

AI Unveils the Books with the Richest Vocabulary

25 classics compared for World Book Day and English Language Day

NEW YORK, UNITED STATES, April 18, 2024 /EINPresswire.com/ -- Key findings:

- The top three scorers in lexical diversity are all works by female authors: Beryl Bainbridge, Edith Wharton, and Emily Brontë. This might suggest a particularly rich use of language among female authors in different periods, highlighting their contribution to the literary landscape with a dense and varied vocabulary.

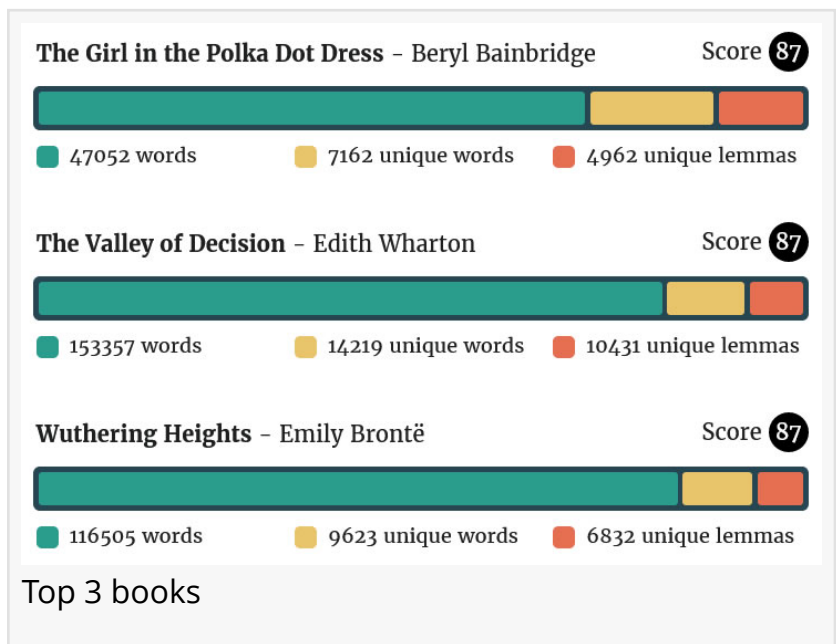
- Despite the varying lengths of their narratives, from Bainbridge's relatively short "The Girl in the Polka Dot Dress" to Wharton's lengthier "The Valley of Decision," these authors manage to maintain a high level of lexical diversity.

- The diversity in publication dates among the top-scoring novels suggests that lexical richness is not confined to a specific era. For instance, "Wuthering Heights" was published in the 19th century, while "The Girl in the Polka Dot Dress" is a more contemporary work. This cross-era representation underscores the timeless value of a rich vocabulary in storytelling.

- For a beginner reader, it might be a good option to start with the books at the bottom of the ranking table and gradually move up, slowly making the reading more complex.

On April 23, commemorating the death of William Shakespeare, both World Book Day and English Language Day are celebrated.

My Poetic Side (<https://mypoeticside.com>) combined these two celebrations in this analysis, which reviews the vocabulary of 25 books to identify the most complex and accessible texts for



readers.

With that goal in mind, they used artificial intelligence and natural language processing software to dissect each work and obtain answers.

My Poetic Side stored the 25 digitally, word by word, in a database. That database contained 4,493,150 words. Then natural language processing (NLP) software and artificial intelligence helped identify and unify similar words through lemmatization. Lemmatization involves reducing all the variants that a word might have to obtain its base lemma. For instance, buy, buys and bought are grouped under the lemma buy. This technique is a tool of natural language processing that ensures the final result is more accurate by only counting lemmas and not all the variations of a word.

Finally, the researchers created their own scoring system to standardize criteria and evaluate lexical diversity.

You can read the [full article here](#).

Julian Yanover

MyPoeticSide

[email us here](#)

Visit us on social media:

[Twitter](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/702368587>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2024 Newsmatics Inc. All Right Reserved.