ROMANIAN JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY
Volume 28, Number 4, 2025, 365–376
doi: 10.59277/ROMJIST.2025.4.05

AI Texts under the Etymologic Lens

Mike H. M. TEODORESCU[1], Cecilia Speranța BOLEA[2], and Horia-Nicolai TEODORESCU[3,*]

[1]University of Washington, Seattle, WA, USA
[2]Institute of Computer Science of the Romanian Academy, Iași, Romania
[3]Technical University "Gheorghe Asachi" of Iași, Iași, Romania
Email: hmteodor@post.harvard.edu,
cecilia.bolea@iit.academiaromana-is.ro, hteodor@etti.tuiasi.ro*
* Corresponding author

Abstract. AI generated texts seem human-written and having a similar quality to that of texts produced by humans. The etymologic analysis of the vocabulary of texts is proposed for examining the complexity of AI generated texts. The etymologic analysis shows that, when asked to emulate an author, AI appears to use an etymologic mixture of the vocabulary less complex than that of the human literary texts. When asked to recount a specific literary work emulating its author, AI produced writings with an etymologic distribution far from that of the narrated text and with aberrations, throwing doubt on AI ability of emulating a specific author at the etymologic level. The tool of etymologic analysis may help distinguish between AI and human generated texts.

Key-words: AI; AI author emulation; AI texts; AI text detection; etymological distribution; etymology comparison; etymology of AI texts.

1. Introductory Matter

There is a fast increasing literature comparing AI capabilities with those of humans, in various domains of endeavor, for example [1–3]. We analyze a facet that has not been addressed, namely the ability of using words with various etymologies in texts with the purpose of creating texts with atmosphere and specific styles.

Etymology is the science of discovering the origin and the evolution of words of a language. Words of a modern language may have originated from one of the ancestral languages that were spoken by the ancestors of the current population speaking the specified language.

Most frequently, a modern language has several ancient "parents". Often, languages have been penetrated by words and cultures from migrating populations that stayed for some time over a territory and then moved away. Other words were borrowed from neighboring languages or even over a distance, when a strong cultural influence was exerted for a long period.

Probably, few if any people think to etymology when they speak or write. As a consequence, one may conjecture that words with various etymologies are used indiscriminately, at random. If so, one may hypothesize that long texts, such as novels, should have a distribution of etymologies coinciding with the distribution in the language at large. These hypotheses prove wrong, as it was found during an analysis of literary works [4–6]. Those studies demonstrated that authors of large literary works use words with quite different etymologic distributions. A Romanian author that literary critics consider to be among the most skilled and to have a poetic style in his novels was found to use a distribution of etymologies much different from the distributions determined for several other authors. It has been proposed that specific etymologic distributions of the vocabulary used in literature works plays a stylistic and literary role [4], [5].

There is no doubt that AI is able to generate interesting texts on given topics, without copying from the extant literature. A text produced by AI is derived from the Internet literature and may seem to be new and informative. A superficial analysis may find that AI behaves as a truly independent, original author. We suggest that it may be worth analyzing the etymology distribution of AI generated texts when AI describes a literary work, or when AI plays the role of a specified writer, and compare it with the distribution for the described literary work. Considering that AI is trained over a large literature and with works of many authors, a first hypothesis is that AI texts may have, on average, the same etymologic distribution as the respective language, as determined for large corpuses. A second hypothesis is more elaborated and stems from the idea that, when asked to represent a literary work by recounting it in a manner close to the human author of that work (AI role-playing), AI will mimic the etymological distribution of the literary work or of its author – assuming that the respective author used a specific etymological distribution.

We have performed two tests to check the above hypotheses. We used the free version of ChatGPT AI in all tests. The several aspects required for a thorough testing are described in the next Section. Section 3 exposes the main results, while the last Section discusses the results and derives conclusions.

2. Method

2.1. The literary work used in the tests

For the tests for AI, a novel was chosen. In two sessions, each with several steps, AI (ChatGPT) was asked to present (first session) and recount (second session) the novel *La Medeleni*, volume Hotarul Nestatornic (At Medeleni – The Unstable Boundary, also translated as Fickle Border) [7]. This is a renown Romanian novel and has a large public even after 100 years since publication. The novel is the first in a trilogy, was initially published in 1925, and is written by Ionel Teodoreanu. It is easy and pleasurably to read and was continuously included in the high school curricula in Romania. In addition, there are innumerable articles about this novel over the Internet, including one on Wikipedia, several articles by literary critics, and many others written by teachers of Romanian literature and by students. There is even a Romanian word, medelenism, for describing the atmosphere in this novel.

The novel is reputed in the Romanian literature for its style, with the author using many metaphors and other literary devices. We hypothesized that AI will recount the work in a similar style, although we recognized the risk that the style of the novel may create difficulties for AI tools. The work has an etymological distribution that is not alike with other works published in the same period [4], [6]; that was found to be even more true when the etymologic distribution of the novel is compared with various contemporary works. This distinctive feature was essential in the choice of the novel as benchmark for the AI ability to learn a specific literary use of the language and to produce texts that mimic a literary work. The language of the novel is not distorted by regionalisms or archaisms. The etymology of the novel has been analyzed and discussed in [4]; the results in [4] were used in this study for comparison with the AI texts.

2.2. Text processing

The text processing was presented in [6] and [8]. The parsing of the text was performed with RACAI parser [9] and the etymology was determined automatically based on the online Dexonline dictionary [10]. For the etymology search, the scraper [11] was used. Manual corrections have been necessary because the novel text used includes words with older orthography, unrecognized by the parser, because some imperfections of the parser, and because AI introduced a few non-extant words [6]. For each text, a list of lemmas, with their etymology or etymologies, and with their number of occurrences was generated. The etymologies of the lemmas were determined based on their parts of speech (PoS) tags; the same lemma may have different meanings and different etymologies when the PoS differs. For example, in Romanian, *toca* may represent a verb with several meanings, derived from the Latin toccare (to chop), or may represent a noun, with French etymology (toque, a type of hat). Similarly, *haina* may be adjective (merciless) with Turkish etymology, or a noun (cloth), with triple etymology (Bulgarian, Serbian, and Croatian). The lists of lemmas were used in the statistical analysis. The multiple etymologies complicate the statistical analysis, as discussed in [6]. In this study, we counted each of the etymologies of a lemma, when there are several; therefore, relative frequencies were determined by dividing the total number of occurrences of a lemma with a specified etymology to the total number of lemmas, where lemmas with several etymologies are counted the respective number of times; details are given in [6].

2.3. The first set of AI tests

In all tests, the web application ChatGPT (https://chatgpt.com/) was used, with no log-in, with no advanced features, such as Canvas, and with accepting only the necessary cookies. While this is the basic variant of using ChatGPT, it is also probably the most used by school students; that may be important for some of the conclusions in this study. The content analyzed (reproduced in the online Appendices at http://iit.academiaromana-is.ro/) was generated at the first prompt; no question addressed to AI was repeated.

The main questions in the first tests can be summarized in English as: “Please, I need a 3-4 pages presentation of the action of the novel La Medeleni, first volume, The Unstable Boundary”. The answers of ChatGPT were quite schematic, imitating the level of a typical description by school students over the Internet. AI interpreted the word “presentation” in the questions as meaning “discussion” and “explanation,” instead of simply showing what happens in the novel. Even the structure of the answers follows that of a typical school homework. (AI texts provided as answers during the first test are included in the online Appendix.)

Due to the somewhat disappointing first answer, the next step was to explain more clearly what is asked for. The story provided was, in our opinion, of the level of a student not yet in high school. Therefore, we insisted asking for a recounting, at a college level. The result was a description by AI that used some less usual words, but was too short, on about two pages. We changed tactics and asked AI for a description for each chapter, with the purpose of obtaining a longer text. The answer has been surprising in that AI found it necessary to explain its approach "... Below you will find a detailed recounting of the first volume of the novel ... structured on chapters." Notice that we have not asked for a "structured" answer. On the other hand, this is, indeed, a requirement for students. Yet, these specificities made the text schematic. The texts generated in this testing session have been analyzed etymologically as three texts apart and then cumulatively. The relatively limited extent of the answers is at least partly attributable to the use of the ChatGPT web free application, which is known to have limited possibilities.

2.4. The second set of tests

For the second tests, we asked ChtGPT to "recount in author's words" the novel. The first answer was not enough detailed. Then, we asked a recount of the first chapter, next the second, the third, and the fourth. After that, AI suggested itself to continue recounting new chapters. We have concatenated the answers and obtained the second AI text for analysis (see online Appendix). The etymological analysis for the texts resulted from the first and the second test are given in Section 3.

2.5. Manual analysis of the vocabulary and parsing errors for AI

The AI text generated in the first test raised few issues during the parsing, with few errors, many related to proper nouns. The large number of parsing errors encountered for the text obtained during the second test imposed a manual analysis of the generated text, of the correctness of the sentences and of the vocabulary, and of the meanings of the sentences. Several types of parsing errors in the AI texts have been discussed in [6].

3. Etymology of AI Texts Recounting a Literary Work and Comparison with the Literary Text

3.1. Distribution of the etymologies in the novel *La Medeleni* (LM)

The etymology distributions of the novel and of the AI generated texts were determined and compared. The novel has a distribution of etymologies of the independent lemmas as shown in Fig. 1. Independent lemmas are lemmas with a unique meaning and a unique PoS role; they are counted once, without repetitions. When repetitions of the independent lemmas are also considered, the respective number is named total number of lemmas.

In the article, the following abbreviations are used: Alb/cognate are actually cognates, but named in some sources Albanian, because they occur also in Albanian; Sv means Old Slavic, Mgh means Magyar, Bg is Bulgarian, Srb – Serbian, Ge – German, Tc- Turkish, Ono – onomatopoeias, NGr – Neo-Greek, Fr – French, Ukn – unknown etymology, Ukr – Ukrainean, Pl – Polish, Ru – Russian, and Cr – Croatian. LM1 designates the first volume of *La Medeleni*.

In Fig. 1, the etymologies of the most frequent 500 lemmas in *La Medeleni* are shown (without stopwords, which in Romanian are all of Latin origin). The dominant etymology is Latin, followed by Old Slavic and French. The data in this subsection correspond to the study [4]. The domain of etymology continuously evolves and the source we used for determining the etymologies, Dexonline [10], is updated. It is possible that minor variations of the figures reported here may appear when the etymologies are determined with the current or future version of Dexonline.

The use of the independent lemmas is given by the number of their occurrences, i.e., their absolute frequency in the text. Fig. 1 shows the number of occurrences of the lemmas with specified etymologies in the text LM1. Fig. 2 corresponds to the respective total numbers.

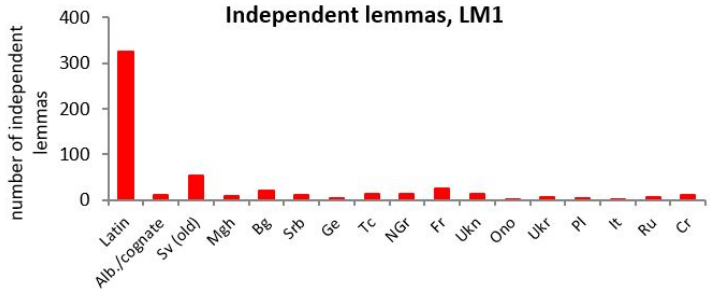

Fig. 1. Number of independent lemmas per etymology (out of the most frequent 500 ones) in *La Medeleni*, first volume (LM1).

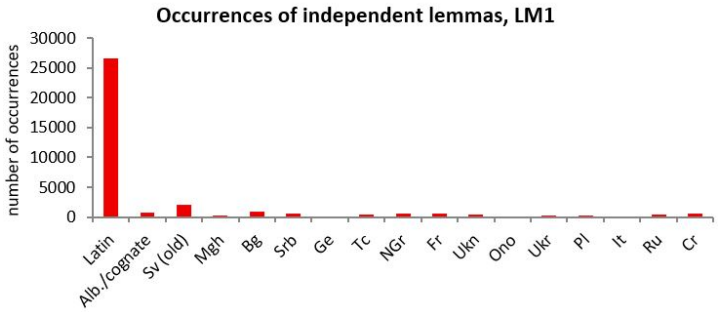

Fig. 2. Total number of occurrences of independent lemmas (out of the most frequent 500 ones) in *La Medeleni* (LM1).

The corresponding Zipf distribution of lemmas is given in [4], with the representation convention that each rank corresponds to an independent lemma, with lemmas having the same count being attributed a random order among them. The exponent in the law is $a = -0.900$ and the intersection is at $A = 8.33$, with a determination coefficient $R^2 = 0.996$ of the predicted line.

For a correct understanding of the differences between the novel *La Medeleni* and the version produced by AI, one needs to appreciate the differences between this novel and other literary works, preferably written during the same epoch and having similarities in the topic. We illustrate these differences using two other works that have self-biographic nature and are written in the same period, *Memoirs of War* (MoW), by Al. Averescu, and *Under Three Kings* (UTK), by N.

Iorga. The differences in etymologies are shown in Fig. 3. Clearly, LM is richer in words of Old Slavic origin, as well as in words of Turkish, Bulgarian, Serbian and Croatian origin. In contrast, it is poorer in words of French, Italian, and German etymologies. Therefore, the etymologic distribution of LM is not due to the epoch when it was written, nor to its self-biographic tendency, which are common for the three works discussed (LM, UTK, MoW). The χ^2 test applied to the etymological distributions of these works shows with a high degree of confidence that they differ. For details on the etymology of the three works see [4].

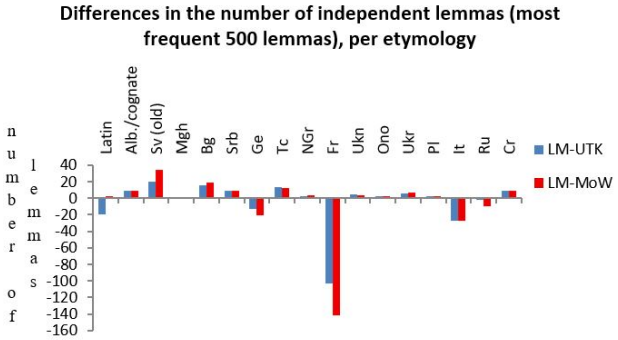

Fig. 3. LM compared, at the number of independent lemmas per etymology, with two self-biographic works from the same period, UTK and MoW (based on [4–6]).

Interestingly, the l-etymologies, where an l-etymology is defined as the total number of occurrences with a specific etymology, l, are distributed in LM approximately according to Zipf law, see Fig. 4. More details on the etymologic distribution of LM are found in [4–6].

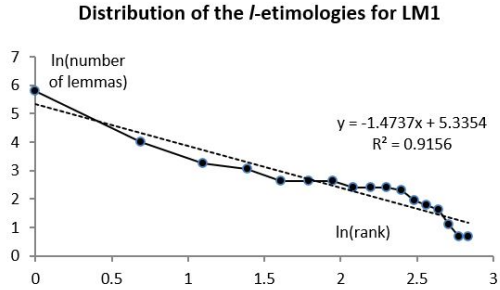

Fig. 4. Zipf's law for the l-etymologies, where l-etymology is defined as the set of words of specified etymology l.

3.2. Distribution of the etymologies in the AI description of the novel *La Medeleni* by AI (first test)

The comparison of the AI description of the novel LM1 is not made at the content level, which AI typically is able to reproduce well based on description over the Internet, but at the etymological level, which is not considered in any algorithm, except [6]], as much as we know.

The etymologies of the independent lemmas in the first three answers (test #1) of AI are shown in Fig. 5, 6, and 7.

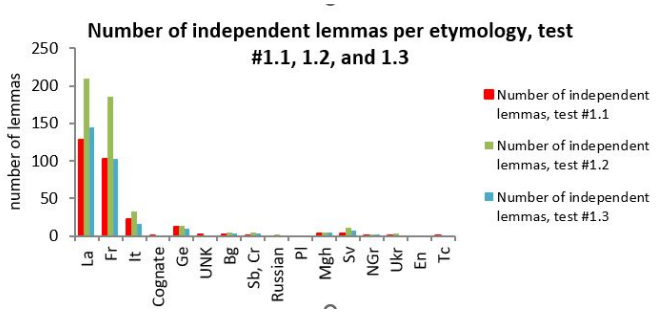

Fig. 5. Comparison of the etymological distributions for the independent lemmas in the texts corresponding to the first set of tests (tests 1.1, 1.2 and 1.3).

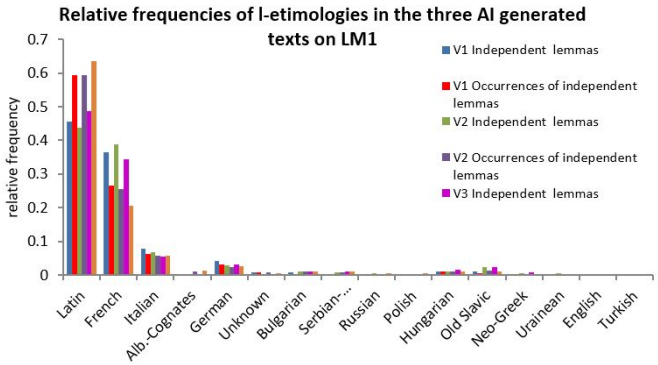

Fig. 6. Overall data, including both the number of independent lemmas and the number of occurrences of the independent lemmas in the three texts produced during the first test.

When the numbers of independent (without repetitions) lemmas of specified etymologies from the three versions of the first tests are added, the resulted distribution is as in Fig. 6. While the number of independent lemmas is relatively small (about 300) for each of the texts in test #1, that is, tens of times less than for the novel, the numbers are large enough for determining with high confidence, using χ^2 test or Fisher exact test, if the etymologic distributions in AI texts and in the novel are identical.

Notice in Figs. 6 and 7 that the lemmas of French etymology are almost as numerous as those for Latin, that lemmas of Italian and German etymologies are on the third and fourth places, and that the lemmas of Old Slavic etymology come only in the fifth position. In the second text of test #1, one word of English etymology occurs (none is in LM). The distinction between the etymology distribution in the first test (all versions independently or cumulated) and the distribution in the novel *La Medeleni* is clear. When reading the three texts authored by ChatGPT, one is able to perceive a difference in vocabulary from that of the novel; yet, the differences shown at the numerical level are still surprising.

There is also a clear difference in the distribution of the occurrences of the lemmas in the texts

generated by AI on the novel and the distribution of the lemmas in the novel: the corresponding approximate Zipf laws have exponents much different, -0.9 for the novel [6], and -0.69 for the AI text, see Fig. 8, which is a large difference for the exponent in Zipf's law.

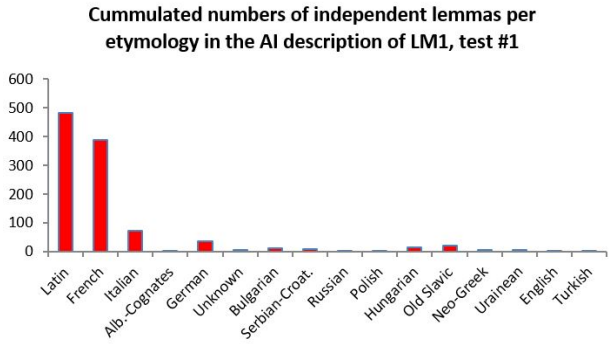

Fig. 7. The (cumulated) numbers of independent lemmas per etymology for the AI texts (first text of test 1).

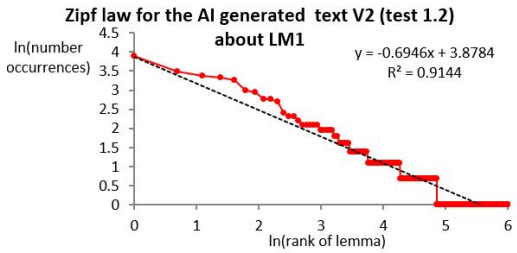

Fig. 8. Zipf's law for the AI generated text V2 (second text in test 1).

The χ^2 test of independence, which is frequently used for comparing distributions, when applied to the etymological distributions of the AI text and of the novel, accepts the hypothesis of identical distribution with a confidence of less than 0.1% ($\chi^2 = 599$, while $\chi^2 = 22$ at $p = 1\%$), we created the category "others," where the lemmas having etymologies with less than or equal to 5 occurrences have been included. This was necessary because one of the conditions for applying χ^2 tests is that of having at least five objects in each category, or at least that the "expected frequencies should be at least 5 for the majority (80%) of the cells" [12]. Another condition of valid use of the χ^2 test is the independence of 'observations'. When the 'observations' refer to distinct lemmas, the independence is not guaranteed. However, here 'observations' refer to the use of different etymologies, they may be considered independent. The samples compared are large enough for the χ^2 test is validly applied. The reason of using the 500 most frequent lemmas in the novel for the etymology comparison consists in that, in a larger text, the most frequent lemmas typically correspond to the whole set of lemmas in a much smaller text.

3.3. Distribution of the etymologies in the AI recounting of the novel *La Medeleni* (second test)

The second test was specifically conceived to assess the ability of AI of recounting an oeuvre in author's words, that is, supposedly, using a similar style and vocabulary, where the meaning of similar vocabulary covers also the etymological distribution. As AI was imposed to generate a text large enough, the error in assessment should be lower than for the first test. The number of independent lemmas in this text is 997 and the overall total number of occurrences is 3668. With respect to the number of occurrences, see Fig. 9, notice that the words of French etymology are almost three time more numerous than those of Old Slavic origin, 387 versus 134. The difference between the weight of the independent lemmas of French etymology and the weight of Old Slavic lemmas is even greater, 266 versus 62, Fig. 10. The number of independent lemmas of French etymology is more than 50%, compared to the number of independent Latin lemmas (266 vs. 482). Still, occurrences of Old Slavic lemmas are more numerous than German and Italian ones.

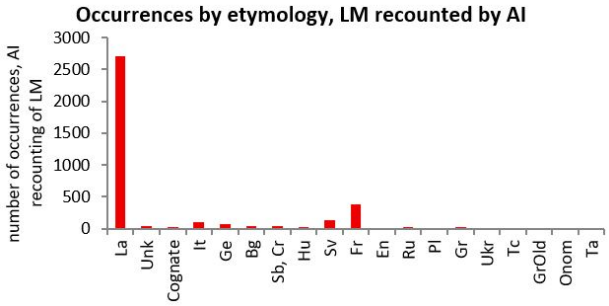

Fig. 9. Etymologic distribution for the occurrences of the lemmas in LM1 recounted "in author's words" by AI, test 2.

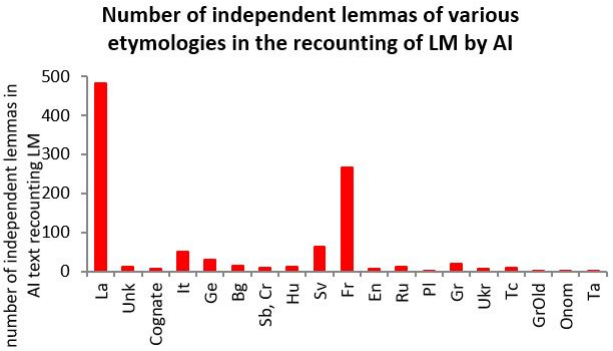

Fig. 10. Etymological distribution of the independent lemmas in the AI text recounting LM1.

A graphical comparison of the original novel *La Medeleni* and its recounting by AI is shown in Fig. 11 in terms of the ratio of numbers of the lemmas of various etymologies to the number of lemmas of Latin etymology, for the novel and for the AI recounted text.

Etymologies, as represented by the number of independent lemmas of specified etymologies, also approximately follow a Zipf law. The exponent is close to 2. The word rank distribution

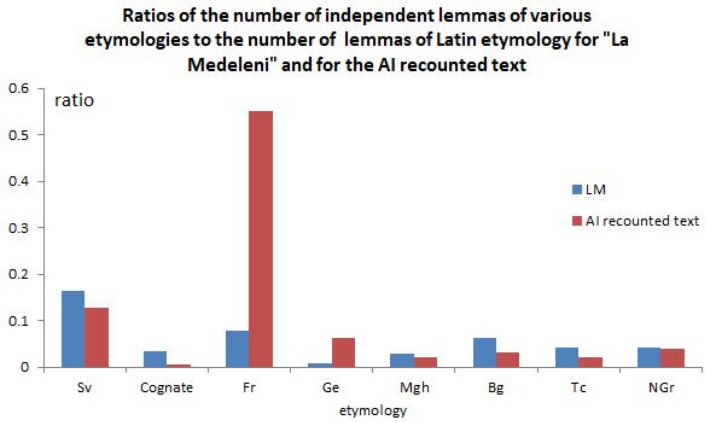

Fig. 11. Comparison of the etymologies of the independent lemmas in *La Medeleni* and in the AI recounted text.

in the AI generated text is, approximately, according to Zipf law, as for human-written large texts, but the exponent in the law differs, indicating a different use of the lexis. Concluding, the etymology distributions of the analyzed AI-generated texts discussing or recounting a literary work were found to be far apart from that of the original, literary text recounted by the AI.

Not only the etymologic distribution of the AI texts differs from that of the novel; in addition, it is found to be much closer to the distribution of some lower level writings discussed in [6]. One could say that AI produced a new type of "langue de bois." But part of the blame may be due to the educational system, to the way the literature oeuvres are presented in class and then are represented by students over the Internet; textbooks may also play a role in the sketchy etymological distribution used by the AI, as the latter is trained possibly on manuals too. Possibly, the etymological analysis could reveal an issue at the deeper level of writing manuals for school, with a schematic language that lacks nuances.

4. Discussion and Conclusions

We asked ChatGPT to recount, "in author's words," a specific Romanian novel that has a particular style and etymological distribution. The results have been unexpected in that AI made several language errors and has used a somewhat primitive language compared with the original. The etymological analysis revealed a different distribution of the etymologies in the AI text, compared with the original. The results suggest that the profound layer of the AI generated text, a layer reflected by the distribution of the etymologies, and the type of vocabulary used, as described by the most used words and by Zipf's law, contrast much from the original text. This may help establishing algorithms that differentiate between texts authored by humans and texts written by AI, in line with [3], [6]. In addition, when not thoroughly driven by the user, AI produces descriptions that are far from what could be expected, with an idiosyncratic and full of stereotypical words in a "wooden tongue" style. This may be an issue with many school students increasingly relaying on AI for learning and for getting a least a hint for their literature and language homework: they may be taught more by AI than by their teachers and may learn schematic models of the native language.

The distribution of etymologies reflects the profound structure of a language. While some of the human writers manipulate this structure with artistical purposes, current AI tools seem unable to do that for now. We suggest that producing texts that are similar to those written by humans may not be enough for AI passing the Turing test in its essence, which involves human-like overall behavior, not attained by AI yet. As Bergstrom and Bak-Coleman [1] say about reviewing,

"even going from initial notes to the final text requires capacities that an LLM simply lacks: our unique perspective, training, values, ethics, domain expertise, understanding of editorial priorities and perceptions of the authors."

Discussing the recent literature on the use of AI in research articles, Mallapaty [2] notices in the abstracts of research articles

"... 'excess words' that started to appear more often than expected after November 2022, when ChatGPT became widely available. ... 454 words appeared much more often in 2024 than in any other year since 2010. They were mostly 'style' words unrelated to the content of the research... Some were common — such as 'findings', 'crucial' and 'potential' — whereas others were more unusual, including 'delves' and 'showcasing'. Excess words that emerged in the second half of 2024 include 'heighten' and 'hinder', ... 'unparalleled' and 'invaluable'" [2].

The increased use of these words is attributed to the use of AI by the respective researchers, when writing their papers. We found a similar pattern, with the word important (same meaning as in English) repeated three times; it is a typical mark of students to add this word and similar ones whenever they are unable to make their essays more worthy in the substance.

However, the above opinions do not contradict the fact that AI is able of marvelous feats, including the generation of novel ideas, as shown by [13], although with some limitations, such as "failures of LLM self-evaluation and their lack of diversity in generation" [13]. For now, an increasing body of literature indicates various aspects that differentiate in a dialogue the machines from humans, including the fact that no human is able to provide answers to questions that cover a huge amount of information. The distinct etymology distribution used by AI may serve today for detecting non-human texts. AI seems to be influenced by the lower level texts on the Internet, which are more numerous and use a linguistically impoverished language. In addition, AI uses a distribution of etymologies that is penurious. Due to the large effect AI has today at the educational level, with the majority of students regularly using AI, the poor AI generated language may become an issue, as the long term effect may be a degradation of the learning and of the languages. This effect may add to the "accumulation of cognitive debt" for those using AI for various tasks [14, 15]. This study suggests a way for improving AI, namely by training it to follow a more natural etymology distribution.

This article may serve as a starting point in a study that should include several stages of analysis. The first stage should extend the examination to a set of authors and of their works, with AI asked to recount (mimic) those works. An extensive comparative analysis is required to check if the results in the present study extend to the AI representation of other literary oeuvres and to other languages. A future statistical study of the etymologies of the answers of AI to the same repeated question(s) might also be interesting. Also, future studies should cover texts generated by several AI, with their full set of available high level tools.

Acknowledgements. The authors thank Acad. Svetlana Cojocaru, Acad. C. Gaindric, Prof. Adriana Vlad, Prof. Mironela Pîrnau, Mr. N. Amalinei, dr. hab. Nina Corcinschi, Dr. Daniela Gifu, and all colleagues who made valuable comments on several preliminary versions of this

study.

Authors' contributions. MT proposed the etymological comparison of AI recounted texts with original works and the comparative statistical analysis of the distributions of the etymologies. CSB parsed the texts, developed the software for assigning the etymology, and performed the etymological data collection. HNT proposed the topic of etymological analysis for texts written by human authors, wrote the first version of the paper, and connected the contributions to this research.

References

[1] C. T. BERGSTROM and J. BAK-COLEMAN. *AI, peer review and the human activity of science*, Nature, Career Columns, 25 June 2025. [Online]. Available: https://www.nature.com/articles/d41586-025-01839-w.

[2] S. MALLAPATY, *Signs of AI-generated text found in 14% of biomedical abstracts last year*, Nature, News, 02 July 2025. [Online]. Available: doi: 10.1038/d41586-025-02097-6.

[3] H. ZHU and L. LEI *Detecting authorship between generative AI models and humans: a Burrows's Delta approach*, Digital Scholarship in the Humanities, DOI: 10.1093/llc/fqaf048, June 2025.

[4] H. N. TEODORESCU and S. C. BOLEA, *A comparative lexical analysis of three Romanian works - The etymological metalepsis role and etymological indices*, Romanian Journal of Information Science and Technology **25**(3-4), 2022, pp. 275–289.

[5] H. N. TEODORESCU, S. C. BOLEA and M. H. M. TEODORESCU, *Use of etymologies as a linguistic instrument in literary works. (Folosirea etimologiei ca instrument lingvistic în operele literare, in Romanian)*, Public presentation, Kishnew / Chisinau, June 16, 2025. Unpublished. [Online]. Available: https://www.youtube.com/watch?v=GO214P7kXCE (accessed July 1, 2025).

[6] M. H. M. TEODORESCU, S. C. BOLEA and H. N. TEODORESCU, *Towards a framework for etymological analysis of literary and AI works*, Proceedings of 29th Annual IEEE High Performance Extreme Computing Virtual Conference (HPEC), Boston, UA, 15 - 19 Sep. 2025.

[7] I. TEODOREANU, *La Medeleni: Hotarul nestatornic*, Editura Cartea Românească, Bucuresti, 1925.

[8] S. C. BOLEA, *An algorithm for automatic analysis of the etymology of Romanian words*, Proceedings of 16th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Iasi, Romania, 2024, pp. 1–5.

[9] RACAI, *Relate Platform*. [Online]. Available: https://relate.racai.ro/.

[10] DEXONLINE. *Dicționar explicativ al limbii române* (in Romanian, Dictionary of the Romanian language). [Online]. Available: https://dexonline.ro/.

[11] DEXONLINE-SCRAPER, [Online]. Available: https://github.com/vxern/dexonline-scraper. MIT License.

[12] K. YEAGER, *SPSS Tutorials: Chi-Square Test of Independence*, July 9, 2025. [Online]. Available: https://libguides.library.kent.edu/spss/chisquare.

[13] C. SI, D. YANG and T. HASHIMOTO, *Can LLMs generate novel research ideas? A large-scale human study with 100+ NLP researchers*. arXiv:2409.04109. [Computation and Language], 6 Sep. 2024. [Online]. Available: https://arxiv.org/abs/2409.04109.

[14] N. JONES, *Does using ChatGPT change your brain activity? Study sparks debate*, Nature. 2025 July;**643**(8070):15–16. doi: 10.1038/d41586-025-02005-y.

[15] N. KOSMYNA, E. HAUPTMANN, Y. T. YUAN, et al., *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task*, arXiv:2506.08872v1 [cs.AI]. [Online]. Available: https://doi.org/10.48550/arXiv.2506.08872. 10 June 2025.