Automatic Detection of Rhetorical Devices in Science Policy Articles

Joshua Crotts, Dr. Nancy L. Green
Department of Computer Science

ABSTRACT:

Science policy articles frequently include rhetorical devices that authors utilize to influence the reader's opinion, whether it be for or against the topic at hand. Manually analyzing rhetoric in these articles is characteristically slow and cumbersome. We are devising algorithms that parse science policy articles to automatically detect rhetorical devices within them. To serve as a benchmark, we have hand-annotated the rhetorical devices in a representative article on the environmental effects of increased ocean acidification [Johnson & White 2014]. First, using a natural language processing toolkit API (NLTK) and Python, we performed a lexical and syntactic analysis of the text. We then implemented an algorithm using the grammatical features to detect several types of rhetorical devices, including parison, antithesis, and positive/negative polarity. The results of the algorithm will be compared to the hand-annotated benchmark to determine the accuracy of the algorithm. Our current and future work is focused on improving the accuracy of the algorithm and extending it to detect other types of rhetorical devices.

INTRODUCTION:

Science policy articles frequently include rhetorical devices authors utilize to influence the reader's opinion, whether it be for or against the topic at hand. Manually analyzing rhetoric in these articles is characteristically slow and cumbersome. Thus, we are devising algorithms that parse science policy articles to automatically detect rhetorical devices within them. To serve as a benchmark, we have hand-annotated the rhetorical devices in a representative article on the environmental effects of increased ocean acidification [1]. To aid the analyst, we have implemented a graphical user interface that displays the rhetorical devices detected by our algorithm (or those that we have manually annotated).

The interface allows the user to search for three different types of rhetorical devices in an article:

- Parison ("corresponding structure in a series of clauses" [6], ex: "He that is to be saved will be saved, and he that is predestined to be damned will be damned" [2]). See Fig. 1 for more examples.
- Antithesis ("the rhetorical contrast of ideas by means of parallel arrangements of words, clauses, or sentences" [4], ex: "Promise her anything, but give her Arpege." [3]). See Fig. 2 for more examples.
- Antimetabole ("reverse lexical repetition" [5], ex: "Ask not what your country can do for you. Ask what you can do for your country." [5]).

In addition to these rhetorical devices, the user can search for particular words or phrases and see statistics such as the average sentence length and number of words in the text (also shown in Fig. 1).

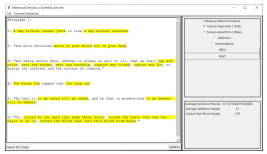


Fig. 1. Parison examples from Nordquist [3]. Note that none come from science-related sources [Nordquist 2019].

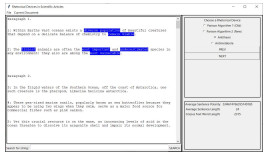


Fig. 2. Hand-annotated XML antithesis examples in Johnson & White [2014].

METHODS:

We created two algorithms for detecting parison. Each has a different way of computing part-of-speech patterns within a sentence. Both algorithms initially tokenize (split) each sentence into a list that contains all corresponding parts-of-speech for each word. The old parison algorithm favors longer instances of parison whereas the new algorithm searches for the most frequently occurring part-of-speech pattern in the sentence. Our new algorithm has a higher chance of returning a false positive compared to the old algorithm but tends to return examples that closely match those in Johnson & White [2014] that we hand-annotated.

Likewise, for antithesis, two algorithms are used, each with a different approach: one utilizes WordNet from NLTK (Natural Language Toolkit [7]), whereas the latter uses a polarity sentiment analysis approach with VADER (Valence Aware Dictionary and sEntiment Reasoner [7]). The former takes two [sub]-sentences and determines if the first has any words that are antonyms of the second, and vice versa. Our second method compares the average polarity of both texts, taking the absolute difference of the two. The idea is that if one section has significantly higher polarity and the other has very low polarity, there is a high probability of antithesis.

Lastly, for antimetabole, first a sentence is tokenized. Then a reversed copy of the tokenized sentence is made. The algorithm highlights those words that appear in the reverse order in the other copy.
In addition, we calculated statistics such as the number of words, average sentence length, and average sentence polarity. The polarity of each sentence was computed using VADER [7].

RESULTS:

We ran all the algorithms on the Johnson and White [2014] article and compared the results to the hand-annotated versions of the article to determine how accurately our algorithms detected rhetorical devices (to determine false-positives and false-negatives). Also we wanted to see how closely the examples of rhetorical devices cited by previous researchers matched those found by our program.

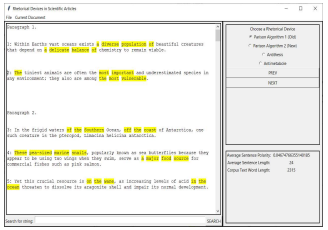


Fig. 3.

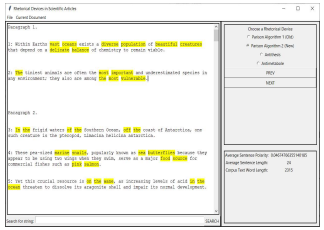


Fig. 4.

Fig. 3 shows the results of the old parison algorithm, whereas *Fig. 4* shows the results of the new parison algorithm. The new algorithm more closely reflects the examples of parison from previous research, shown in *Fig. 1*.

In addition, we ran the antithesis detection algorithm. Our approach was to highlight the entire sentence that has a high probability of being an antithesis example, instead of the specific words in sentence (See Fig. 5). In the original article, we did not hand-annotate antimetabole, but we decided to run the algorithm to see what it would detect as an antimetabole (as shown in Fig. 6). The only instances of antimetabole found by the algorithm were common conjunction words, as opposed to noun-phrases or adjectives. However, when looking at the article, we found very few rhetorically significant instances of antimetabole.

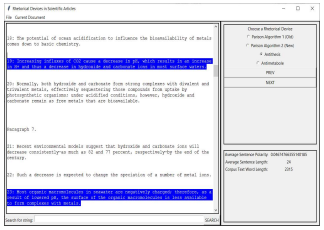


Fig. 5.

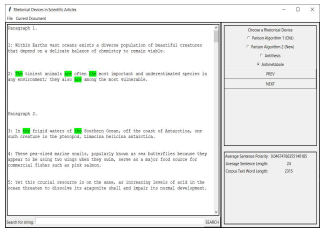


Fig. 6.

Fig. 5 shows the antithesis examples detected by our algorithm are highlighted, whereas *Fig. 6* shows the antimetabole instances detected by our algorithm.

Lastly, we found that our approach to detecting polarity using current tools such as VADER did not meet expectations. Using VADER, a positive polarity (close to 1.0) indicates a positive tone, a negative polarity (close to -1.0) indicates a negative tone. A neutral tone is recorded as close to zero from either side (positive or negative). Our first attempt to recognize the polarity from the Johnson and White [2014] article proved that we could not rely on what it detected as the polarity (0.0467), because it did not match our interpretation of the article as having an overall negative tone.

DISCUSSION AND CONCLUSION:

As we have shown, both parison algorithms produce similar results that when combined will closely match what we have hand-annotated in the Johnson and White [2014] article as rhetorical devices. There are some false positives marked from both algorithms. However, it is clear that the new algorithm better matches the examples of parison found by previous researchers, and the old algorithm finds more instances than the new algorithm.

Future work includes improving our pattern-detection (syntactic patterns specifically) and removing parison examples that are typically marked as false positives from the new algorithm.

There is still progress to be made with our antithesis algorithm; our approach could possibly be refined by changing how antonyms are detected, or by changing the algorithm used to compute the corpus polarity. We are limited by what NLTK determines are antonyms for a specific word, so by using a different resource for identifying antonyms, we may be able to detect antithesis at a higher rate of success.

We plan to not only improve the algorithms we currently have, but to include additional types of rhetorical devices. Readers need to have a clear understanding of how authors attempt to persuade them so they can determine the legitimacy of the authors' arguments.

REFERENCES:

[1] Ashanti Johnson and D. Natasha White. 2014. Ocean acidification: the other climate change issue. American Scientist, Vol. 102, no. 1, p. 60+.

[2] Cooper, James Fenimore. 2017. The Last of the Mohicans: A Narrative of 1757. Ancient Wisdom Publication.

[3] Nordquist, Richard. 2019. Definition and Examples of Parison. Retrieved from thoughtco.com/parison-rhetoric-term-1691577.

[4] R. Brian Stone and Lean Wahlin. 2018. The Theory and Practice of Motion Design Critical Perspectives and Professional Practice. Routledge, 711 Third Avenue, New York, NY.

[5] Randy Allen Harris and Cliff O'Reilly, 2009. Antimetabole and Image Schemata – ontological and vector space models. Angila Ruskinson University, UK., University of Waterloo, Canada.

[6] Randy Allen Harris, Chrysanne Di Marco, Sebastian Ruan, and Cliff O'Reilly. 2018. An Annotation Scheme for Rhetorical Figures, Cheriton School of Computer Science, University of Waterloo, Canada. 155-175.

[7] Steven Bird, Ewan Klein, and Edward Loper. 2009. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit (1st ed.). O'Reilly Media Inc, Sebastopol, CA.

ACKNOWLEDGEMENTS:

We would like to acknowledge support through the UNCG Undergraduate Research, Scholarship and Creativity Office.

Fig. 1

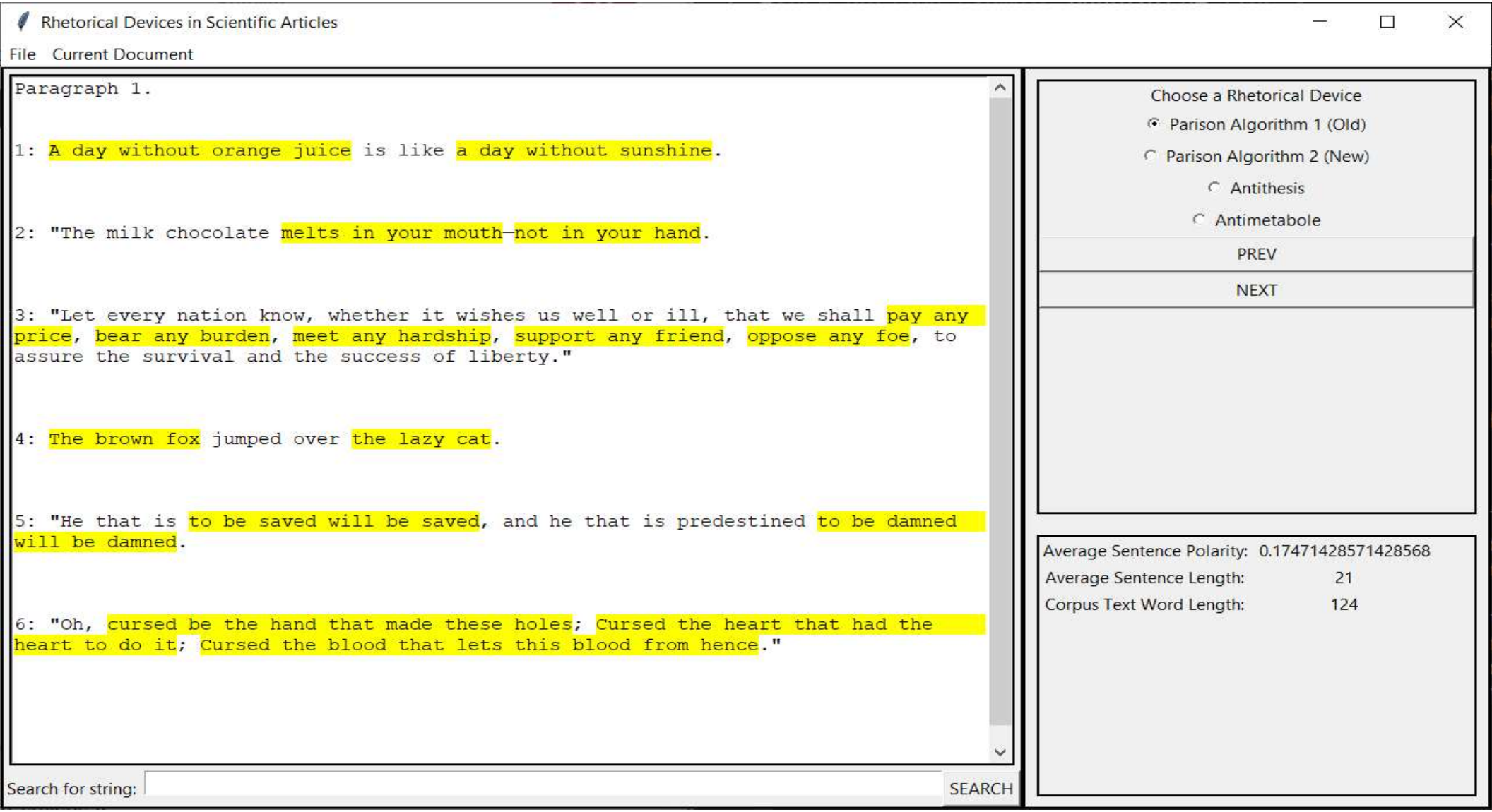

Rhetorical Devices in Scientific Articles
File
Current Document
Paragraph 1.
1: A day without orange juice is like a day without sunshine.
2: "The milk chocolate melts in your mouth—not in your hand.
3: "Let every nation know, whether it wishes us well or ill, that we shall pay any price, bear any burden, meet any hardship, support any friend, oppose any foe, to assure the survival and the success of liberty."
4: The brown fox jumped over the lazy cat.
5: "He that is to be saved will be saved, and he that is predestined to be damned will be damned.
6: "Oh, cursed be the hand that made these holes; Cursed the heart that had the heart to do it; Cursed the blood that lets this blood from hence."
Search for string:
SEARCH
Choose a Rhetorical Device
Parison Algorithm 1 (Old)
Parison Algorithm 2 (New)
Antithesis
Antimetabole
PREV
NEXT
Average Sentence Polarity: 0.17471428571428568
Average Sentence Length: 21
Corpus Text Word Length: 124

Fig. 2

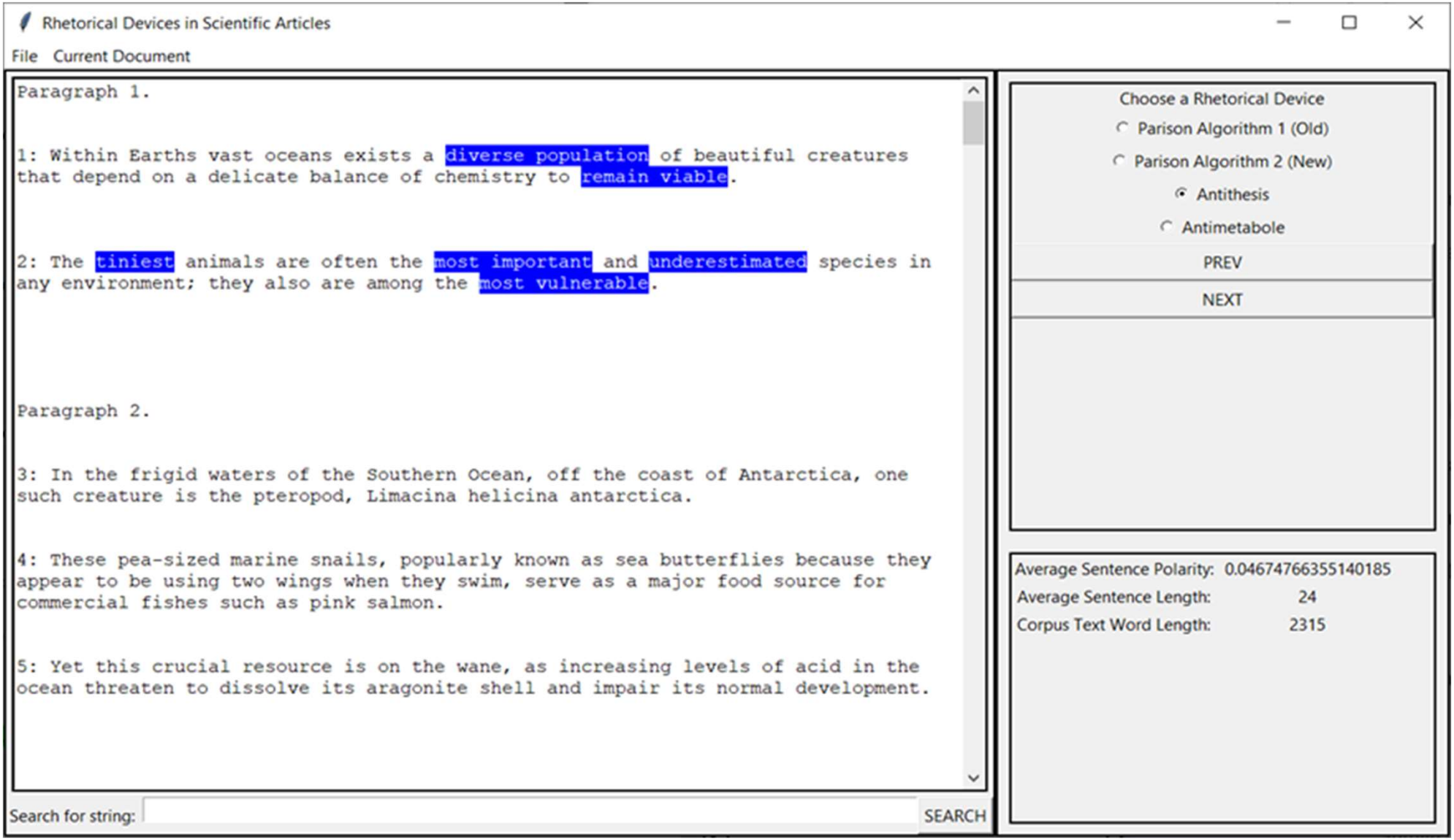

Rhetorical Devices in Scientific Articles
File
Current Document
Paragraph 1.
1: Within Earths vast oceans exists a diverse population of beautiful creatures that depend on a delicate balance of chemistry to remain viable.
2: The tiniest animals are often the most important and underestimated species in any environment; they also are among the most vulnerable.
Paragraph 2.
3: In the frigid waters of the Southern Ocean, off the coast of Antarctica, one such creature is the pteropod, Limacina helicina antarctica.
4: These pea-sized marine snails, popularly known as sea butterflies because they appear to be using two wings when they swim, serve as a major food source for commercial fishes such as pink salmon.
5: Yet this crucial resource is on the wane, as increasing levels of acid in the ocean threaten to dissolve its aragonite shell and impair its normal development.
Search for string:
SEARCH
Choose a Rhetorical Device
Parison Algorithm 1 (Old)
Parison Algorithm 2 (New)
Antithesis
Antimetabole
PREV
NEXT
Average Sentence Polarity: 0.04674766355140185
Average Sentence Length: 24
Corpus Text Word Length: 2315

Fig. 3

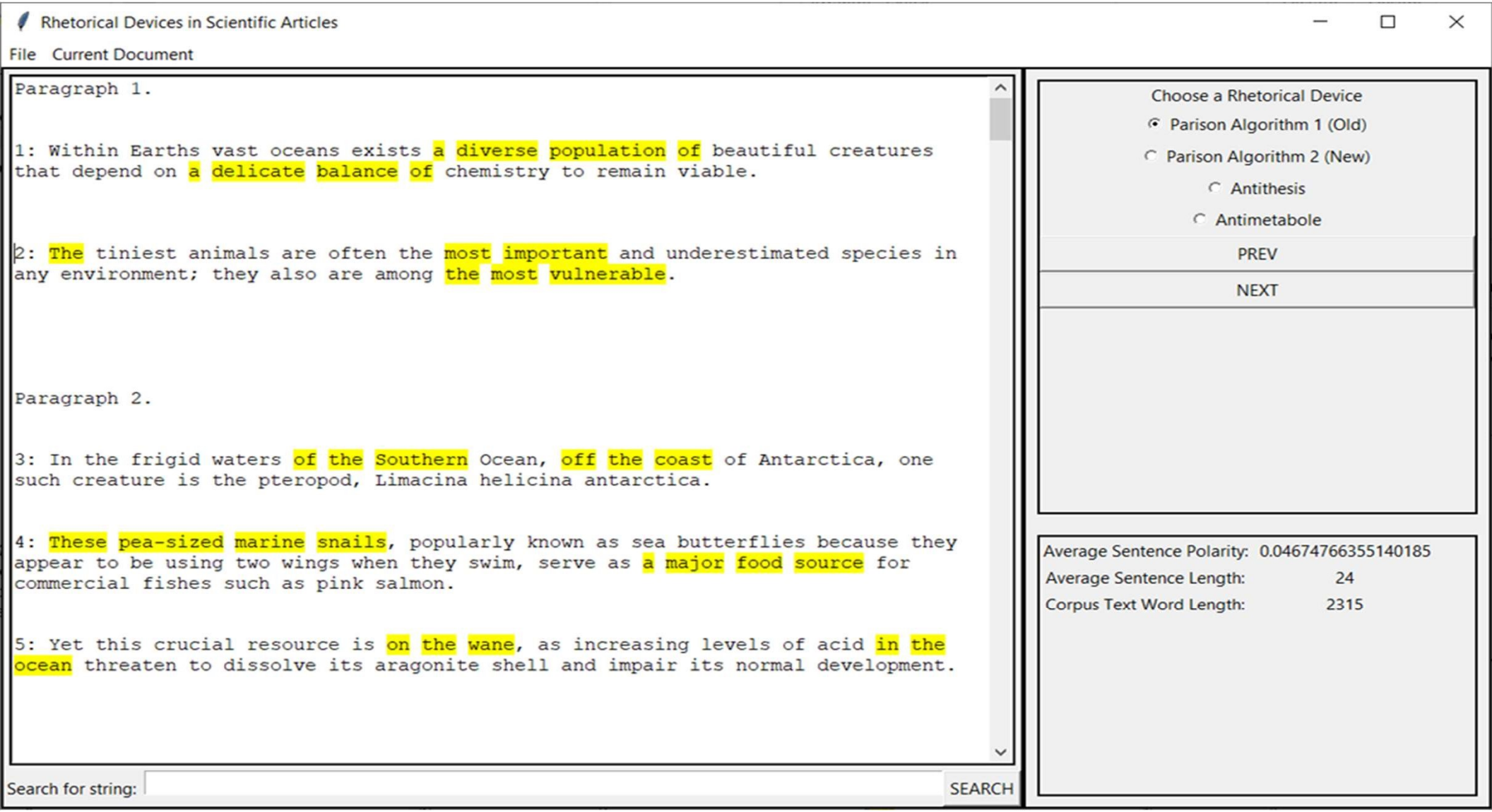

Rhetorical Devices in Scientific Articles
File
Current Document
Paragraph 1.
1: Within Earths vast oceans exists a diverse population of beautiful creatures that depend on a delicate balance of chemistry to remain viable.
2: The tiniest animals are often the most important and underestimated species in any environment; they also are among the most vulnerable.
Paragraph 2.
3: In the frigid waters of the Southern Ocean, off the coast of Antarctica, one such creature is the pteropod, Limacina helicina antarctica.
4: These pea-sized marine snails, popularly known as sea butterflies because they appear to be using two wings when they swim, serve as a major food source for commercial fishes such as pink salmon.
5: Yet this crucial resource is on the wane, as increasing levels of acid in the ocean threaten to dissolve its aragonite shell and impair its normal development.
Search for string:
SEARCH
Choose a Rhetorical Device
Parison Algorithm 1 (Old)
Parison Algorithm 2 (New)
Antithesis
Antimetabole
PREV
NEXT
Average Sentence Polarity: 0.04674766355140185
Average Sentence Length: 24
Corpus Text Word Length: 2315

Fig. 4

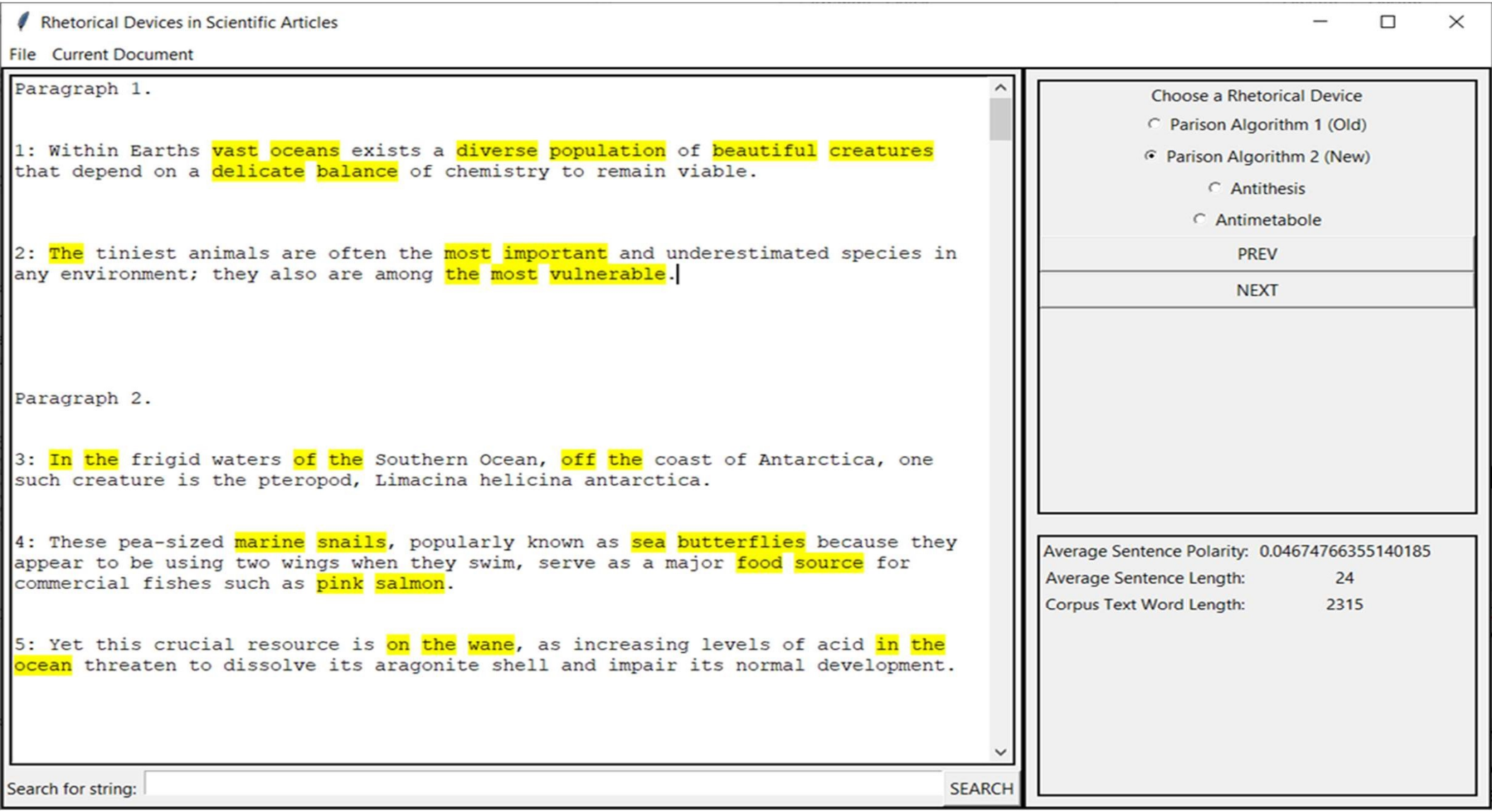

Rhetorical Devices in Scientific Articles
File
Current Document
Paragraph 1.
1: Within Earths vast oceans exists a diverse population of beautiful creatures that depend on a delicate balance of chemistry to remain viable.
2: The tiniest animals are often the most important and underestimated species in any environment; they also are among the most vulnerable.
Paragraph 2.
3: In the frigid waters of the Southern Ocean, off the coast of Antarctica, one such creature is the pteropod, Limacina helicina antarctica.
4: These pea-sized marine snails, popularly known as sea butterflies because they appear to be using two wings when they swim, serve as a major food source for commercial fishes such as pink salmon.
5: Yet this crucial resource is on the wane, as increasing levels of acid in the ocean threaten to dissolve its aragonite shell and impair its normal development.
Search for string:
SEARCH
Choose a Rhetorical Device
Parison Algorithm 1 (Old)
Parison Algorithm 2 (New)
Antithesis
Antimetabole
PREV
NEXT
Average Sentence Polarity: 0.04674766355140185
Average Sentence Length: 24
Corpus Text Word Length: 2315

Fig. 5

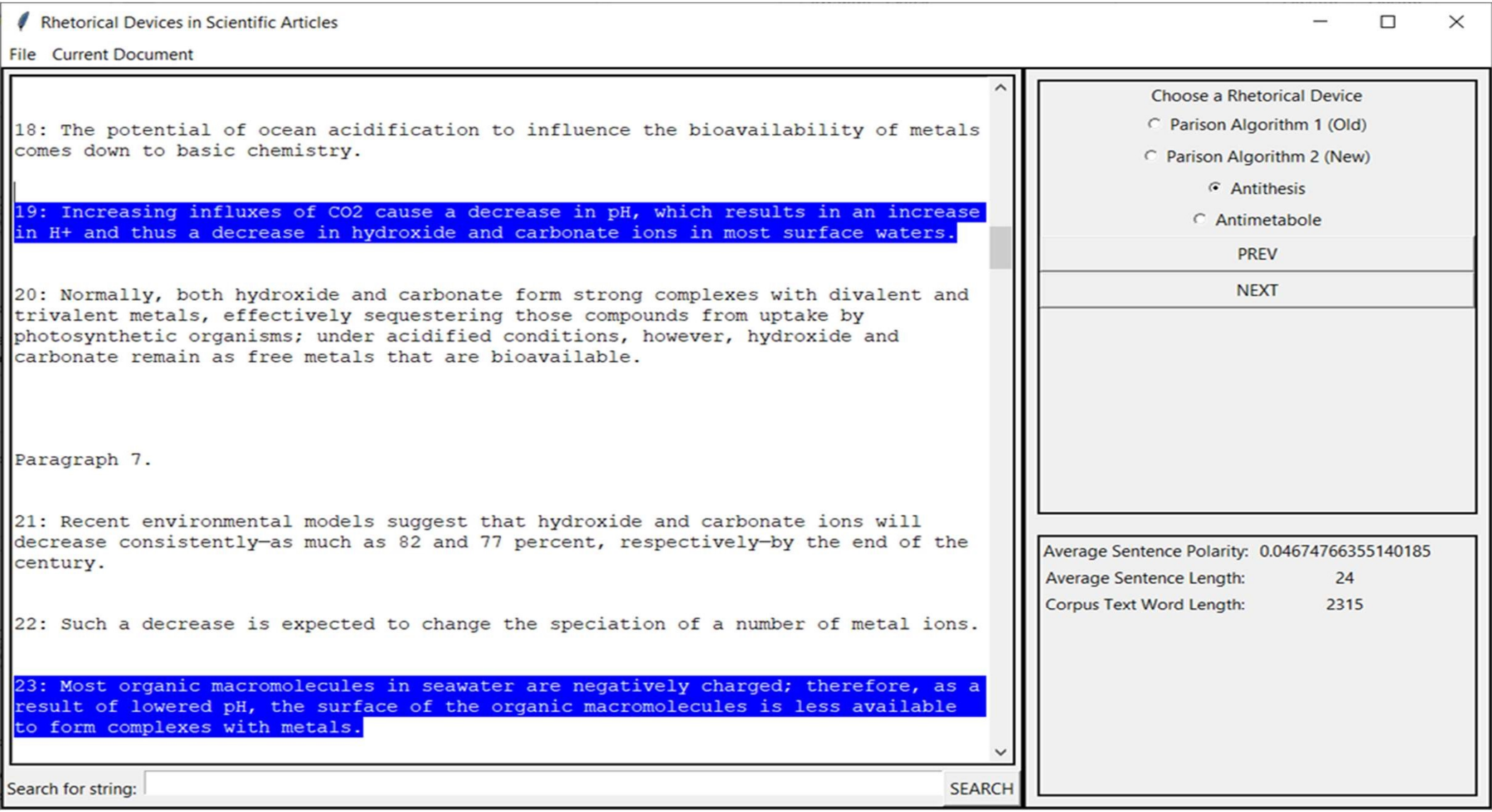
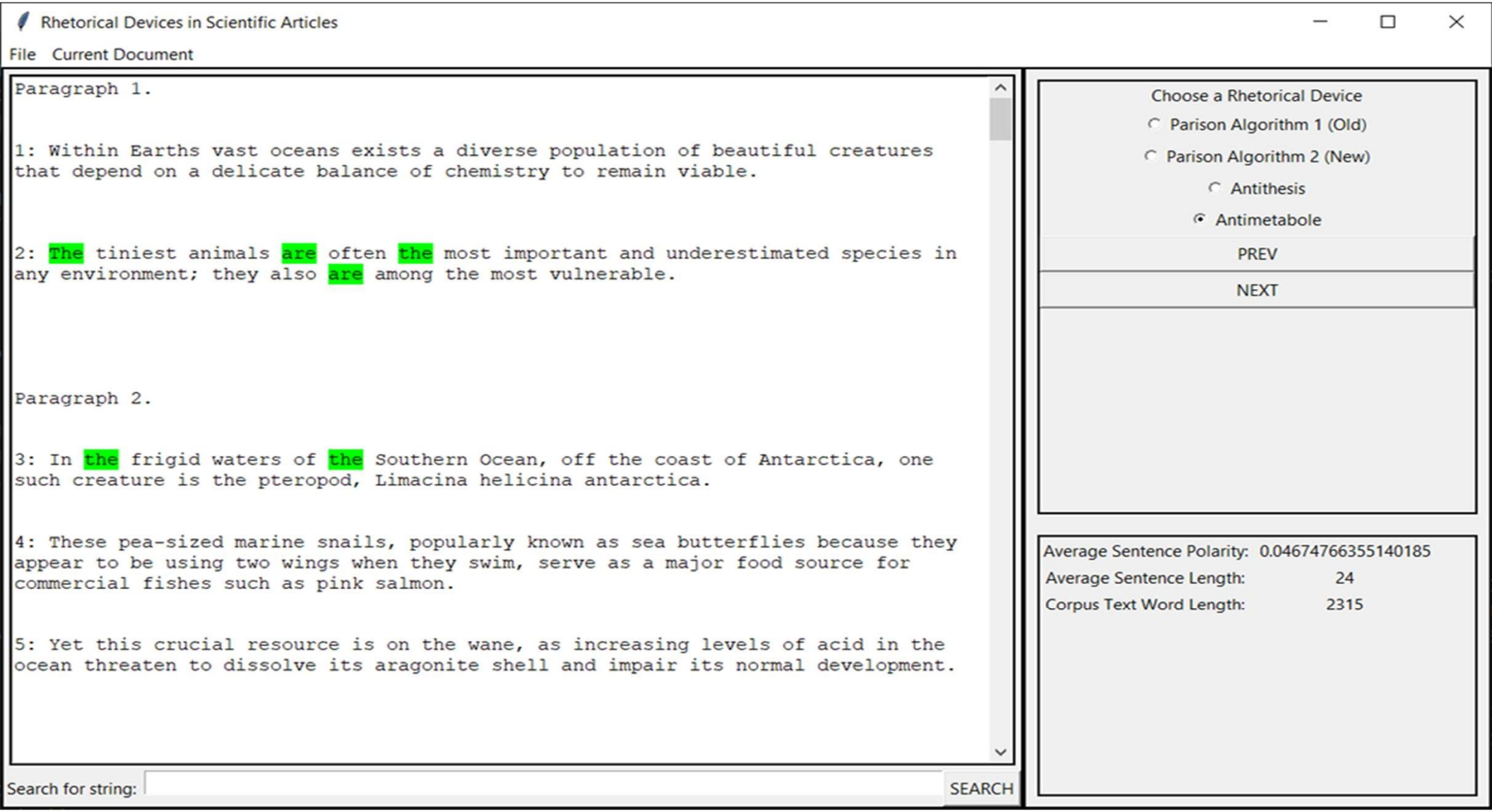

Rhetorical Devices in Scientific Articles
File
Current Document
18: The potential of ocean acidification to influence the bioavailability of metals comes down to basic chemistry.
19: Increasing influxes of CO2 cause a decrease in pH, which results in an increase in H+ and thus a decrease in hydroxide and carbonate ions in most surface waters.
20: Normally, both hydroxide and carbonate form strong complexes with divalent and trivalent metals, effectively sequestering those compounds from uptake by photosynthetic organisms; under acidified conditions, however, hydroxide and carbonate remain as free metals that are bioavailable.
Paragraph 7.
21: Recent environmental models suggest that hydroxide and carbonate ions will decrease consistently—as much as 82 and 77 percent, respectively—by the end of the century.
22: Such a decrease is expected to change the speciation of a number of metal ions.
23: Most organic macromolecules in seawater are negatively charged; therefore, as a result of lowered pH, the surface of the organic macromolecules is less available to form complexes with metals.
Search for string:
SEARCH
Choose a Rhetorical Device
Parison Algorithm 1 (Old)
Parison Algorithm 2 (New)
Antithesis
Antimetabole
PREV
NEXT
Average Sentence Polarity: 0.04674766355140185
Average Sentence Length: 24
Corpus Text Word Length: 2315

Fig. 6


Rhetorical Devices in Scientific Articles
File
Current Document
Paragraph 1.
1: Within Earths vast oceans exists a diverse population of beautiful creatures that depend on a delicate balance of chemistry to remain viable.
2: The tiniest animals are often the most important and underestimated species in any environment; they also are among the most vulnerable.
Paragraph 2.
3: In the frigid waters of the Southern Ocean, off the coast of Antarctica, one such creature is the pteropod, Limacina helicina antarctica.
4: These pea-sized marine snails, popularly known as sea butterflies because they appear to be using two wings when they swim, serve as a major food source for commercial fishes such as pink salmon.
5: Yet this crucial resource is on the wane, as increasing levels of acid in the ocean threaten to dissolve its aragonite shell and impair its normal development.
Search for string:
SEARCH
Choose a Rhetorical Device
Parison Algorithm 1 (Old)
Parison Algorithm 2 (New)
Antithesis
Antimetabole
PREV
NEXT
Average Sentence Polarity: 0.04674766355140185
Average Sentence Length: 24
Corpus Text Word Length: 2315