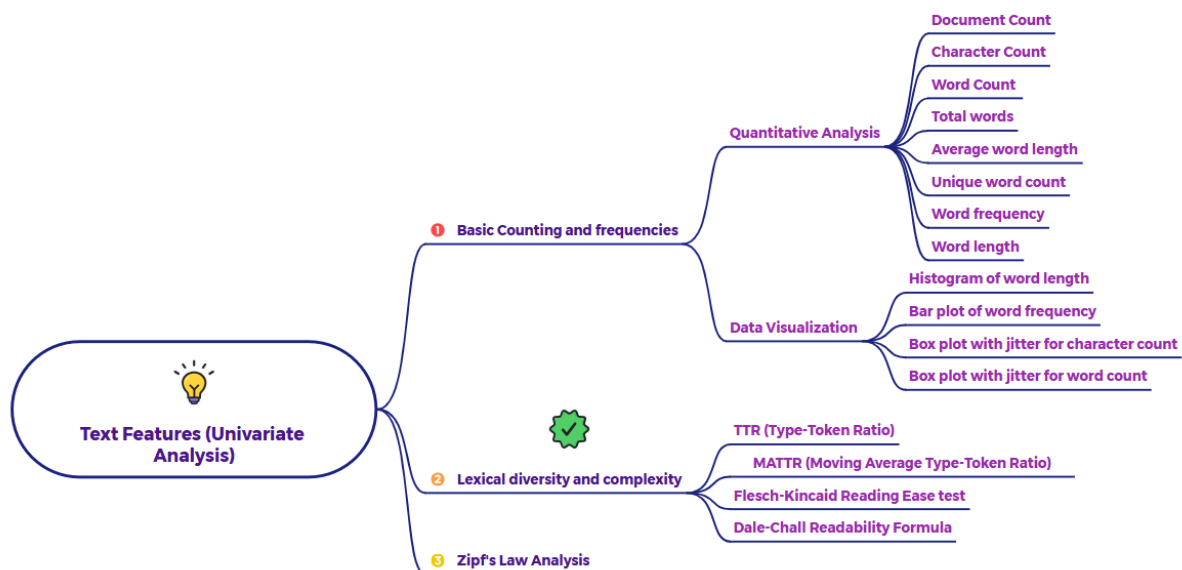


What is lexical diversity and complexity?



Lexical diversity refers to the **range of unique words** used in a text or across a set of texts. A text with high lexical diversity employs a wider variety of vocabulary, indicating a richer and potentially more sophisticated use of language. It's often measured by metrics that compare the number of unique words (types) to the total number of words (tokens).

Lexical complexity is a broader concept that encompasses not just the variety of words but also the **sophistication and density of information** conveyed through the vocabulary. It can involve factors like the use of less common words, words with multiple syllables, and the density of content words versus function words. While lexical diversity focuses on the sheer number of different words, lexical complexity delves into the nature and richness of that vocabulary.

Definitions of the key measures:

- **TTR (Type-Token Ratio):** A simple measure of lexical diversity calculated as the number of unique words (types) divided by the total number of words (tokens).

- **MATTR (Moving Average Type-Token Ratio):** A more robust measure of lexical diversity calculated by averaging the TTR over a fixed-size moving window of text.
- **Flesch-Kincaid Reading Ease test:** A readability formula that estimates how easy a piece of writing is to understand, with higher scores indicating easier readability.
- **Dale-Chall Readability Formula:** A readability formula that estimates the grade level required to understand a piece of writing, based on a list of familiar words.