

Distinguishing AI-Generated and Human-Written Text Through Psycholinguistic Analysis

Chidimma Opara[✉]

School of Computing, Engineering & Digital Technologies
Teesside University, Middlesbrough, UK
c.opara@tees.ac.uk

Abstract. The increasing sophistication of AI-generated texts highlights the urgent need for accurate and transparent detection tools, especially in educational settings, where verifying authorship is essential. Existing literature has demonstrated that the application of stylometric features with machine learning classifiers can yield excellent results. Building on this foundation, this study proposes a comprehensive framework that integrates stylometric analysis with psycholinguistic theories, offering a clear and interpretable approach to distinguishing between AI-generated and human-written texts. This research specifically maps 31 distinct stylometric features to cognitive processes such as lexical retrieval, discourse planning, cognitive load management, and metacognitive self-monitoring. In doing so, it highlights the unique psycholinguistic patterns found in human writing. Through the intersection of computational linguistics and cognitive science, this framework contributes to the development of reliable tools aimed at preserving academic integrity in the era of generative AI.

Keywords: AI generated text · stylometry · psycholinguistics · plagiarism detection · education · authorship verification · cognitive load theory.

1 Introduction

AI-generated texts, powered by advanced models such as Openai’s GPT series, have significantly reshaped content creation, enabling artificial intelligence to produce text that closely resembles human writing. While these systems enhance efficiency and creativity, such as improving learning experiences [14], they also introduce new challenges in distinguishing AI-generated content from human-authored work, particularly in education, where written assignments serve as a key measure of critical thinking and subject expertise [2]. In response to these challenges, recent state-of-the-art research has focused on AI-text detection using deep learning techniques, including Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory networks, and Transformer-based models such as RoBERTa and GPT. These models have demonstrated excellent performance in classification tasks [19]. However, they are often criticised for their “black box” nature, offering high accuracy without transparent reasoning.

A more interpretable and linguistically grounded approach is **stylometry**, the quantitative analysis of writing style, traditionally used in literary authorship attribution. Stylometric techniques analyse measurable features such as lexical diversity, sentence complexity, and readability to identify patterns indicative of authorship [17,11,13]. Although stylometric features often yield high classification accuracy when applied on machine learning algorithms, they typically lack explanatory capacity regarding the cognitive processes that give rise to these features in human writing.

Understanding human authorship requires a deeper engagement with the **psycholinguistic** mechanisms that shape written expression. Stylistic features in human-authored texts are not arbitrary; they reflect complex cognitive activities such as self-monitoring, discourse planning, and lexical retrieval [8,10]. For instance, higher syntactic complexity may indicate intentional structuring of ideas. By aligning stylistic analysis with psycholinguistic theories, it becomes possible to develop more interpretable and cognitively informed approaches to distinguishing human-authored and AI-generated text.

1.1 Research Contribution

This study builds upon prior research [17] and extends the AI-generated text detection model (**StyloAI**) by grounding stylistic features in psycholinguistic theory. 31 stylistic features are systematically mapped to cognitive and linguistic processes, including Cognitive Load Theory, Metacognition and Self-Monitoring, Lexical Access and Retrieval, and Discourse Planning and Cohesion. This theoretical alignment enhances the understanding of why each of the stylistic features serves as an effective discriminator between human and AI-generated writing.

The remainder of the paper is organised as follows: The next section reviews psycholinguistic theories relevant to authorship detection. Section 3 provides an in-depth analysis of six stylistic feature categories: Lexical, Syntactic, Sentiment, Readability, Named Entity, and Uniqueness, each linked to corresponding psycholinguistic processes. Section 4 concludes the paper.

2 Psycholinguistic Theories in Human and AI-Generated Writing

2.1 Cognitive Load Theory

Cognitive Load Theory (CLT), introduced by Sweller in 1988 [22], posits that human working memory has limited capacity. When writing tasks overwhelm this capacity, through the simultaneous demands of idea generation, sentence construction, and argument organisation, performance declines and error likelihood increases [1]. This strain manifests in natural markers such as pauses, revisions, and stylistic fluctuations, which contribute to the uniqueness of human writing. For instance, under increased cognitive load, writers may simplify sentence structures or omit details [9]. These features are detectable through stylistic analysis and offer insight into the cognitive processes involved in human authorship.

By contrast, AI systems do not experience cognitive load in the human sense. Large Language Models (LLMs) generate coherent, fluent text without pausing, self-monitoring, or strategic planning [21]. While AI is bound by computational limits, these do not mirror the constraints of human cognition. As a result, AI-generated writing tends to be polished and consistent, yet often lacks the subtle imperfections typical of human expression. Efforts to simulate human-like limitations, such as bounded pragmatic speaker models [15], remain artificial and do not replicate authentic cognitive processes.

2.2 Metacognition and Self-Monitoring

Metacognition, as defined by Flavell (1979) [7], refers to one's ability to reflect on and regulate cognitive processes. In writing, this includes revising content to improve clarity, coherence, and

alignment with audience expectations. Such self-monitoring is evident in abrupt stylistic changes, iterative corrections, and context-sensitive modifications during drafting.

AI models, by contrast, lack fundamental metacognitive capabilities. Although models like GPT-4 can simulate reasoning through techniques such as chain-of-thought prompting [4,12], they do not engage in conscious reflection or deliberate revision. Instead, their refinements are statistical, not intentional. This limitation is especially evident in phenomena like hallucinations, overconfident but inaccurate outputs [20]. Unlike AI, human authors revise to correct meaning, adjust tone, or refine structure, making self-monitoring a key psycholinguistic marker of human authorship [23].

2.3 Lexical Access and Retrieval

Word selection in human writing involves a complex interplay between semantics, grammar, memory, and personal experience [6]. This cognitive process can result in occasional retrieval errors, such as tip-of-the-tongue moments or unintended repetition, often caught and corrected through self-monitoring. These subtle markers reflect real-time mental effort.

AI models, however, select words based on probabilistic associations derived from training data. While capable of producing text that mimics emotional richness, AI lacks lived experience and contextual depth, limiting the authenticity of its lexical choices [12]. Neuroimaging studies confirm that human writing activates brain areas related to memory and emotion dimensions AI cannot replicate [9,24]. Though AI may demonstrate lexical diversity, it often struggles to maintain tonal consistency and semantic cohesion over longer passages [4], making lexical variation another useful marker for authorship attribution.

2.4 Discourse Planning and Cohesion

Effective writing requires logical progression of ideas, as discussed by Halliday and Hasan (1976) [3]. Human writers use cohesive devices such as pronouns, transitions, and structured argumentation to build narratives hierarchically: from broad ideas to fine-grained sentence-level refinements. Yet, under cognitive load, this planning may falter, leading to abrupt topic changes or weak transitions.

AI-generated texts emulate coherence via statistical predictions rather than conceptual planning [12,24]. While initially appearing structured, these texts often contain contradictions or repetitiveness due to shallow contextual modelling [16]. In contrast, human writing typically reflects more intentional and consistent discourse planning, making cohesion and logical progression key differentiators [4].

3 Framework Mapping Stylometric Features to Psycholinguistic Theories

This section presents a systematic analysis of writing by integrating stylometric features with psycholinguistic theory.

The [StyloAI](#) model proposed in the previous work [17] incorporates 31 stylometric features, including 12 novel metrics specifically developed for detecting AI-generated texts. These features are organised into six distinct categories: Lexical Diversity, Syntactic Complexity, Sentiment and Subjectivity, Readability, Named Entities, and Uniqueness and Variety. For an extensive discussion on the theoretical rationale and detailed feature descriptions, readers are referred to the [previous work](#) and Table 2 in the Appendix.

Table 1 summarises the mapping of 18 out of the 31 of these features to psycholinguistic theories.

Stylometric Category	Psycholinguistic Category	Count	Features
Lexical Features	Cognitive Load	1	word_count
	Metacognition & Self-Monitoring	1	hapax_legomenon_rate
	Lexical Access & Retrieval	4	unique_word_count, ttr, avg_word_length, hapax_legomenon_rate
	Discourse Planning & Cohesion	1	char_count
Syntactic Features	Cognitive Load	2	avg_sentence_length, complex_sentence_count
	Metacognition & Self-Monitoring	3	punctuation_count, question_count, contraction_count
	Discourse Planning & Cohesion	1	complex_sentence_count
Uniqueness Features	Discourse Planning & Cohesion	1	syntax_variety
Sentiment Features	Cognitive Load	1	subjectivity
	Metacognition & Self-Monitoring	2	emotion_word_count, vader_compound
Readability Features	Cognitive Load	1	gunning_fog
	Metacognition & Self-Monitoring	1	gunning_fog
	Discourse Planning & Cohesion	1	flesch_reading_ease
Named Entity Features	Metacognition & Self-Monitoring	1	first_person_count
	Discourse Planning & Cohesion	1	direct_address_count

Table 1. Mapping Stylometric Features to Psycholinguistic Theories.

3.1 Detailed Feature Mapping

Lexical Features encompass the fundamental elements of language, words, and their usage patterns. This category includes metrics such as total word count, vocabulary diversity, average word length, and the frequency of unique words. In academic contexts, these features provide insight into cognitive processes such as lexical retrieval, semantic depth, and the ability to structure and integrate knowledge effectively [13,18].

For instance, the following is an excerpt of a human-written text from the dataset used in the **StyloAI** model, that statistically has a high value of unique word count:

"as you suspect a digestive problem in your rabbit, take him to your veterinarian immediately. If your rabbit has diarrhea, your veterinarian will test the feces to identify the specific organism (e.g., Clostridium)"

Psycholinguistic Interpretation: Human authors draw on semantic memory, lived experiences, and an awareness of audience expectations when choosing words. This often results in greater vocabulary diversity, precise language use, and contextually refined phrasing. Using stylometry, these behaviours are captured in features such as *Unique Word Count*, *TTR* and *Hapax Legomenon rate*. Psycholinguistically, these behaviours reflect active *lexical retrieval* and *metacognitive self-monitoring*. In contrast, AI models, although capable of statistically simulating lexical variation, often struggle to produce contextualised or original word choices, especially across longer passages.

Syntactic Features: Syntactic features examine sentence structure and grammatical complexity. Sentence construction places demands on working memory and discourse organisation, making it a key area where human and AI writing differ [22].

Psycholinguistic Interpretation: Human writers demonstrate flexibility in managing syntactic structure, tailoring sentence length, punctuation, and phrasing based on communicative goals such as clarity, emphasis, or engagement. These adjustments are enabled by all psycholinguistic theories discussed in Section 2, especially *self-monitoring* and *discourse planning* processes that respond dynamically to context and audience. In contrast, while syntactically fluent, AI-generated texts often exhibit a more rigid, repetitive structure. Without intrinsic self-monitoring, AI cannot vary syntax for rhetorical or communicative effect. As a result, its writing tends to lack the stylistic modulation typical of human authorship.

Sentiment Features: Sentiment analysis evaluates the emotional tone conveyed in a text. For example, the *Emotion Word Count*, which counts the total number of words associated with emotions in the text, is one of the stylometric features extracted in this category. In academic writing, expressions of enthusiasm, critique, or personal reflection can indicate human cognition [11,12]. While AI models can mimic emotional language, their expressions typically lack intrinsic emotional depth and intentionality, often resulting in mechanical or contextually misplaced language. For instance, the following is a generated text from the dataset used in the [StyloAI](#) model, demonstrating a mechanical expression of emotion:

"My car wont start, it doesnt turn over, it gives a clicking sound, keeps turning over but wont start, HELP."

The excerpt demonstrates the AI's attempt to convey emotion through specific linguistic choices. The use of "HELP" adds urgency and signals distress. The short, fragmented phrases create a choppy rhythm that mirrors frustration, while the repetition of "won't start" emphasises the speaker's emotional intensity and sense of helplessness.

Psycholinguistic Interpretation: Human authors naturally embed emotional layers in their writing, drawing from lived experience, empathy, and reflective processes. These expressions evolve organically within the text, drawing on *metacognition* and *cognitive load*, showing contextual variation and dynamic sentiment progression. An example of human-written text from the dataset used in the [StyloAI](#) model, that has a high emotion word count, but embeds it in layers:

"If you find yourself struggling with patience as you work to change your reputation, remind yourself that your reputation isn't either mature or immature like a light switch is on or off, but is instead on a continuum that varies in maturity..."

In contrast, AI-generated sentiment is based on statistical patterns rather than experiential grounding. As a result, emotional expressions in AI texts often appear flat, stereotyped, or overly uniform.

Readability Features Readability metrics assess the complexity and accessibility of a text, often combining indicators of lexical richness and syntactic structure. For example, *Gunning fog*, which estimates the years of formal education needed to understand a text on the first reading, is one of the stylometric features extracted in this category. These features offer insight into a writer's ability to adapt language for clarity and audience comprehension [13,11]. Human writers frequently adjust their writing based on reader expectations, while AI-generated texts may fluctuate between overly simple and unnecessarily complex content. The excerpt below of a generated text from the dataset used in the [StyloAI](#) model, that has a high value of Gunning fog, which means the text will be difficult to follow by a general audience:

"The Cooperation Agreement PREAMBLE The Parties, Recalling their common will to promote, protect and ensure the rights and freedoms of persons with disabilities and to promote..."

Psycholinguistic Interpretation: Human authors regulate text complexity through deliberate planning, ongoing self-monitoring, and revision, balancing accessibility with depth. These processes reflect *cognitive loadmetacognitive awareness, discourse-level planning*. In contrast, AI lacks true audience modelling capabilities and may produce writing that alternates unpredictably between overly dense and overly simplistic phrasing, without a clear rhetorical strategy.

Named Entity Features Named entities such as people, organisations, locations, and dates offer important contextual clues about how authors ground their writing in real-world knowledge, personal experience, or domain-specific understanding [13,11]. Human writers often incorporate actual references, drawing from memory and context. In contrast, AI-generated texts may include fabricated, vague, or inconsistently integrated entities.

Psycholinguistic Interpretation: Human-authored writing reflects deliberate discourse planning, often integrating references to real events, individuals, or timelines. These references emerge from episodic memory, thematic coherence, and personal narrative strategies. These characteristics reference *metacognition and discourse planning cognitive theories*. AI models, while capable of mimicking named entities, lack experiential grounding and may introduce irrelevant or inaccurate references due to limitations in factual recall or context awareness.

Uniqueness Features Uniqueness refers to how distinct a piece of writing is from commonly observed linguistic patterns. This may involve novel word combinations (e.g., bigrams or trigrams), varied sentence structures, or stylistic originality [11]. Human authors typically produce more original phrasing, influenced by creativity, domain familiarity, and context-sensitive reasoning. AI-generated texts, on the other hand, often rely on high-probability sequences drawn from training data, resulting in more predictable or generic outputs.

Psycholinguistic Interpretation: Original expression in human writing is underpinned by semantic memory, flexible lexical retrieval, and the ability to adapt tone and structure to audience or purpose. These unique characteristics demonstrate *lexical retrieval and discourse planning cognitive theories*. AI models, even when prompted for variety, generate phrases that are statistically probable but rhetorically shallow, lacking personal voice or adaptive strategy.

4 Conclusion

This study demonstrates that integrating stylometric analysis with psycholinguistic theory offers an interpretable framework for distinguishing AI-generated from human-authored texts. By mapping 31 stylometric features to underlying cognitive processes, such as lexical retrieval, discourse planning, cognitive load management, and metacognitive self-monitoring, this research enhances the understanding of the psycholinguistic signatures embedded in human writing. As AI tools become more integrated into educational and professional writing, it is essential to identify synthetic text while safeguarding the cognitive effort, originality, and ethical responsibility that characterise human authorship.

References

1. Baddeley, A.: Working memory: looking back and looking forward. *Nature reviews neuroscience* **4**(10), 829–839 (2003)
2. Bailey, R.A.: Undergraduate students' perceptions the role and utility of written assessment feedback. *Journal of Learning Development in Higher Education* (1) (2009)
3. Carrell, P.L.: Cohesion is not coherence. *TESOL quarterly* **16**(4), 479–488 (1982)
4. Collins, K.M., Wong, C., Feng, J., Wei, M., Tenenbaum, J.: Structured, flexible, and robust: benchmarking and improving large language models towards more human-like behavior in out-of-distribution reasoning tasks. In: Proceedings of the Annual Meeting of the Cognitive Science Society. vol. 44 (2022)
5. De Jong, T.: Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional science* **38**(2), 105–134 (2010)
6. Dell, G.S., O'Seaghda, P.G.: Stages of lexical access in language production. *Cognition* **42**(1-3), 287–314 (1992)
7. Flavell, J.: Theories of learning in educational psychology. *American Psychologist* **34**, 906–911 (1979)
8. García-Díaz, J.A., Beydoun, G., Valencia-García, R.: Evaluating transformers and linguistic features integration for author profiling tasks in spanish. *Data & Knowledge Engineering* **151**, 102307 (2024)
9. Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S.A., Feder, A., Emanuel, D., Cohen, A., et al.: Shared computational principles for language processing in humans and deep language models. *Nature neuroscience* **25**(3), 369–380 (2022)
10. Huang, B., Chen, C., Shu, K.: Authorship attribution in the era of llms: Problems, methodologies, and challenges. *ACM SIGKDD Explorations Newsletter* **26**(2), 21–43 (2025)
11. Kumarage, T., Garland, J., Bhattacharjee, A., Trapeznikov, K., Ruston, S., Liu, H.: Stylometric detection of ai-generated text in twitter timelines. *arXiv preprint arXiv:2303.03697* (2023)
12. Mahowald, K., Ivanova, A.A., Blank, I.A., Kanwisher, N., Tenenbaum, J.B., Fedorenko, E.: Dissociating language and thought in large language models. *Trends in Cognitive Sciences* (2024)
13. Mindner, L., Schlippe, T., Schaaff, K.: Classification of human-and ai-generated texts: Investigating features for chatgpt. In: International Conference on Artificial Intelligence in Education Technology. pp. 152–170. Springer (2023)
14. Ngo, T.T.A.: The perception by university students of the use of chatgpt in education. *International Journal of Emerging Technologies in Learning (Online)* **18**(17), 4 (2023)
15. Nguyen, K.X.: Language models are bounded pragmatic speakers. In: First Workshop on Theory of Mind in Communicating Agents (2023)
16. Nolfi, S.: On the unexpected abilities of large language models. *Adaptive Behavior* **32**(6), 493–502 (2024)
17. Opara, C.: Styloai: Distinguishing ai-generated content with stylometric analysis. In: International conference on artificial intelligence in education. pp. 105–114. Springer (2024)
18. Paas, F., van Merriënboer, J.J.: Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science* **29**(4), 394–398 (2020)
19. Sarvazyan, A.M., González, J.Á., Franco-Salvador, M., Rangel, F., Chulvi, B., Rosso, P.: Overview of autextification at iberlef 2023: Detection and attribution of machine-generated text in multiple domains. *arXiv preprint arXiv:2309.11285* (2023)
20. Smith, A.L., Greaves, F., Panch, T.: Hallucination or confabulation? neuroanatomy as metaphor in large language models. *PLOS Digital Health* **2**(11), e0000388 (2023)
21. Suresh, S., Mukherjee, K., Yu, X., Huang, W.C., Padua, L., Rogers, T.T.: Conceptual structure coheres in human cognition but not in large language models. *arXiv preprint arXiv:2304.02754* (2023)
22. Sweller, J.: Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction* **4**(4), 295–312 (1994)
23. Talboy, A.N., Fuller, E.: Challenging the appearance of machine intelligence: Cognitive bias in llms and best practices for adoption. *arXiv e-prints* pp. arXiv–2304 (2023)
24. Tuckute, G., Kanwisher, N., Fedorenko, E.: Language in brains, minds, and machines. *Annual Review of Neuroscience* **47** (2024)

Feature	Description	Psycholinguistic Rationale	Stylistic Category
WordCount	Total number of words in the text	WordCount reflects cognitive load balancing: more words often imply efficient idea generation and expression, as outlined by Sweller's Cognitive Load Theory [22].	Lexical Features
UniqueWordCount	Number of distinct words used	UniqueWordCount taps into lexical access and retrieval processes [6]. A broader range of vocabulary implies active retrieval from semantic memory and flexibility in language use, both requiring fluent cognitive control.	
CharCount	Total number of characters	CharCount correlates with discourse planning—longer texts usually present more cohesive arguments [3].	
AvgWordLength	Average length of words	AvgWordLength indicates the level of vocabulary sophistication retrieved during writing [6]. Longer words may indicate access to more technical or abstract lexical items, but if cognitive load increases, writers might opt for simpler, shorter words to conserve cognitive resources [5].	
TTR	Ratio of unique words to total words	A high TTR demonstrates dynamic lexical retrieval [6] and active self-monitoring [7]. Low TTR can point to repetitive phrasing under mental strain.	
HapaxLegomenonRate	Proportion of words used only once	Hapax Legomena highlights creative, context-specific word use and metacognitive planning to avoid redundancy [6,7].	
SentenceCount	Total number of sentences	Captures how authors chunk ideas, balancing cognitive [22] load and reader comprehension [3].	
AvgSentenceLength	Average number of words per sentence.	Longer sentences reflect higher planning effort; shorter ones can signal simplification under cognitive strain [22].	
PunctuationCount	Total punctuation marks	PunctuationCount shows metacognitive control over sentence boundaries, rhythm, and readability [7].	
StopWordCount	Total count of common function words	StopWordCount balances grammatical fluency and cohesion [3].	
ComplexSentenceCount	Number of sentences with more than one clause	Complex sentences show layered discourse planning [3] and advanced working memory management [22], allowing multiple ideas to be integrated cohesively.	Syntactic Features
QuestionCount	Total question marks	Questions reflect rhetorical engagement and audience-awareness strategies [7].	
ExclamationCount	Total exclamation marks	Exclamations signal emotional emphasis or spontaneity in discourse [3].	
ContractionCount	Number of contractions (e.g., don't, can't)	Contractions demonstrate stylistic awareness and adaptation to context, reflecting self-monitoring [7] and flexible tone control.	
AbstractNounCount	Number of intangible concept nouns	Abstract nouns highlight conceptual abstraction abilities, engaging higher-level cognitive functions like thematic integration and abstract discourse planning [3].	
ComplexVerbCount	Less frequent verbs	The use of complex verbs reflects semantic sophistication and flexible lexical access [6], often associated with subject-matter expertise.	
SophisticatedAdjectiveCount	Complex adjectives (e.g., “-ive,” “-ous”)	Highlights rich descriptive abilities tied to discourse planning and lexical retrieval, showing an intention to convey precision [3].	
EmotionWordCount	Number of emotion-laden words	Emotion word usage reflects authentic emotional processing and discourse planning aimed at emotional engagement [7].	
Polarity	Overall sentiment orientation	Polarity tracks shifts between positive and negative tones in argument flow [7].	
Subjectivity	Degree of opinion vs factual language	Subjectivity in text highlights personal stance and metacognitive involvement [22].	
VaderCompound	Overall sentiment score	Quantifies subtle emotional regulation throughout the text [3].	Readability Features
FleschReadingEase	Ease of reading score	Reflects cognitive effort in adjusting complexity for audience comprehension [7] and working memory management [22].	
GunningFog	Education level needed to understand text	Indicates how writers balance depth of ideas against clarity and flow [3].	
FirstPersonCount	First-person pronouns	Reflects self-referential awareness and narrative framing [7].	
DirectAddressCount	Direct reader engagement	Indicates audience-awareness and planning strategies [3].	Named Entity Features
PersonEntities	Names of individuals	Taps into episodic memory retrieval [6,22] and authentic context integration.	
DateEntities	Temporal markers mentioned	Reflects the integration of personal or historical timelines in discourse [6,22].	
UniqueNgramCount	New word combinations (bigrams/trigrams)	Captures associative creativity and flexible lexical retrieval [6], shaped by personal memory and discourse needs.	
SyntacticVariety	Range of sentence structures	Reflects flexible discourse planning and adaptability [3].	Uniqueness Features

Table 2. Stylistic Features and Their Psycholinguistic Rationales, by Category