Computer-support for the analysis of rhetorical devices in texts

Ștefan Trăușan-Matu
University Politehnica of Bucharest
313 Splaiul Independentei,
Bucharest, Romania
and
Research Institute for Artificial Intelligence
and
Academy of Romanian Scientists
stefan.trausan@upb.ro

Adrian-Dinu URSE
University Politehnica of Bucharest
313 Splaiul Independentei,
Bucharest, Romania
adrian_dinu.urse@stud.acs.upb.ro

ABSTRACT
In this paper we present the background ideas and theories of an implemented prototype of an artificial intelligence, natural language processing application that takes a text as input and outputs the analysis of several classes of repetitions, which are types of rhetorical devices. The application is targeted for students, for researchers, and for any person that aims to improve writing or oratorically skills. It was used for analysing chapters from two well-known novels and speeches, the paper containing also excerpts of some of the results.

Author keywords
Artificial intelligence; natural language processing; text analysis; rhetoric devices; phonaesthetics

ACM Classification Keywords
I 2.7 Natural Language Processing. Text analysis

DOI: 10.37789/ icusi.2024.14

INTRODUCTION

Speech and text are a major interaction manner among humans and computers, having also an important influence on our way of thinking [1]. In the nowadays context of artificial intelligence (AI), generative systems such as ChatGPT [2] seem able to produce any required text and their performances might induce the idea that it is obsolete to study literature and other humanistic disciplines, to learn how to write texts, and to read stories and books authored by humans.

However, this paper has as a major motivation that humans should still study humanistic disciplines, for example, learning and practicing literary analysis, and that AI can help these endeavours. Moreover, humans should be encouraged to write texts in a personal style, that express their personality and their specific human features: empathy and creativity [3]. Texts are communication means, which, in addition to reflect one's style and personality, they can create a soul-to- soul connection. Texts should be enjoyed by the readers, they should provide an aesthetical experience, and in some cases, they should have an influence and even have a persuasive force, which is the object of study of rhetoric.

Very important features of a text are discourse coherence [4], ease of reading, and attractivity (aesthetics). These features are related to music attributes, between text and music being strong relations [5]. Musicality of texts is important for practical reasons, considered in two related domains: rhetoric – how to have an impact on the reader/listener using rhetorical devices [6] - and aesthetics – how to generate enjoyment to the reader/listener - J. R. R. Tolkien, the author of The Hobbit and The Lord of the Rings being credited to have used the term phonaesthetics [7].

Inducing a musical dimension in texts needs the presence of rhythm and, in complex writings, polyphony [8]. There are many examples of polyphonic texts, including novels such as those of James Joyce, Anthony Burgess, and others [9]. Rhythm and polyphony are based on various kinds of repetitions, which are a particular case of rhetorical devices.

Using rhetorical devices is beneficial in writing, "they can help you communicate your ideas more clearly and coherently. By using devices such as repetition, anaphora, and epistrophe, you can emphasize your main points and create a sense of unity and structure in your writing … they can help you enhance your style and creativity … use alliteration to create a sound effect, to emphasize a word or a phrase, or to make a catchy slogan." [10] Therefore, learning how to use rhetorical devices is important not only for authors of literature but also for academic writing, for politicians, journalists, advertisements, and not the least, for fostering creativity in various domains.

In this paper we present a prototype of an AI, natural language processing (NLP) application that takes a text as input and outputs the analysis of several repetition-based rhetorical devices. Such text analysis applications based on AI techniques can aid students, researchers, and any person wishing to improve her writings to both comparatively analyse texts and learning how to write texts with a higher impact and a human touch.

```
 - With this faith we will be able to work together, to pray together, to struggle together,
 to go to jail together, to stand up for freedom together, knowing that we will be free one d
ay.
```

Figure 1. Rhetorical repetitions in Martin Luther King Jr.'s 'I Have a Dream' speech

The paper continues with a state-of-the-art section followed by the presentation of the application and of results of the analysis of two fragments from two novels, with different degrees of using rhetorical devices.

STATE OF THE ART

Rhetoric, the study of writing or speaking as a means of communication, of influencing or persuasion readers or listeners, is a major subject in many contexts: law trials, advertisements, politic discourses, and literature. See in Figure 1 an example from a famous rhetorical speech. This domain is used in NLP for several purposes, such as generating coherent and not boring (attracting) texts (including phonaesthetics), summarization, identifying authors, and generating text with a particular style.

In NLP, an important theory, which has a major role in assuring a coherent and attractive text is the Rhetorical Structure Theory (RST) [4], which identifies a repertory of rhetorical structures that link in various ways text segments. In addition to RST, we consider that another related subject of study and implementation in NLP should be the rhetorical devices, which are also directly related to music. There are extremely few cases of NLP systems that can generate simple examples of some of rhetorical devices [11, 12], some work was done on identifying them in texts, but we did not identify systems that can visualise their presence in texts for research, learning, and training purposes.

Rhetorical devices can be classified in three classes:

1. sound-related (considering rhythmic or phonetic sound in words or phrases, for example alliterations, assonances, consonances, and sibilances),
2. order-related (obtained by a specific ordering of words in phrases or sentences, for example, anaphoras, anadiplosis, epistrophes, etc.),
3. meaning related, for example, metaphors, metonymies, hyperboles, etc.

In the implementation of the present paper only the first two classes were considered, which are related also to music. Here are definitions of the main rhetorical devices implemented, which can induce phonaesthetics in texts.

- Anaphora - a word or a group of words repeated at the beginning of two or more successive phrases, adding emphasis and/or unity.
- Anadiplosis - the last word of a clause or sentence is repeated as the first word of the clause or sentence.
- Polyptotons - repetitions of words with the same root, identified by NLP stemming the words [13].
- Epistrophe - a word or a group of words at the end of two successive sentences are the same.
- Epizeuxis - words or short phrases that are repeated in succession with no other words in between.
- Alliteration – repetition of the first sounds of words within sentences.
- Assonance - repeated vowel sounds within a group of words.
- Sibilance - repeated hissing sounds, such as 's', 'sh', 'z', 'ch', 'x', within a group of words.
- Consonance - repeated consonant sounds within a group of words, excluding common suffixes, such as “ing”, ”ed”, “es”, “s”.

IMPLEMENTATION AND RESULTS

The implementation of the application has been done in Python, using the Natural Language Toolkit Library (NLTK) [13]. Some preprocessing steps were performed: conversion to lowercase, punctuation removal, stopwords elimination, tokenization, stemming. The CMU Pronouncing Dictionary
[14] was used for the sound-related rhetorical devices. The maximum number of words in a group that was checked was 4.

Several genres of text were analysed: chapters of novels, speeches, and text generated by ChatGPT. The most representative text analysed was the Sirens chapter from the Ulysses novel written by James Joyce (awarded with the Nobel Prize), which is widely recognized as a musical and even polyphonic text. This novel is considered as one of the outstanding novels of the 20th century. The Sirens chapter in Ulysses is widely considered as containing many devices related to musicality. Even James Joyce himself said of this chapter that it is like a “fuga per canonem”. A fugue is a polyphonic discourse with origins in music in which independent several threads of ideas/concepts/words (“voices” in music) occur in parallel entering in divergences/dissonances and convergences/ consonances, counterpoints inducing creativity and a sense of life [8].

Numbers\Texts	Sirens	A Tale of Two Cities
Characters	70610	50182
Anaphora	48	12
Anadiplosis	23	1
Polyptoton	29	18
Epistrophe	45	10
Epizeuxis	48	8
Alliteration	1413	1625
Assonance	2115	2140
Sibilance	591	462
Consonance	2652	2317

Table 1. Statistics of rhetorical devices

In addition to the Sirens chapter, a fragment of the “A Tale of Two Cities” novel by Charles Dickens was analysed together with some speeches and a text generated by ChatGPT.

A comparison of the number of rhetorical devices found in the two fragments from novels is presented in Table 1. As seen, the Sirens chapter has, with two exceptions, more rhetorical devices than the fragment of Dickens. The differences are significant in the case of order-related rhetorical devices. However, it should be mentioned that the dimensions of the two texts are different, the first one having more characters.

Excerpts of four categories of rhetoric devices detected by the implemented application for the Sirens chapter of Ulysses are shown in Figure 2. In Figure 3 are shown three categories from Dickens' A Tale of Two Cities.

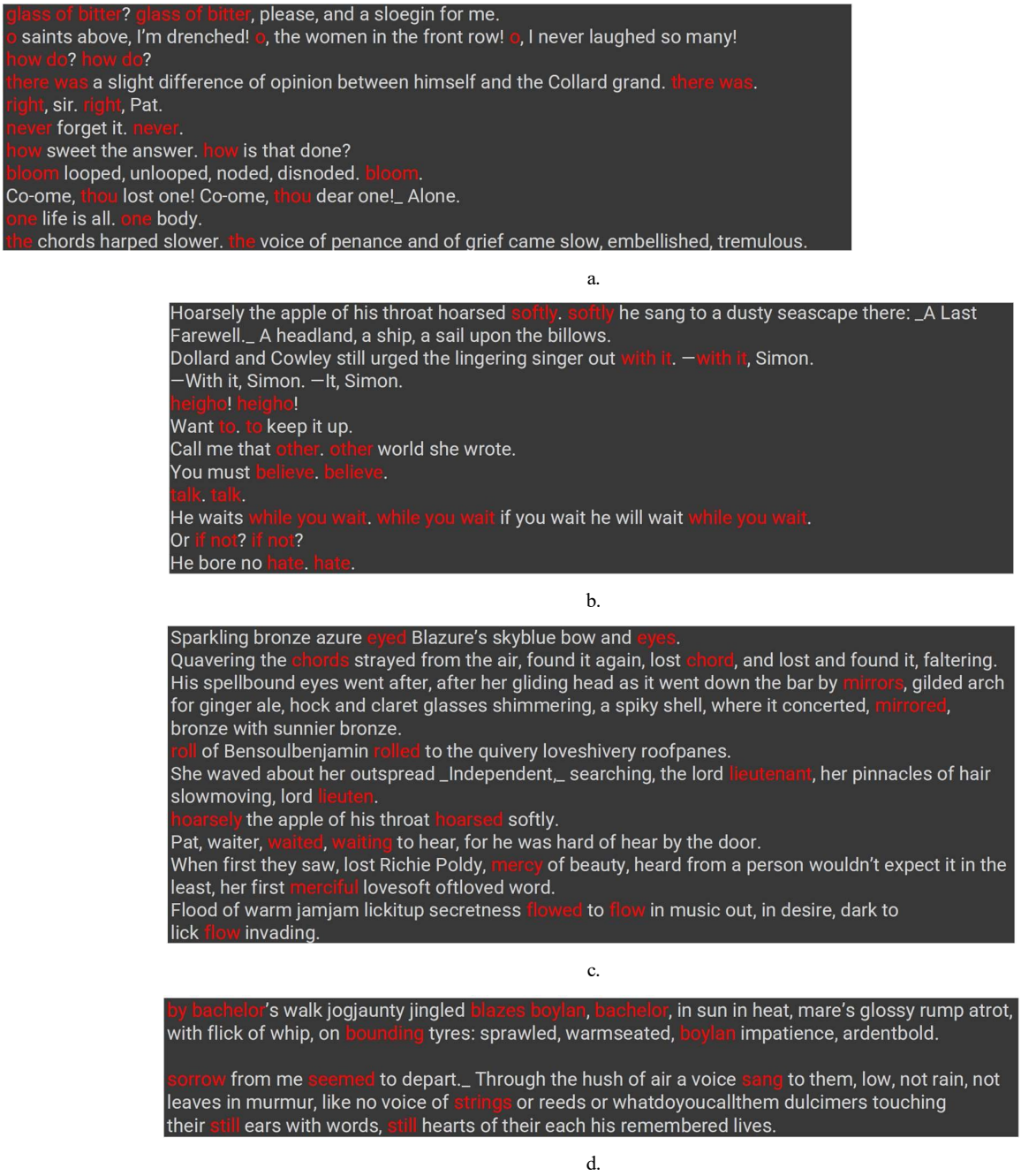

Figure 2. Examples of rhetorical devices in Sirens (a. anaphora, b. anadiplosis, c. polyptoton, d. aliteration)

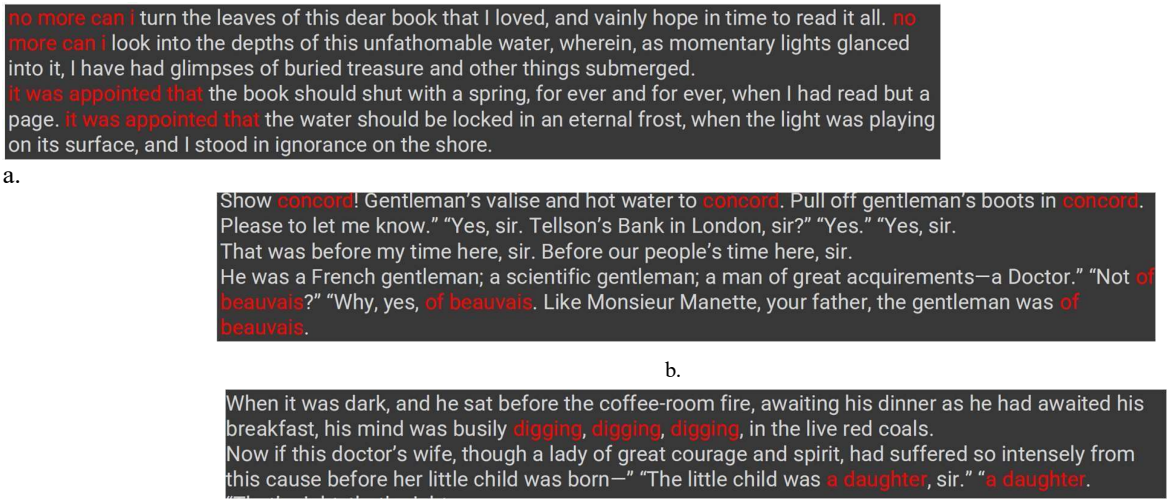

Figure 3. Examples of rhetorical devices in A tale of two cities (a. anaphora, b. epistrophe, c. epizeuxis)

CONCLUSIONS

The paper discussed the importance of the education and training of humans to write texts that contain a personal style, to use rhetorical devices, to promote creativity and musicality, in the context of the automatic text generation of ChatGPT-like applications. For supporting such aims, the prototype of an AI NLP application was implemented that identifies and visualize rhetorical devices in texts. This application is targeted for students, for researchers, and for any person that aims to improve writing or oratorically skills. It was used for analysing chapters from two well-known novels and speeches. The presented research will be extended with in the direction of considering the musicality of texts [5], including polyphony [3].

REFERENCES

1. Lindley, K.E. How words influence thought (and decisions), *Medium*, (2020) https://medium.com/@CoachDeveloper/how-words-influence-thought-and-decisions-65f89c2662f8, accessed on 1 July 2024
2. OpenAI (2023) GPT-4 Technical Report. *arXiv:2303.08774*
3. Trausan-Matu, S., Hossu, A. Analysis of Polyphonic Interanimation in Novels, a Support for Understanding Beyond ChatGPT Answers. *International Journal of User-System Interaction* 16(1), (2023), 11-24.
4. Jurafsky, D. & Martin, J.H. Speech and Language Processing, 3rd edition draft, (2024) https://web.stanford.edu/~jurafsky/slp3/, accessed on 1 July 2024
5. Trausan-Matu, S. Connections between human music and natural language, in Milena Slavcheva (ed.) *Knowledge, Language & Models*, INCOMA Ltd. Shoumen, , ISBN 978-954-452-062-5, (2020), 302-317.
6. ***, What James Joyce Taught Me About Rhetorical Devices,https://strangeandcuriousthings.blogspot.com/2015/11/w hat-james-joyce-taught-me-about.html, accessed on 1 July 2024
7. Nordquist, R., Phonaesthetics, https://www.thoughtco.com/phonaesthetics-word-sounds-1691471, accessed on 1 July 2024
8. Trausan-Matu, S., The Polyphonic Model of Collaborative Learning, in Mercer, N., Wegerif, R., & Major, L. (eds.), *The Routledge international handbook of research on dialogic education,* Routledge, New York, NY (2020), 454-468.
9. Clarke, J. The 'Baroque Weaving Machine': Contrasting Counterpoint in James Joyce and Anthony Burgess. In Carpentier, M.C. (eds) *Joycean Legacies*. Palgrave Macmillan, London, (2015), 127–149, https://doi.org/10.1057/9781137503626_8
10. ***, https://www.linkedin.com/advice/3/what-benefits- using-rhetorical-devices-your-writing, accessed on 22 August 2024
11. ***, Figures of Speech. Spice up your language. Generate figures of speech with AI, https://classx.org/ai-figure-of-speech-generators/, accessed on 1 July 2024
12. ***, https://www.ai4chat.co/pages/alliteration-generator, accessed on 1 July 2024
13. ***, Natural Language Toolkit, https://www.nltk.org/, accessed on 1 July 2024
14. ***, http://www.speech.cs.cmu.edu/cgi-bin/cmudict, accessed on 1 July 2024