

Text Analysis Report: theory_of_relativity_wikipedia

1. Basic Statistics

Metric	Value
Word Count	515
Sentence Count	26
Character Count	3633
Unique Word Count	270
Avg Word Length (chars)	5.85
Avg Sentence Length (words)	19.81
Avg Sentence Length (chars)	139.73
Sentence Length Std Dev (words)	12.18
Lexical Diversity (TTR)	0.5243
MATTR (Window=50)	0.7914

2. Entropy & Complexity

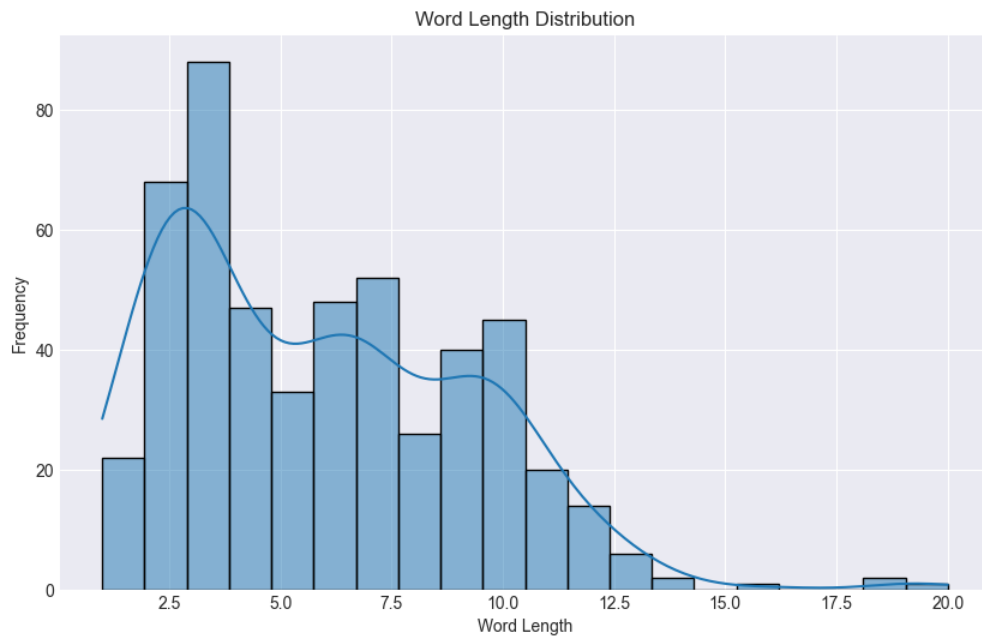
Metric	Value
Shannon Entropy (Words)	7.2313 bits
Shannon Entropy (Chars)	4.6032 bits
Function Word Ratio	0.3340

3. Readability Scores

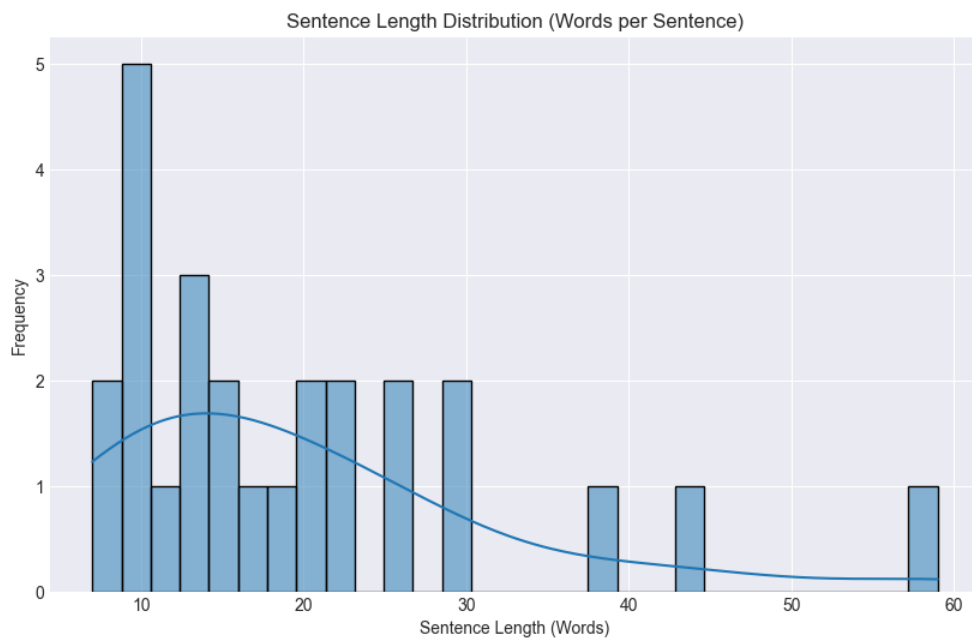
Index	Score
Flesch Reading Ease	9.79
Flesch-Kincaid Grade Level	16.60
Gunning Fog Index	15.70
SMOG Index	16.90
Coleman-Liau Index	17.93
Automated Readability Index	17.80

4. Visualizations

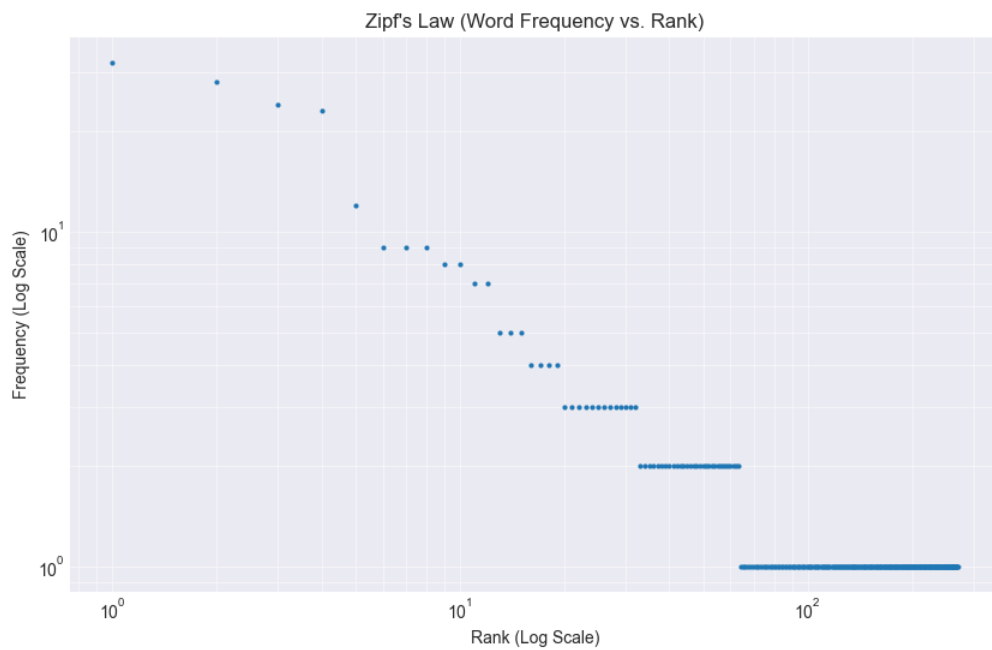
Word Length Distribution

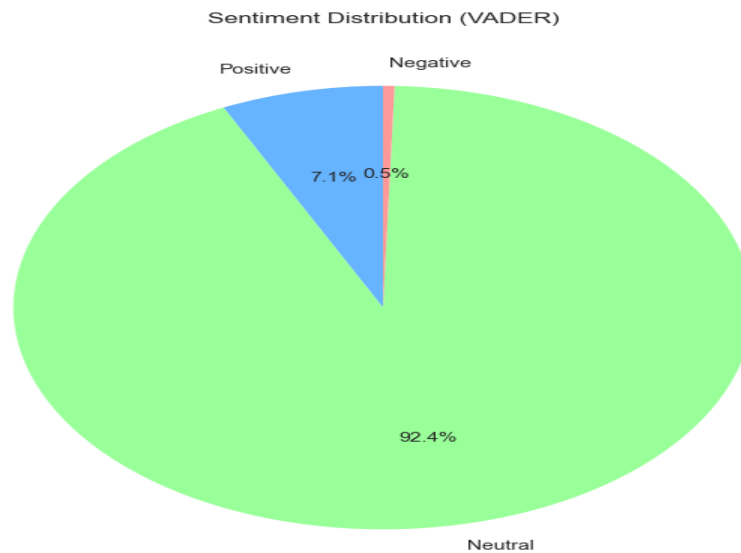


Sentence Length Distribution

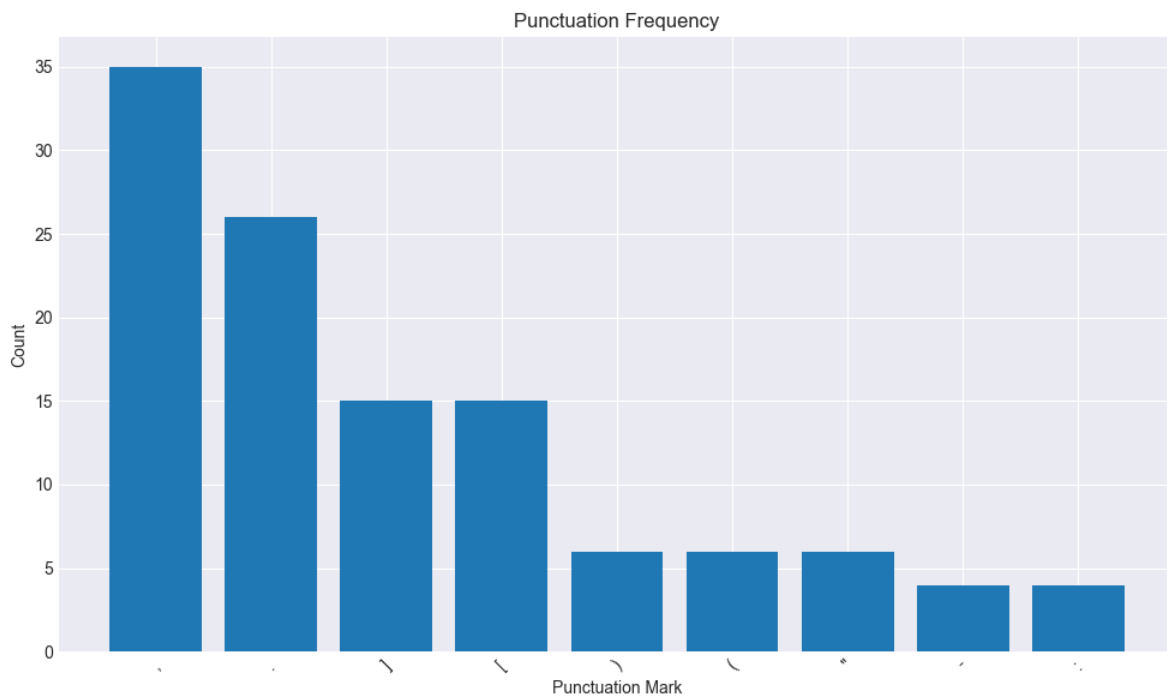


Zipf's Law Analysis (Word Frequency vs Rank)





Punctuation Frequency



5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

the (32), and (28), of (24), relativity (23), theory (12), general (9), in (9), to (9), physics (8), by (8), as (7), 3 (7), for (5), special (5), a (5)

Most Common Punctuation:

',' (35), '.' (26), '[' (15), ']' (15), '"' (6), '(' (6), ')' (6), ':' (4), '-' (4)

6. N-gram Analysis

Common 2-grams (Top 10):

'general relativity' (9), 'theory of' (6), 'the theory' (6), 'of relativity' (5), 'special relativity' (5), 'in the' (5), '3 the' (3), 'physics and' (3), 'for the' (2), 'of the' (2)
2-gram Repetition Rate: 0.1148

Common 3-grams (Top 10):

'theory of relativity' (4), 'the theory of' (2), 'special relativity and' (2), '3 the theory' (2), 'physics and astronomy' (2), '3 4 5' (2), 'of special relativity' (2), 'of general relativity' (2), 'this article is' (1), 'article is about' (1)
3-gram Repetition Rate: 0.0195

Common 4-grams (Top 10):

'this article is about' (1), 'article is about the' (1), 'is about the scientific' (1), 'about the scientific concept' (1), 'the scientific concept for' (1), 'scientific concept for philosophical' (1), 'concept for philosophical or' (1), 'for philosophical or ontological' (1), 'philosophical or ontological theories' (1), 'or ontological theories about' (1)
4-gram Repetition Rate: 0.0000

Character N-gram Analysis:

Common Character 3-grams (Top 10):

'the' (61), '_th' (49), 'ati' (43), 'al_' (37), 'lat' (33), 'he_' (32), '_re' (31), 'nd_' (31), 'vit' (30), 'and' (30)

Common Character 4-grams (Top 10):

'_the' (49), 'the_' (32), 'lati' (32), 'rela' (29), 'elat' (29), '_rel' (28), '_and' (28), 'and_' (28), 'ativ' (27), 'tivi' (25)

Common Character 5-grams (Top 10):

'_the_' (29), 'relat' (29), '_rela' (28), 'elati' (28), '_and_' (28), 'lativ' (27), 'ativi' (25), 'tivit' (24), 'ivity' (23), 'vity_' (21)