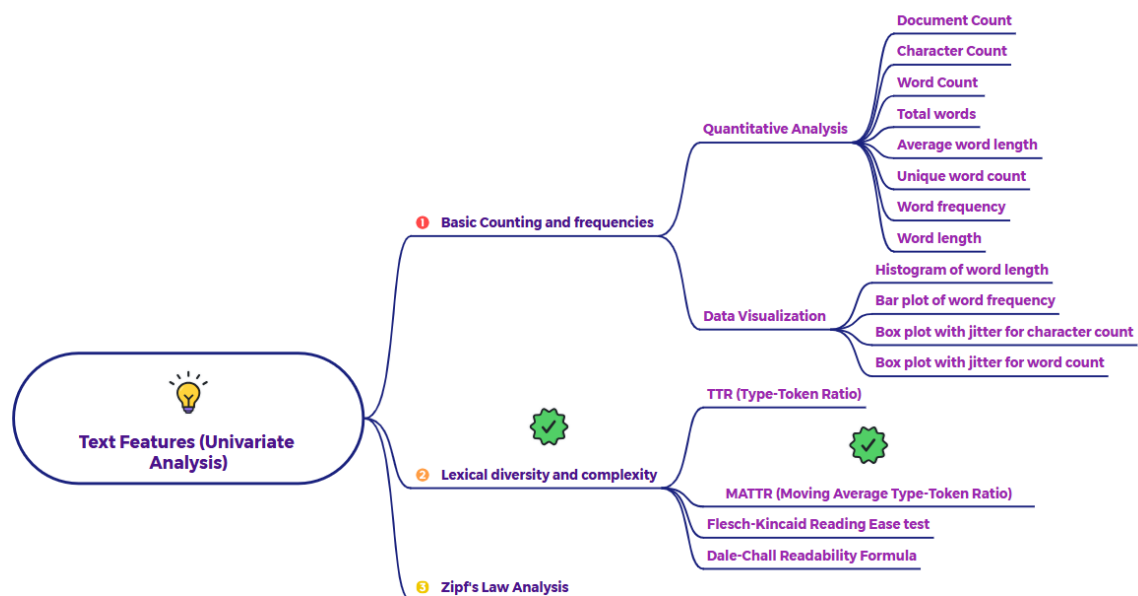# What is MATTR (Moving Average Type-Token Ratio)?



**Moving Average Type-Token Ratio (MATTR)** is a more sophisticated and robust measure of **lexical diversity** that attempts to address the issue of text length dependency inherent in the basic Type-Token Ratio (TTR).

## Calculation:

MATTR is calculated by:

- **Dividing the text into a series of contiguous, non-overlapping segments (windows) of a fixed size (e.g., 50 words).** The choice of window size can influence the resulting MATTR value, but a common size is often used for consistency.
- **Calculating the TTR for each of these segments.** For each window, the number of unique words (types) is divided by the total number of words (tokens) within that window.
- **Averaging the TTR values obtained from all the segments.** The final MATTR score is the mean of these individual TTR values.

MATTR = (Sum of TTR for all windows) / (Number of windows)

### How to Interpret Moving Average Type-Token Ratio (MATTR):

The resulting MATTR value typically ranges between 0 and 1 (or can be expressed as a percentage).

- **Higher MATTR:** A higher MATTR value indicates **higher lexical diversity**. This suggests that across different segments of the text, a greater proportion of words are unique within those local contexts. It implies a more consistently varied vocabulary throughout the text, less repetition within short spans, and a richer overall use of language.
- **Lower MATTR:** A lower MATTR value indicates **lower lexical diversity**. This suggests that across different segments of the text, a smaller proportion of words are unique within those local contexts. It implies more repetition of words within short spans and a less varied vocabulary throughout the text.

### Advantages of MATTR over TTR:

- **Reduced Length Dependency:** The primary advantage of MATTR is that by calculating TTR over fixed-size windows, it largely mitigates the issue of text length affecting the score. Whether the overall text is short or long, the MATTR focuses on the local lexical diversity within those consistent segments. This allows for more meaningful comparisons of lexical diversity between texts of different lengths.
- **More Stable Measure:** MATTR tends to be a more stable and reliable measure of lexical diversity, especially for longer texts, as it averages out local variations in vocabulary use.

### Interpretation Considerations:

- **Window Size:** The choice of window size can impact the MATTR value. Smaller window sizes might be more sensitive to local variations, while larger window sizes might smooth out these variations more. It's important to be aware of the window size used when interpreting and comparing MATTR scores.
- **Focus on Local Diversity:** MATTR essentially measures the "local" lexical diversity within the chosen window size. While it provides a good overall indication of vocabulary variation, it doesn't directly capture the total number of unique words in the entire text as TTR

does (though indirectly, a higher MATTR often correlates with a larger overall vocabulary).

- **Comparison Across Texts:** MATTR allows for more direct and meaningful comparisons of lexical diversity between texts of different lengths, which is a significant improvement over TTR.

In summary, MATTR provides a more robust and length-independent measure of lexical diversity by calculating the average TTR across fixed-size moving windows of text. A higher MATTR indicates a more consistently varied and richer vocabulary throughout the text, making it a valuable tool for comparing lexical diversity across different texts, regardless of their length.