analysis toward the two perspectives of semantics and pragmatics. Examinations of actual materials show that the semantic entropy and syntactic variation of human texts are much higher than those of AI-generated texts [2]. In general, human writing contains more unexpected semantic changes, personalized pragmatic phenomena, and emotional intensity, while AI texts show overly even semantic transitions, a basically neutral emotion, and a lack of special pragmatic forms. In addition, when AI-generated content encounters complex issues, it may misconstrue logic or fabricate facts, but such errors are often covered by surface coherence and are not intuitively visible defects.

Taking prose fragments on the same topic as an example, the difference between the two can be seen directly. Human-written texts often present expressions such as “the wind rolled the scent of osmanthus into the window ledge; most of the tea in the cup had cooled before the afternoon letter still pressed under the inkstone was remembered,” which are jump-like, contain personalized details and semantic turns, and vary in sentence length. The corresponding AI-generated text is often like “the autumn breeze carried the fragrance of osmanthus into the window; the tea on the table gradually cooled, reminding people that there were unfinished letters to reply to.” This type of expression is smooth, coherent, and regular in sentence form, with highly predictable semantic transitions and obviously lower personalized detail and syntactic variation.

3. Research Methods

The methodological system in this field has been gradually constructed around a three-level evidence chain of “textual features - behavioral performance - neural responses.” Existing studies mainly use five methods: computational linguistic text analysis, behavioral discrimination experiments, eye tracking, EEG recording,

and multimodal synchronization. These methods involve different observation levels and complement one another.

3.1 Quantification of Computational Linguistic Textual Features

Textual feature quantification is the material basis for subsequent behavioral and neural experiments. Existing studies generally first calculate multidimensional linguistic indicators for the two types of human and AI texts. Surface indicators include sentence-length distribution, syntactic complexity, word-frequency distribution, and lexical richness, while deeper indicators include semantic entropy, syntactic variation, emotional density, and the frequency of pragmatic markers. In terms of the computational process, texts are generally first segmented, part-of-speech tagged, and analyzed for dependency syntax; then each quantitative indicator is extracted on this basis. Some studies additionally introduce measures based on large language models, such as perplexity and semantic coherence, to characterize the degree of predictability of the text. These quantitative results can be used not only to describe feature differences between human and AI texts, but also as independent variables explaining changes in behavioral and neural indicators in subsequent correlation analysis.

3.2 Behavioral Discrimination Experiments

Behavioral experiments mainly adopt a forced-choice discrimination paradigm to examine the subjective ability of people to distinguish human and AI texts. In experiments, participants are generally randomly presented with human-written text fragments and AI-generated text fragments, and are required to complete binary classification judgments within a limited time. Discrimination accuracy and reaction time are recorded. In terms of variable control, existing studies usually match the topic, genre, and length of the two types of texts, and control participants' age, educational background, and AI-use experience. Some studies additionally include confidence ratings or reason-recall procedures to explore the types of bases for discrimination. The limitation of behavioral experiments is that they can only reflect final discrimination results and are difficult to use to distinguish whether differences come from syntactic processing, semantic integration, or pragmatic perception. Therefore, they need to be combined with eye-movement and EEG methods to analyze the internal process.

3.3 Eye-Tracking Technology

Eye tracking captures online processing behavior during natural reading with high temporal resolution, and is the intermediate observation link connecting textual features and neural responses. Experiments generally adopt a natural reading paradigm and do not limit participants' reading speed, so as to approximate real reading scenes. The core observation indicators include first fixation duration, total fixation duration, number of regressions, and regression proportion. Some studies also record finer indicators such as saccade amplitude and fixation-position distribution. In terms of data processing, trials with blinking, tracking loss, and deviation from the area of interest are generally removed first, and then statistics are aggregated by target words or areas of interest. The cognitive-load differences reflected by eye-movement indicators can convert linguistic feature differences into observable behavioral indicators, providing behavioral evidence for subsequent analysis of neural mechanisms.

3.4 EEG Recording and Fixation-Related Potential Analysis

EEG recording is used to obtain neural electrical activity in the brain during reading. Traditional event-related potentials (ERP) studies mostly use the rapid serial visual presentation paradigm, present stimuli word by word, and take the moment of word appearance as the waveform zero point, but this paradigm differs greatly from real continuous text reading. To solve this problem, natural reading studies often use synchronous recording of eye movements and EEG and introduce the analysis method of FRPs. The FRP method takes the moment when the first fixation on the target word occurs as the time zero point of the EEG waveform, and processes the data according to the idea of "first-fixation window separation + separate superposition of regressions." It separately superposes the EEG waveform induced by the first fixation and the waveform induced by regression rereading, so as to avoid signal-overlap interference caused by saccades and regressions. Later analysis generally extracts the mean amplitude and peak latency of N400, P600, and the late positive

component (LPC). Some studies also analyze EEG power changes in the α , β , and γ frequency bands to quantify the cognitive arousal and processing intensity of the brain when reading human and AI texts.

3.5 Multi-Source Data Association and Statistical Analysis

Multi-source data association analysis aims to connect the complete transmission chain of “linguistic features - eye-movement behavior - neural responses.” Existing studies generally conduct joint statistics on text quantification indicators, eye-movement behavioral indicators, and EEG ERP component indicators, and use methods such as correlation analysis, regression modeling, or mediation-effect testing to examine the associations and transmission paths among the three. In terms of moderating variables, factors such as text genre, participant background, and model parameters are generally included, and hierarchical regression or mixed-effects models are used to test the stability of effects under different conditions. These analytical methods help clarify the formation mechanism and boundary constraints of cognitive differences between human and AI texts, but at present the complete quantitative association chain has not yet been fully established and remains a key direction for future research in this field.

4. Behavioral and Eye-Movement Evidence for Text-Processing Differences

Differences in linguistic features will ultimately appear in readers’ online processing and final discrimination performance. Behavioral and eye-movement studies provide direct evidence for differences in the processing depth of human and AI texts.

4.1 Behavioral Performance in Text Discrimination

Behavioral literature has used forced-choice discrimination methods to examine people’s ability to distinguish AI-written texts. From the report by Clark et al. [3], it can be seen that when participants judged AI-generated news or story texts, the accuracy they obtained was generally not very high, and there were obvious differences among domains. Current studies generally agree on one point: in terms of text genre, discrimination accuracy varies greatly; narrative or prose-like texts are easier to identify, while correct identification of academic texts is relatively difficult. These differences can basically be attributed to the high degree of standardization in academic language, which limits the freedom of individual expression and makes human and machine expressions close to the same style. In addition, readers’ habits of daily AI use, the domains involved, and their familiarity with the texts can also affect discrimination results. However, the accuracy and reaction time obtained behaviorally cannot fully explain the details of the cognitive mechanism, and cannot determine whether semantics or syntax takes effect first.

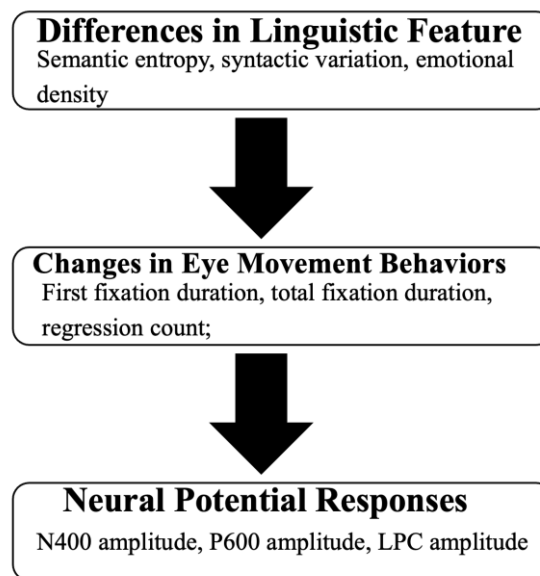
4.2 Eye-Movement Patterns in Natural Reading

Eye-tracking technology can observe actual reading behavior with high temporal resolution, and the data obtained are online indicators of cognitive processing load. Typical eye-movement indicators include first fixation duration, total fixation time, number of regressions, and regression proportion. Existing studies have pointed out that when people read texts written by other people, both first fixation duration and overall fixation time are longer, regressions are more frequent, and the regression proportion is higher. These indicators all suggest that the human texts being read require deeper thinking and more semantic integration. By contrast, AI-produced texts are smoother to read, with shorter fixations and fewer regressions; that is, the cognitive burden is smaller and the result is more predictable. Relevant statistics also prove that eye-movement indicators have a clear relationship with the semantic entropy and syntactic complexity of language, and that linguistic differences affect reading patterns. Combining the four levels of linguistic features, eye-movement performance, neural responses, and discrimination ability, the multidimensional core differences between human and AI texts can be summarized as Table 1.

The above differences in Table 1 do not exist independently, but form a complete mechanism of action along the path of “linguistic feature input - cognitive processing transmission - neural potential output.” The transmission mechanism of cognitive processing differences between human and AI texts is shown in Figure 1. Differences in semantic entropy, syntactic variation, and emotional density at the linguistic level are transmitted to the eye-movement behavioral level through changes in cognitive processing depth, and finally appear as systematic differences in neural potentials.

Table 1: Summary of Multidimensional Core Differences Between Human and AI Texts

Dimension	Human-Written Text	Corresponding AI Text
Linguistic features	Semantically more complex; greater syntactic variation; stronger emotion; irregular pragmatic usage	More unified sentence patterns; smooth semantic connection; basically neutral tone; frequent templated expression
Eye-movement performance	Longer fixation time; more regressions; heavier information processing load	Shorter fixation time; fewer regressions; smoother reading
Neural responses	Larger amplitudes of N400, P600, and LPC; strong cognitive arousal	Smaller ERP amplitudes; shallower cognitive investment
Discrimination ability	Narrative genres are easy to identify, and judgments are relatively reliable	Academic genres are difficult to distinguish, and accuracy is not high

Figure 1: Schematic Diagram of the Transmission Mechanism of Cognitive Processing Differences Between Human and AI Texts

Human-written texts and the corresponding AI texts differ as follows. Linguistic features: human-written texts are semantically more complex, show greater syntactic variation, have stronger emotion, and contain irregular pragmatic usage; AI texts have more unified sentence patterns, smooth semantic connections, a basically neutral tone, and frequent templated expressions. Eye-movement performance: human-written texts involve longer fixation time, more regressions, and heavier information processing load; AI texts involve shorter fixation time, fewer regressions, and smoother reading. Neural responses: human-written texts produce larger amplitudes of N400, P600, and LPC, and stronger cognitive arousal; AI texts produce smaller ERP amplitudes and shallower cognitive investment. Discrimination ability: narrative genres are easier to identify and judgments are relatively reliable, while academic genres are difficult to distinguish and accuracy is not high.

5. Moderating Factors and Research Boundaries

EEG recording methods are well suited for analyzing brain electrical phenomena involved in cognition. This technology is now generally used together with eye-tracking methods in the examination of natural reading, and can also be used to consider neural differences in how humans and machines process texts.

5.1 Core ERP Components in Language Processing

Existing studies on classic ERPs have already connected three core components with language processing. The component referred to as N400 is a negative deflection at about 300-500 ms, and basically corresponds to the integration of lexical meaning or deviation from contextual prediction. The amplitude observed is a measure of semantic-processing load [4]. P600 is a positive deflection occurring within the range of 500-800

ms, and is related to syntactic reanalysis and changes in semantic state. Its amplitude is an indicator of the difficulty of structural analysis. The LPC is observed in the 600-1000 ms time window, and is the processing manifestation of continuous information, thinking intensity, and emotion. These three components are precisely the basic neurophysiological points for comparing human and AI text processing.

In simple terms, these three ERP EEG components are core electrical signals of different stages of language processing in the brain, and are also key neural indicators for comparing processing differences between human and AI texts. N400 (300-500 ms) reflects the difficulty of semantic matching; the less the content fits the context, the larger the amplitude. P600 (500-800 ms) corresponds to the difficulty of syntactic and semantic-logical reconstruction; the more abnormal the expression, the higher the amplitude. The LPC (600-1000 ms, positive wave) reflects the intensity of sustained deep processing of information, and is related to degree of thinking and emotional response.

5.2 Fixation-Related Potential Analysis in Natural Reading

The word-by-word rapid serial visual presentation used in traditional ERP is not close to actual reading and cannot well meet the processing needs of continuous texts. FRP analysis is one method to solve this type of problem. The FRP method takes the fixation time corresponding to the first appearance of the target word as the waveform starting point, and uses the idea of “first-fixation window separation + separate superposition of regressions” to limit waveform overlap caused by saccades or regressions [5]. More simply, FRPs are a method specially used to solve the problem that “EEG signals overlap and cannot be seen clearly.” When a person reads a sentence, the eyes jump back and forth and often go back to reread, which causes EEG signals at different stages to mix together and prevents accurate analysis. The idea of FRPs is very direct: the moment when the eyes first fixate on the target word is used as the timing starting point, and the EEG signals corresponding to “first reading” and “going back to reread” are separately counted and organized. In this way, overlap interference can be eliminated and clearer and more accurate results can be obtained. The method itself has already received considerable experimental support, especially from the perspective of cognitive neuroscience, and it can also serve as a reasonable technical means for examining human-AI text-processing differences under high ecological validity.

5.3 Neural Response Differences Between Human and AI Texts

Preliminary neuroimaging research has already pointed out that readers’ processing of human texts and AI-type texts is stably different in EEG. According to the current data, for AI-related texts, the N400, P600, and LPC mean amplitudes observed when reading human-written texts are generally more obvious and are concentrated in central, parietal, and temporal regions [2]. From a cognitive perspective, higher semantic entropy in human-written texts means that words are harder to predict and semantic integration is more difficult, thereby producing a larger N400 amplitude. Greater syntactic variation imposes a larger burden on syntactic processing, and the P600 amplitude accordingly increases. The richer the individual expression and emotional content, the easier it is to maintain attention or empathy, and therefore the amplitude of the LPC becomes larger. Power analysis in the α , β , and γ frequency bands also support the conclusion that cognitive arousal and processing effort are higher when people read human texts.

6. Moderating Factors and Research Boundaries

The cognitive differences between human and AI texts are not constant, but are regulated by multiple factors such as textual attributes and reader characteristics. Clarifying these boundary conditions helps in understanding the stability and scope of application of the differences.

6.1 Text Genre and Domain Effects

Text genre is now the moderating variable most often confirmed by research. Existing literature jointly points out that prose or narrative genres tend to involve larger differences in human-AI language and cognitive responses, while academic and formal explanatory texts generally show a smaller magnitude of difference [1]. Since academic writing is subject to more constraints and its language has less freedom, human individual expression is partly compressed, and thus its closeness to AI text is also greater. Correspondingly, readers’ accuracy in judging academic AI texts is not high, and the differences indicated by EEG and eye-movement

indicators are also weaker. Conclusions obtained from the perspective of genre show that discrimination methods should be properly distinguished according to specific applications.

6.2 Boundary Conditions Related to Readers and Models

In addition to the attributes contained in the text itself, reader characteristics are also factors in processing differences. If AI is considered by more experienced readers, their sensitivity to the textual features of AI will decline, and there will be fewer differences in both judgment accuracy and neural responses. People with relevant textual knowledge also show specific processing patterns. In terms of models, settings such as the temperature value used, that is, the parameter that regulates the randomness of AI-generated content, where a lower value produces more stable output and a higher value is more creative, as well as model size and training materials, will affect the characteristics of language output. Larger models combined with higher temperature can often generate texts that are semantically more disordered, namely higher in entropy. Such results are closer to human expression and are also more difficult to identify clearly.

7. Conclusion

The present paper systematically reviews the relevant research progress on human-AI text differences from three dimensions: linguistic features, behavioral and eye-movement performance, and EEG neural responses. It summarizes the core findings, theoretical explanations, and moderating factors in existing research.

7.1 Summary of Core Findings

Existing research has formed a basic consensus: human and AI texts show systematic differences from linguistic features to cognitive processing. At the linguistic level, human texts have higher semantic entropy, syntactic variation, and emotional density, while AI texts show templated and standardized features. At the behavioral and eye-movement level, human texts induce deeper processing, longer fixation duration, and more regressions, and discrimination is more difficult under formal genres. At the neural level, human texts induce larger amplitudes of N400, P600, and the LPC, reflecting higher semantic integration, syntactic processing, and sustained cognitive investment. Multidimensional indicators form a consistent evidence chain and jointly support the core mechanism of “differences in cognitive processing depth.”

8. Theoretical Extension and Future Prospects

8.1 Theoretical and Practical Value

At the theoretical level, research in this direction extends the research scope of linguistic cognitive processing theory to the era of generative AI, and reveals how the human brain processes probabilistic language output that differs from natural human expression. Compared with traditional computational linguistic research that only focuses on textual attributes, the cognitive neuroscience perspective opens the “black box” of reading processing and provides a deeper explanation for the essential differences between human and AI texts.

At the practical level, the research conclusions can provide empirical evidence for AI text discrimination technology, academic integrity governance, and large-model optimization. Cognitive-driven detection methods based on neural indicators can complement traditional feature-detection models, especially in scenarios where surface features are highly similar. In terms of model optimization, clarifying the cognitive mechanism of text naturalness can guide large language models to generate expressions with a more human quality.

8.2 Research Limitations

Although preliminary progress has been achieved, current research still has obvious limitations. First, existing studies mostly use young college students as samples, and the age and educational background ranges are limited, so the universality of the conclusions still needs further verification. Second, most experiments in this field are currently conducted with English materials and Western participants as the core. Human-AI text cognition research oriented toward Chinese contexts and Chinese populations is still very scarce. Existing studies cover only a small number of text genres and a single type of model, lacking systematic comparisons

under different models and different application scenarios. Third, the complete quantitative association chain from linguistic features to eye-movement behavior and then to neural responses has not yet been fully established.

8.3 Future Prospects

Future research can be advanced in four directions. First, the sample diversity of Chinese participants should be expanded to cover groups with different age levels, occupational backgrounds, and AI-use experience, so as to systematically test the influence of demographic variables on differences in human-AI text processing. Second, the scope of Chinese text genres and generation models should be expanded to cover typical Chinese scenarios such as news commentary, daily narrative, official-document writing, and academic abstracts, and the cognitive difference patterns of different mainstream Chinese large models should be compared to verify the stability and applicability of the conclusions. Third, a whole-chain quantitative model of “linguistic features - eye-movement indicators - neural responses” should be constructed, and the feasible path for integrating cognitive neural features into automatic Chinese AI text detection should be explored, so as to develop a more robust and more interpretable Chinese AI text discrimination method. Fourth, the application of cognitive neural indicators in evaluating the human similarity of Chinese large-model generated texts should be explored. Objective physiological indicators from EEG and eye movements can supplement traditional manual subjective scoring and provide a new standard with greater ecological validity for the quantitative evaluation of the naturalness of Chinese large-model generation.

References

- [1] Shen, X., & Wang, L. (2023). A study on the detection of artificial intelligence-generated academic journal text. *Science-Technology & Publication*, 42(8), 56-62.
- [2] Zhu, S., Zhong, Y., Ye, Z., Du, B., Zhou, Y., Ai, Q., & Liu, Y. (2026). How do humans process AI-generated hallucination contents: A neuroimaging study. arXiv. <https://doi.org/10.48550/arXiv.2605.16953>
- [3] Clark, E., August, T., Serrano, S., Haduong, N., Gururangan, S., & Smith, N. A. (2021). All that’s ‘human’ is not gold: Evaluating human evaluation of generated text. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 7282-7296). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.acl-long.565>
- [4] Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2013, August). Word surprisal predicts N400 amplitude during reading. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 878-883). <https://aclanthology.org/P13-2152/>
- [5] Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: analyses and review. *Journal of experimental psychology: General*, 140(4), 552. <https://doi.org/10.1037/a0023885>

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights

Copyright for this article is retained by the author (s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).