Annales Mathematicae et Informaticae
61 (2025) pp. 248–260
DOI: 10.33039/ami.2025.10.013
URL: https://ami.uni-eszterhazy.hu

Syntactic comparison of human and AI-written scientific texts

Erika B. Varga[a], **Attila Baksa**[b]

[a]Institute of Information Technology
erika.b.varga@uni-miskolc.hu

[b]Institute of Applied Mechanics
Faculty of Mechanical Engineering and Informatics
University of Miskolc, H-3515 Miskolc, Hungary
attila.baksa@uni-miskolc.hu

Abstract. The spread of large language models (LLMs) has transformed scientific writing, enabling the generation of fluent and convincing text with minimal human input. This development poses significant challenges for authorship verification, especially when AI-generated or AI-assisted content is embedded in academic manuscripts. While most existing detection approaches rely on surface-level lexical features or stylometric clues, our study proposes a novel syntactic-level method to distinguish between human-authored, translated, and AI-generated scientific texts. We constructed a controlled corpus of 24 scientific articles in the field of computer science, divided into four categories: native-authored, human-translated, ChatGPT 4.0-generated, and ChatGPT 4o-generated with deep research. Each corpus was processed using part-of-speech (POS) and dependency parsing, followed by statistical profiling and sentence-structure discovery via process mining. Our results reveal that AI-generated texts differ significantly in their use of modal verbs, participles, coordination, and syntactic complexity. We demonstrate that process-mined graphs of syntactic transitions provide an interpretable and robust fingerprint of authorship, enabling us to detect AI-generated patterns and differentiate them from translated or native writing. The proposed framework contributes a novel methodological perspective to the growing field of AI authorship detection.

Keywords: AI-generated text detection, syntactic analysis, sentence structure modeling, process discovery

AMS Subject Classification: 68T50, 68T20, 68U15

Accepted: October 8, 2025
Published online: October 28, 2025

1. Introduction

AI tools built on large language models (LLMs) such as ChatGPT have become increasingly widespread in academic writing. From brainstorming and drafting to paraphrasing and summarization, LLMs are now embedded in the workflow of researchers and students alike. This shift in authorship practices has raised important questions about the authenticity of scientific writing and the ability to distinguish AI-generated text from human-authored content.

Current approaches to AI detection focus predominantly on surface-level characteristics such as word frequency patterns, perplexity scores, and stylometric irregularities [9, 16]. While these methods offer fast and scalable detection, AI-generated texts can slip through these systems by means of paraphrasing, translation, or editing, which obscure the statistical fingerprints of AI models. Moreover, stylometric features tend to treat grammatical elements as isolated tokens rather than analyzing how sentences are structured syntactically [6].

In this paper, we argue that syntactic patterns, especially sentence-level grammatical structures, offer a deeper and more robust basis for detecting AI-generated writing. We focus on sentence structure as a distinctive linguistic fingerprint and investigate how it varies across different types of text: native-authored, human-translated, ChatGPT-generated, and ChatGPT 4o-generated (in deep research mode). To this end, we created a four-part scientific text corpus, each group consisting of six scientific papers written or generated under controlled conditions in the computer science domain. We apply two complementary methods: (1) statistical profiling based on part-of-speech (POS) tags and dependency roles, and (2) process mining to discover generalized syntactic flow patterns in sentence construction using the Heuristics Miner algorithm [14]. While the former quantifies grammatical characteristics, the latter visualizes structural tendencies through heuristic process graphs. This experiment introduces a novel application of process mining in linguistic analysis.

2. Related works

Detecting AI-generated or AI-influenced text has become a prominent research field, particularly with the widespread use of large language models (LLMs) in academic writing. Stylometric methods relying on features such as function-word usage, part-of-speech (POS) distributions, and sentence-length metrics are proving effective in distinguishing AI-generated content from human-authored text. Notably, Zaitsu & Jin [16] achieved 98 % accuracy in classifying GPT-generated versus human-written Japanese scientific text using POS bigrams and function-word frequencies. Prova (2024) [9] presents a hybrid detection model using feature-based classifiers (XGBoost, SVM) alongside a BERT-based architecture. Trained on a balanced dataset of AI- and human-generated samples, the BERT model achieved 93 % accuracy, outperforming XGBoost (84 %) and SVM (81 %). The recent StyloAI model [6] applies 31 stylometric features, including new grammatical markers,

using Random Forest classifier and achieves up to 98 % accuracy in multi-domain detection tasks.

However, these approaches treat syntactic structure as static features, such as isolated POS or dependency tag frequencies. Researchers have advanced the field by using POS n-grams and syntactic n-grams. For instance, Pokou et al. in [7] introduced variable-length POS patterns for authorship attribution, capturing frequent POS sequences rather than single-tag statistics. Posadas-Durán et al. in [8] used complete syntactic n-grams from dependency trees to profile writing style, demonstrating improved performance over lexical n-grams. Still, these methods remain static, focusing on patterns without modeling the sequential structure of syntactic roles.

Structured dependency stylometry has also gained attention. Murauer & Specht in [5] proposed DT-grams, language-independent dependency-tree-gram features for cross-language authorship identification, proving their utility in capturing grammatical style when transferring texts across languages. However, like POS and syntactic n-grams, DT-grams treat dependency substructures in isolation and do not model the flow of syntactic roles within sentences.

In parallel, AI-influenced translation, where human text is revised or translated by LLMs, poses new detection challenges. Systems like GPTZero struggle to identify AI-assisted rewriting, with performance dropping to 28 % F1 in paraphrased cases and producing false positives in human writing [3]. Krishna et al. in [4] report similar challenges, noting that AI-assisted paraphrasing often avoids detection due to preservation of human syntactic patterns.

Our earlier work addresses part of this challenge through lexical profiling. In [13], we introduced a synonym set based approach using WordNet synsets to measure conceptual recursion and redundancy. We found that translated texts exhibit higher lexical density than AI texts, while native texts share greater concept overlap with AI output. We also analyzed lexical redundancy and conceptual overlap, showing clear distinctions between AI- and human-authored scientific writing in [12].

While stylometry outlines linguistic fingerprints, it does not capture the dynamic flow of sentence construction. Process mining has been recently combined with NLP for extracting structured models from event logs, with applications such as semantic role labeling from textual logs [10] and exploratory process model discovery via language models [11]. However, no prior work directly applies process mining to syntactic dependency sequences in text.

Our present approach applies Heuristic Miner to build dependency-role transition graphs from sentence-level parse sequences, yielding graphical models that reflect typical structures. This method extends traditional syntactic analysis by representing how roles (e.g., subject → verb → object → modifier) sequentially co-occur, offering a novel, process-oriented view. To our knowledge, it is the first attempt at such structural profiling. In the context of translated texts, which may combine human sentence planning with AI-driven surface-level phrasing, our process-mining approach reveals patterns such as syntactic chaining or padding

that are indicative of AI involvement. These structural patterns supplement lexical measures, enabling a more robust detection of AI assistance in translation.

3. Methodology

3.1. Corpus design

The text corpus used in this study consists of 24 scientific articles from the Computer Science domain, divided into four categories, each containing six texts. The first category includes human-written papers authored by native English speakers and published between 2000 and 2015, prior to the emergence of LLMs. Their counterparts form the second category: articles generated by ChatGPT-4. For these, the title and abstract of each human-written paper were provided as prompts, together with an instruction to write a full scientific article of about 10,000 words. In practice, however, ChatGPT-4 produced outputs of approximately 5,000 words, and none of the generated texts exceeded this length.

The third category consists of translations of papers written by non-native researchers after 2020. These texts were created with moderate AI support, which was used only to improve linguistic quality, such as correcting grammatical errors, smoothing the discussion flow, and enhancing stylistic variety. The fourth category contains the AI-generated counterparts of these translations, produced with ChatGPT-4o's deep research function. As in the previous case, the model was given the title and abstract of each translated paper and instructed to write a full scientific article of about 10,000 words. In this case, ChatGPT-4o complied with the requested length and generated significantly longer articles, typically over 10,000 words.

The use of these two different AI models explains the variation in text length between the paired categories. All texts nevertheless exceed 3,000 words, ensuring sufficient syntactic complexity and topic depth for the analysis. The basic characteristics of the papers in the collection, as well as their lexical comparison, are reported in a recent paper [13].

3.2. Syntactic annotation

Text files were first cleaned and segmented into individual sentences using a combination of regular expressions and the NLTK python library's sentence tokenization function. Non-linguistic elements such as titles, metadata, and references were excluded. Sentences were then normalized by: (1) removing punctuation, numeric tokens, and special characters; (2) filtering short, non-informative lines (e.g., headings) and retaining sentences with sufficient lexical and grammatical complexity. Next, each sentence was POS-tagged and dependency-parsed using spaCy's `en_core_web_sm` model to enable sentence-level syntactic profiling.

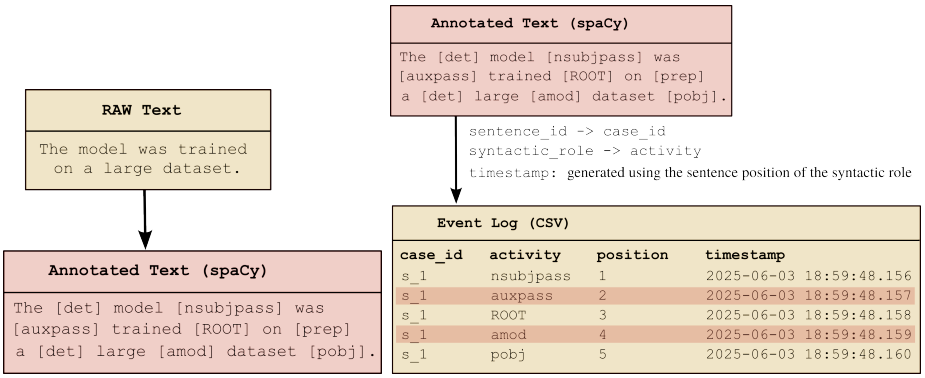
3.3. Sentence structure discovery

In this study, we introduce a novel methodological contribution by applying process mining techniques to explore and visualize the structural complexity of natural language sentences. Traditionally used in business process management and system logs, process discovery allows the extraction of structured workflow models from event logs [2]. Here, we adapt this technique to syntactic dependency data, treating each sentence as a trace and each dependency label as an event within that trace. For this approach, all texts had to be transformed into an event log that captures the sequence of syntactic roles. To illustrate the concept, a simplified example is shown below:

$$\mathcal{L} = \{\langle A, B, C, D\rangle, \langle A, C, D\rangle, \langle A, B, D\rangle\}$$

In this example, log $\mathcal{L}$ contains three traces, each representing a sentence as an ordered list of dependency labels. A, B, C, and D are placeholder symbols standing for different dependency labels that represent syntactic roles (e.g., subject, verb, object, modifier).

Event logs typically contain at least a case ID (a unique identifier for a process instance), an activity label (representing an event), and a timestamp. In our setting, the timestamps do not carry linguistic information but are required by the process mining algorithm to determine the order of events within each trace. They were therefore generated automatically according to the sequential order of the dependency labels in a sentence. The syntactic annotation and log generation pipeline is shown in Figure 1.



(a) Syntactic annotation. **(b)** Event log generation.

Figure 1. Text transformation into event log.

During the construction of the event logs, not all syntactic roles (dependency labels) were considered equally relevant. To focus the analysis on core sentence structure and reduce noise from frequent but less informative elements, several dependency types were deliberately excluded from the logs. The omitted labels include common grammatical markers such as determiners (det), prepositions and

case markers (case), subordinating and coordinating conjunctions (mark, cc), preconjunctions (preconj) and punctuation (punct). These roles typically represent grammatical scaffolding rather than semantic or structural functions within a sentence. By excluding them, the resulting log better captures the backbone of syntactic constructions, such as subject–verb–object relationships and clause-level dependencies.

Common process discovery algorithms include (1) Alpha Miner [1], that is suitable for simple, noise-free logs, (2) Heuristics Miner [14] which is robust against exceptional or infrequent behavior, and (3) Inductive Miner [15] which produces sound, block-structured models. For this study, the Heuristics Miner algorithm was selected for two main reasons. Firstly, natural language sentence structures exhibit high variability. Nearly every sentence can be seen as a unique case with different syntactic sequences. This results in a highly exceptional log with few frequent patterns. The Heuristics Miner is designed to handle such noisy, non-repetitive event logs, making it ideal for linguistic applications. Secondly, unlike other algorithms that produce complex Petri nets, the Heuristics Miner generates a graphical model, called heuristic net, that emphasizes the most statistically significant transitions between events. This representation is more interpretable for syntactic analysis, where the goal is not strict process conformance but insight into sentence construction tendencies like how subjects relate to objects, or the placement of modifiers.

4. Results

This section presents the results of sentence-level grammatical profiling across native-authored, translated, and AI-generated scientific texts. The analysis addresses two main objectives:

1. to quantify the statistical differences in grammatical structure between the text categories, and

2. to identify distinctive sentence structure patterns that can differentiate human-authored and translated texts from AI-generated ones.

4.1. Quantitative analysis of grammatical differences

To quantify the differences between the four groups of texts, we have applied four statistical measures. All grammatical comparisons are based on normalized or sentence-level average metrics to eliminate distortions caused by differing document lengths.

4.1.1. POS distribution

The normalized POS tag frequencies in Figure 2 reveal clear lexical patterns that differentiate the three text groups. Notably, AI-generated texts stand out for their higher frequency of modal verbs (MD), present participles (VBG), adjectives (JJ),

plural nouns (NNS), and adverbs (RB). These patterns suggest a tendency toward generalization, elaboration, and structural padding. These are typical features of model-generated language that aims for academic style without strong referential grounding. Native-authored texts lead in the use of proper nouns (NNP), infinitival markers (TO), and personal pronouns (PRP), reflecting a more referential and agent-oriented writing style. Not surprisingly, translated texts share features with both groups. They resemble AI texts in their frequent use of common nouns (NN, NNS) and TO, which suggests compact syntax likely arising from translation conventions or simplification. On the other hand, they align more closely with native texts in their use of past participles (VBN) and proper nouns (NNP), the features associated with academic conventions like passive constructions and citations.

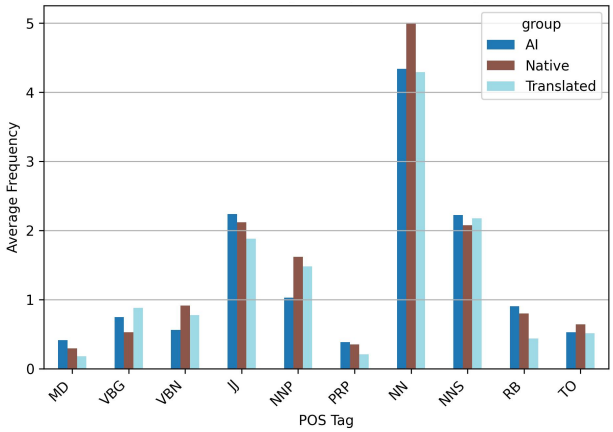


Figure 2. Average POS distributions.

4.1.2. Syntactic role distribution

The syntactic structure of sentences was analyzed by grouping dependency roles into higher-level categories (e.g., *Subject*, *Object*, *Modifier*). The radar chart in Figure 3 visualizes the average number of each role per sentence, aggregated by text group. Remarkably, AI-generated and native-authored texts show similar syntactic role distributions across most categories. This suggests that AI writing tools successfully replicate human-like sentence structuring in scientific text. The most notable divergence is seen in the Modifier role. Human-written texts consistently use more modifiers, indicating a tendency to enrich noun phrases or insert descriptive elements. AI texts, by contrast, apply modifiers more conservatively, potentially favoring structural clarity over elaboration. In terms of Coordination, the pattern is reversed. AI texts exhibit higher coordination than human-written ones, suggesting a preference for parallel or additive structures. Native texts rely less on coordination, possibly favoring more hierarchical or subordinated constructions. The Object role also shows a distinct contrast. Native-authored texts feature

the highest object frequency, which may reflect more diverse transitive constructions or greater syntactic density. AI texts exhibit lower values here, hinting at more simplified or formulaic sentence building. Other roles such as *Subject*, *Predicate*, and *Adverbial* remain relatively stable across all groups, indicating a shared core structure in scientific writing regardless of authorship.

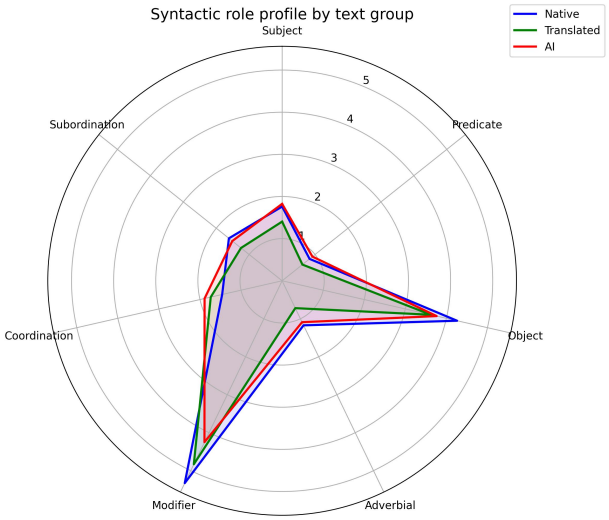


Figure 3. Average distribution of syntactic roles.

4.1.3. Position of key syntactic roles

The most frequent positions of the main syntactic roles (`nsubj`, `ROOT`, `dobj`) shown in Figure 4 further distinguish the groups. Although native texts have the longest sentences, averaging over 25 tokens, the subject (`nsubj`) typically appears in the first position, while the predicate (`ROOT`) appears in the second position, which aligns with canonical English word order (*Subject–Verb–Object*). In AI-generated texts, these roles are delayed, likely due to the frequent use of fronted modifiers. Direct object (`dobj`) shows the most notable difference. In AI-generated texts, it appears later (usually at position 7), while in both native and translated texts the object appears at position 5. This suggests that AI systems tend to insert more modifiers between the verb and the object, creating syntactically padded structures.

4.1.4. Sentence complexity profiles

The comparison of normalized sentence structure distributions in Figure 5 highlights distinct tendencies across author groups. Interestingly, AI-generated texts rely heavily on active constructions, with nearly 80 % of sentences being active.

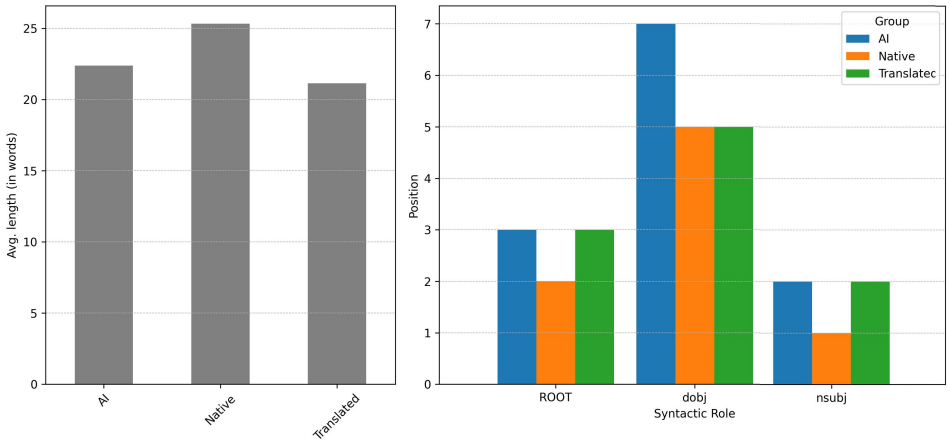


Figure 4. (a) Average sentence length (b) Most frequent positions of key syntactic roles.

They also show minimal use of complex and passive structures, suggesting syntactic simplicity and regularity. This reflects the model's preference for clarity and reduced grammatical embedding. Native-authored texts, in contrast, display greater structural diversity. While active voice still dominates, these texts feature more passive and complex sentences, indicating a higher degree of syntactic flexibility and subordination. Translated texts exhibit a hybrid behavior. Their use of complex and fragmentary sentences is slightly higher than in native writing, possibly due to literal rendering of source syntax or segmentation artifacts introduced by translation tools. Passive constructions occur less frequently than in native texts, but more than in AI-generated content, pointing to some retention of authentic academic style.

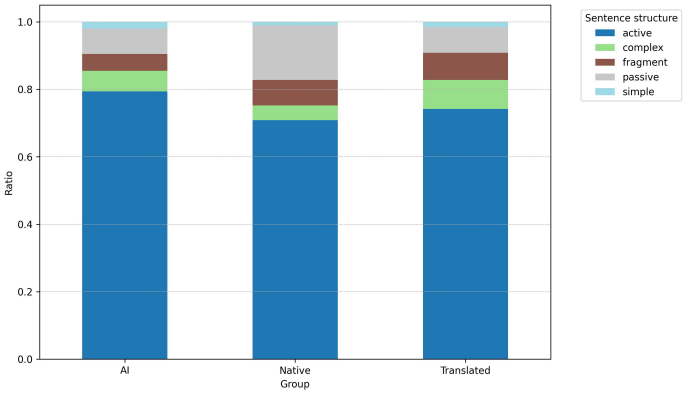
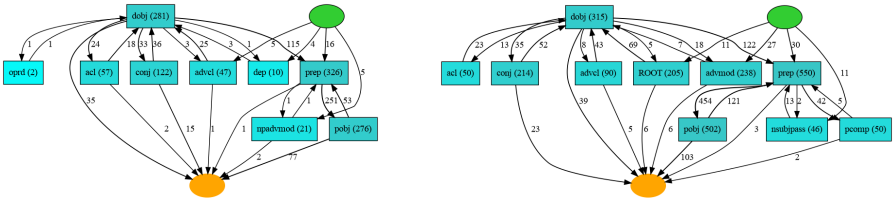


Figure 5. Normalized sentence structure ratios.

While the statistical profiles above capture frequency-based syntactic tendencies, process mining allows us to explore how these grammatical elements are sequenced and interact within individual sentences.

4.2. Sentence structure discovery

To identify structural patterns typical of each author group, we applied the Heuristic Miner algorithm to generate dependency-role process graphs for every single text in the corpus. These individual graphs model the transitions between major dependency roles, capturing both their frequency and structural positioning within sentences. We also experimented with creating aggregated process graphs for entire groups; however, these models became overly large and heterogeneous, making them less suitable for meaningful interpretation. For this reason, the paper presents one representative graph from each group, selected to illustrate characteristic sentence-level syntactic tendencies identified in that category. Specifically, Figure 6a shows an example from the human-translated texts, Figure 6b from the ChatGPT-4 generated texts, Figure 7a from the native-authored texts, and Figure 7b from the ChatGPT-4o deep research texts. The development of more effective methods for constructing generalized group-level process models is left for future work.

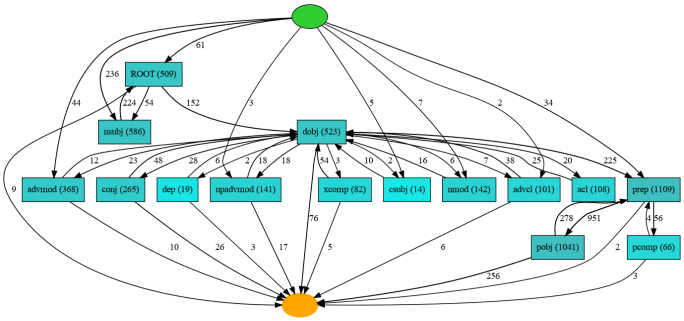


(a) Translated text. **(b)** ChatGPT 4.0 generated text.

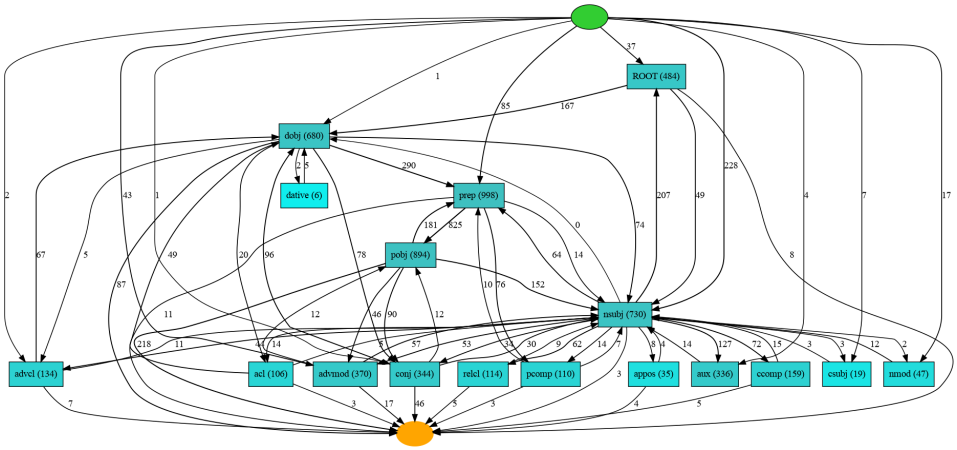
Figure 6. Heuristic dependency structure graphs of translated and ChatGPT 4.0-generated texts.

We can observe that the process graphs of human-translated and ChatGPT 4.0 texts are structurally similar. Both graphs display relatively shallow syntactic structures which indicates a tendency toward simpler and flatter clause chains. These texts exhibit frequent use of prepositional phrases which reflects compact sentence construction and dense nominal modification (i.e., sequences of stacked adjectives or noun modifiers within noun phrases). This is considered typical in translated texts and baseline AI. The lower presence of subordinate structures (`xcomp`, `advcl`, and `csubj`) suggests limited clause embedding. It is also worth noting, that both groups rely heavily on direct object constructions, indicating strong SVO alignment and topic-focus in sentence planning.

The process graphs of native writings and texts generated by ChatGPT 4o deep research function show deeper syntactic complexity and stronger structural variation. Subordinate clause structures (`advcl`, `xcomp`, and `acl`) appear more often



(a) Native-authored text.



(b) ChatGPT 4o deep research generated text.

Figure 7. Heuristic dependency structure graphs of native and ChatGPT 4o deep research generated texts.

and in richer configurations. Especially, subject and object (`nsubj` and `dobj`) are embedded more frequently inside these clauses, which indicates syntactic elaboration and hierarchical depth. Also, both graphs include more dependency roles (`conj` and `ccomp`) and more transitions, which suggests the high use of coordinated or parallel clause structures.

5. Discussion and conclusion

This study set out to identify grammatical features that can help infer the extent of AI involvement in human-translated scientific texts. Through a combination of frequency-based syntactic profiling and structural process mining, we have uncovered clear patterns that distinguish AI-generated content from both native-authored and translated texts.

Our analysis revealed that AI-generated texts exhibit distinctive surface-level tendencies. These include an elevated use of modal verbs, participles, and coordinated structures, combined with reduced use of passive voice and embedded clauses. Although these patterns aim to mimic academic tone, they often result in overly regular sentence structures with less syntactic depth. By contrast, native-authored texts demonstrated more referential grounding and syntactic variability, marked by higher frequencies of proper nouns, passive constructions, and subordinate clauses.

Not surprisingly, translated texts emerged as a hybrid category. In many aspects, such as the prevalence of common nouns, infinitival markers, and prepositional phrases, they aligned with baseline AI output. This suggests the use of translation systems which tend to apply simplification and structural compression. However, their use of passive voice, proper nouns, and past participles more closely resembled human writing, particularly in stylistic conventions of scientific discourse.

Notably, the most insightful patterns emerged from process mining. The structural graphs showed that baseline AI (ChatGPT 4.0 default) and translated texts rely on flatter syntactic chains with limited embedding, while native and advanced AI (ChatGPT 4o with deep research) texts reveal richer clause interactions which reflect a deeper level of syntactic planning.

These findings suggest that it is possible to develop diagnostic criteria to detect AI involvement in translated texts. Texts that display (1) high coordination but low subordination, (2) frequent direct object placement at later sentence positions, (3) dense nominal modification with limited clause embedding, and (4) reduced use of referential markers (e.g., proper nouns, personal pronouns), are more likely to contain AI-generated segments.

While translated texts may inherit some of these attributes from the source language or the translation process itself, an over-concentration of such features, especially when aligned with statistical outliers in modifier usage or syntactic role position, can serve as a heuristic indicator of AI usage.

6. Limitations and future work

A limitation of the present study is that syntactic roles were automatically assigned using the `spaCy` parser, which may introduce annotation inaccuracies. Another limitation concerns the corpus size: with 24 papers divided into four categories, the dataset is adequate for a proof-of-concept but relatively small for drawing broader conclusions. While each paper is long enough to provide syntactic depth, the limited number of texts constrains the representativeness of the findings. As a direction for future research, we aim to expand the corpus and to develop methods for constructing more interpretable generalized process graphs at the group level, so that broader structural tendencies can be captured.

References

[1] W. van der Aalst, T. Weijters, L. Maruster: *Workflow mining: discovering process models from event logs*, IEEE Transactions on Knowledge and Data Engineering 16.9 (2004), pp. 1128–1142, doi: 10.1109/TKDE.2004.47.

[2] W. van der Aalst: *Process Mining*, Springer Berlin, Heidelberg, 2016, isbn: 978-3-662-49850-7, doi: 10.1007/978-3-662-49851-4.

[3] D. Brown, D. Jensen: *GPTZero vs. Text Tampering: The Battle that GPTZero Wins*, in: ed. by M. Shelley, V. Akerson, M. Unal, International Conference on Social and Education Sciences, Las Vegas, NV, USA. ISTES Organization, 2023, pp. 734–744, url: https://files.eric.ed.gov/fulltext/ED656089.pdf.

[4] K. Krishna, Y. Song, M. Karpinska, J. Wieting, M. Iyyer: *Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense*, 2023, url: https://arxiv.org/abs/2303.13408.

[5] B. Murauer, G. Specht: *DT-grams: Structured Dependency Grammar Stylometry for Cross-Language Authorship Attribution*, 2021, url: https://arxiv.org/abs/2106.05677.

[6] C. Opara: *StyloAI: Distinguishing AI-Generated Content with Stylometric Analysis*, 2024, url: https://arxiv.org/abs/2405.10129.

[7] Y. J. M. Pokou, P. Fournier-Viger, C. Moghrabi: *Authorship Attribution using Variable Length Part-of-Speech Patterns*, in: Proceedings of the 8th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART, INSTICC, SciTePress, 2016, pp. 354–361, isbn: 978-989-758-172-4, doi: 10.5220/0005710103540361.

[8] J.-P. Posadas-Duran, G. Sidorov, I. Batyrshin: *Complete Syntactic N-grams as Style Markers for Authorship Attribution*, in: Human-Inspired Computing and Its Applications, ed. by A. Gelbukh, F. C. Espinoza, S. N. Galicia-Haro, Cham: Springer International Publishing, 2014, pp. 9–17, isbn: 978-3-319-13647-9, doi: 10.1007/978-3-319-13647-9_2.

[9] N. Prova: *Detecting AI Generated Text Based on NLP and Machine Learning Approaches*, 2024, url: https://arxiv.org/abs/2404.10032.

[10] A. Rebmann, H. van der Aa: *Extracting Semantic Process Information from the Natural Language in Event Logs*, 2021, url: https://arxiv.org/abs/2103.11761.

[11] A. Rebmann, F. D. Schmidt, G. Glavaš, H. van der Aa: *Evaluating the Ability of LLMs to Solve Semantics-Aware Process Mining Tasks*, 2024, url: https://arxiv.org/abs/2407.02310.

[12] E. B. Varga, A. Baksa: *A Comparative Analysis of Human and Machine Written Scientific Texts*, in: ed. by E. Xhaferra, Proceedings of the 1st International Conference on Computer Science and Information Technology, Durrës, Albania, 2025, url: https://csitconf.org/Book_of_Proceedings_CSIT2025.pdf.

[13] E. Varga, A. Baksa: *Exporing Lexiacal Variation Through Synonym Sets in Human and AI-Written Scientific Texts*, submitted to journal Multidisciplinary Sciences, 2025.

[14] A. Weijters, W. Aalst, A. Medeiros: *Process Mining with the Heuristics Miner-algorithm*, BETA Working Paper Series, WP 166, Eindhoven University of Technology, Eindhoven, 2006.

[15] M. W. Wibisono, A. P. Kurniati, G. A. A. Wisudiawan: *Process Mining using Inductive Miner Algorithm to Determine the actual Business Process Model*, JURIKOM (Jurnal Riset Komputer) (2022), doi: 10.30865/jurikom.v9i4.4769.

[16] W. Zaitsu, M. Jin: *Distinguishing ChatGPT(-3.5, -4)-generated and human-written papers through Japanese stylometric analysis*, PLOS ONE 18.8 (2023), e0288453, doi: 10.1371/journal.pone.0288453.