Vocabulary gradient

Tom Rochette <tom.rochette@coreteks.org>

July 24, 2025 — daae079c

1   Problem

I write a lot of articles and I want them to be understood by most people. How do I use the most common language possible?

2   Solution

My approach is to write whatever I want to write about using whatever language I come up with first. Then I use a tool I've developed which I've called the vocabulary gradient. It is a very simple tool where you will generally copy and paste the article you've written and look at the result of the analysis. The tool uses a word frequency list as specified in the README.md. This list was built using the Project Gutenberg library, which makes the word frequency list a bit outdated.

The report generated by the tool presents the minimum, average, maximum and standard deviation of the index of the words used in the text you provided. Those numbers give you a rough overview of the difficulty of your text based on word frequency alone. The lower your average and maximum is, the simpler the article should be to understand An histogram is also generated, where the bins are based again on the index of the word in the frequency list. Finally, the provided text is rendered with each word index as a subscript. Words that are unknown are highlighted in yellow, while words for which the index is high are shaded with a darker shade of gray as their index increases.

With this information in hand, you can spot the words that have high word frequency indexes and try to replace them with lower index words.

3   Reference

- https://github.com/tomzx/vocabulary-gradient