

Text Analysis Report: modernist_literature_wikipedia

1. Basic Statistics

Metric	Value
Word Count	588
Sentence Count	21
Character Count	3773
Unique Word Count	99
Avg Word Length (chars)	5.28
Avg Sentence Length (words)	28.00
Avg Sentence Length (chars)	179.67
Sentence Length Std Dev (words)	9.34
Lexical Diversity (TTR)	0.1684
MATTR (Window=50)	0.8252

2. Entropy & Complexity

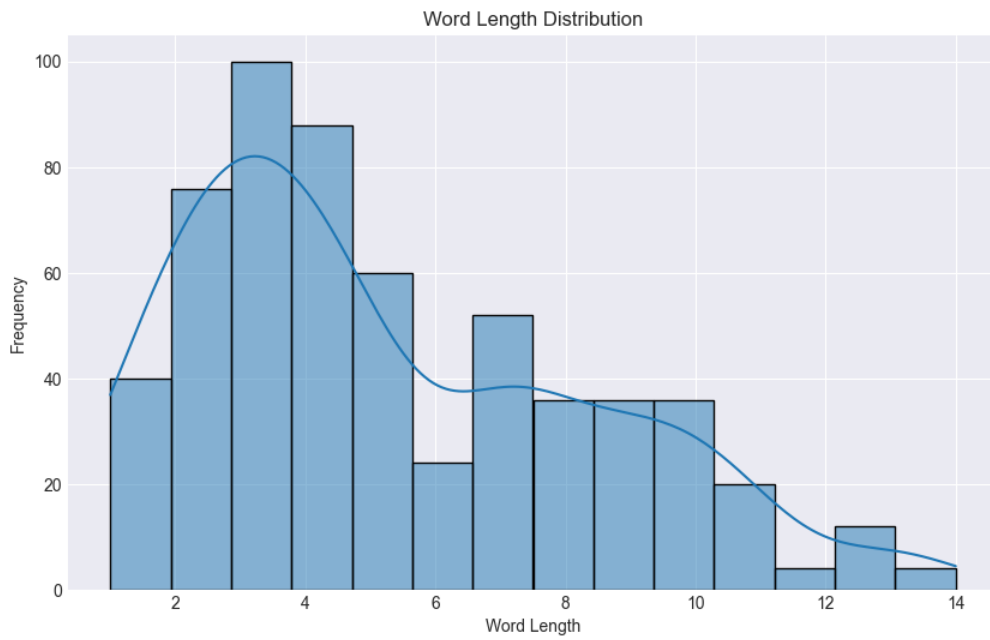
Metric	Value
Shannon Entropy (Words)	6.2719 bits
Shannon Entropy (Chars)	4.3871 bits
Function Word Ratio	0.3673

3. Readability Scores

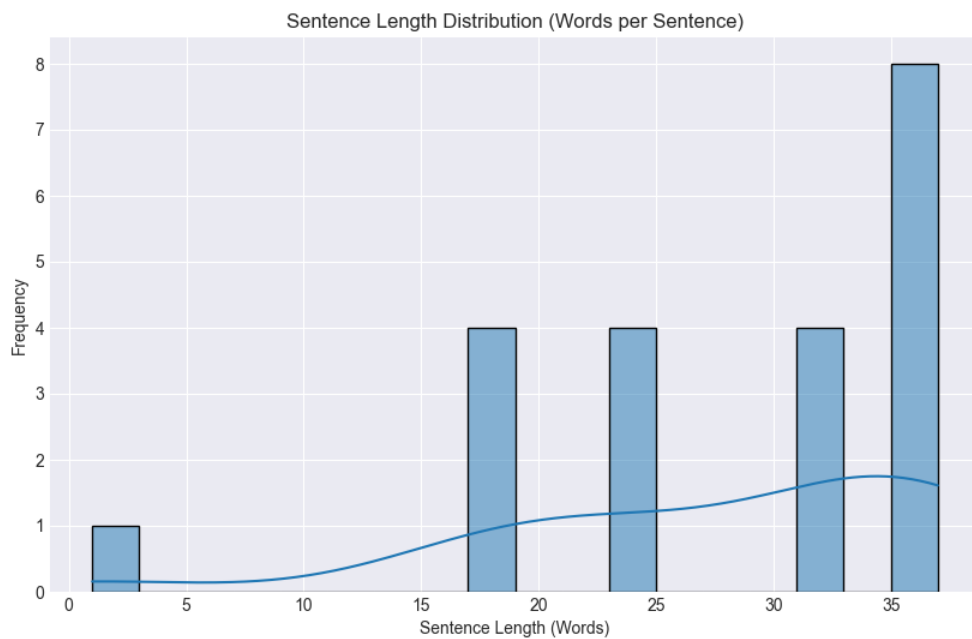
Index	Score
Flesch Reading Ease	17.47
Flesch-Kincaid Grade Level	17.80
Gunning Fog Index	13.34
SMOG Index	18.70
Coleman-Liau Index	14.92
Automated Readability Index	19.50

4. Visualizations

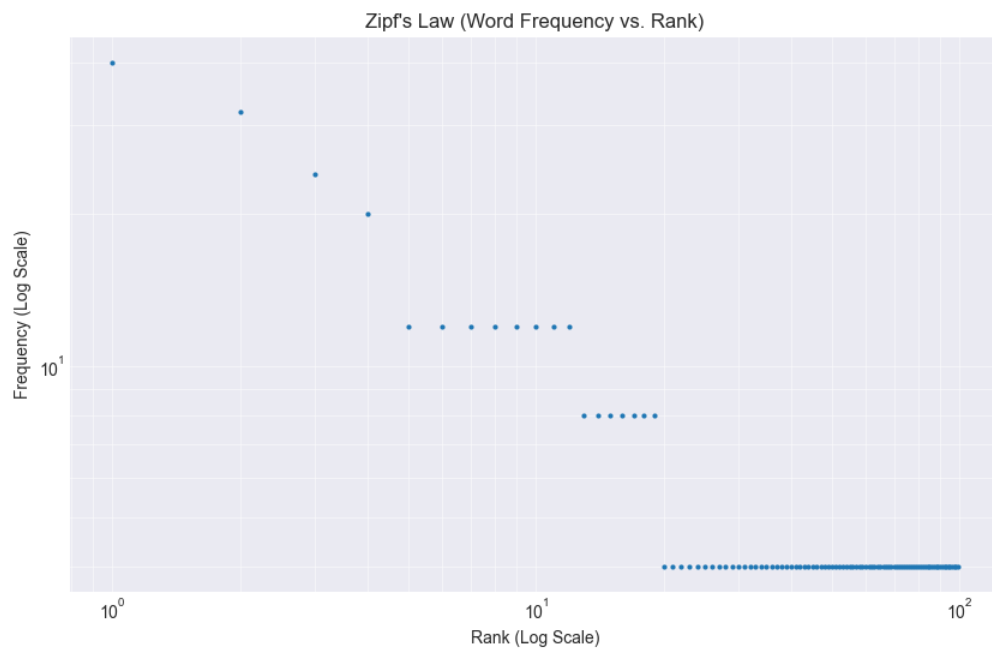
Word Length Distribution



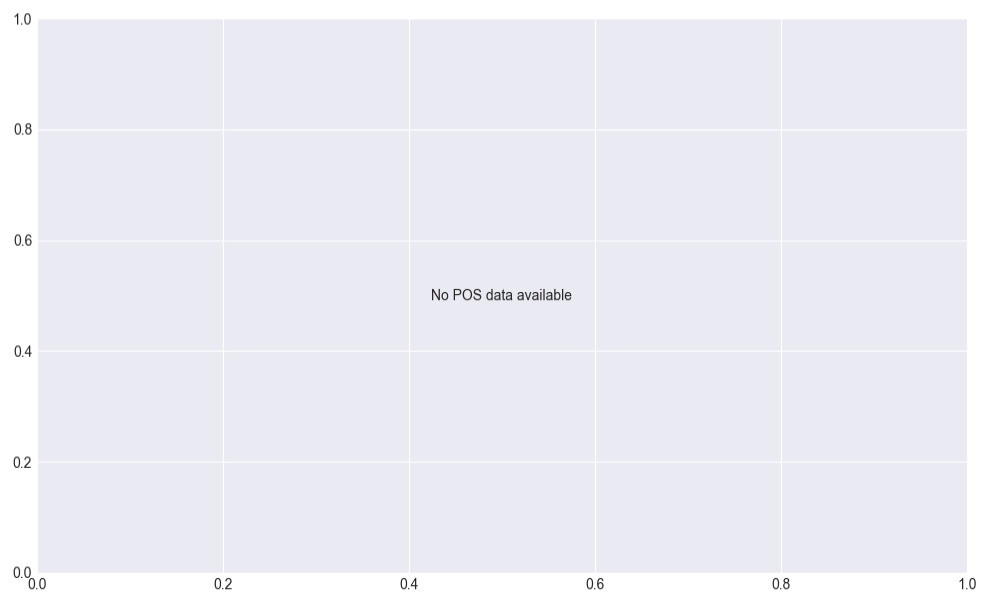
Sentence Length Distribution



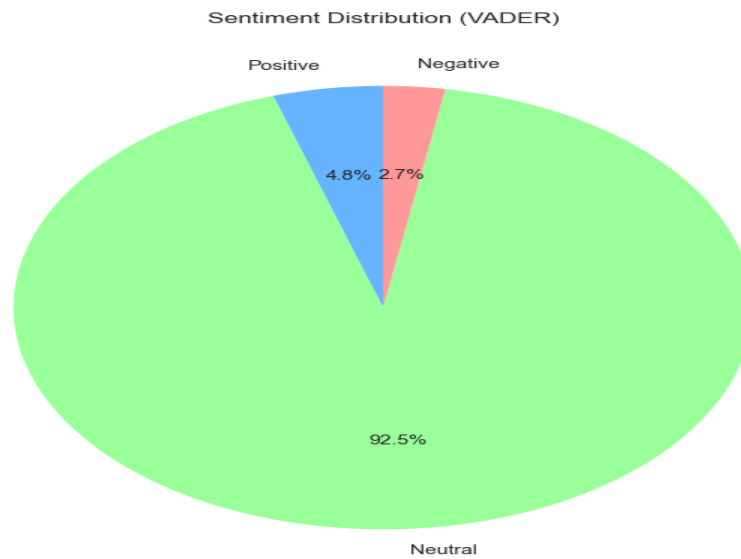
Zipf's Law Analysis (Word Frequency vs Rank)



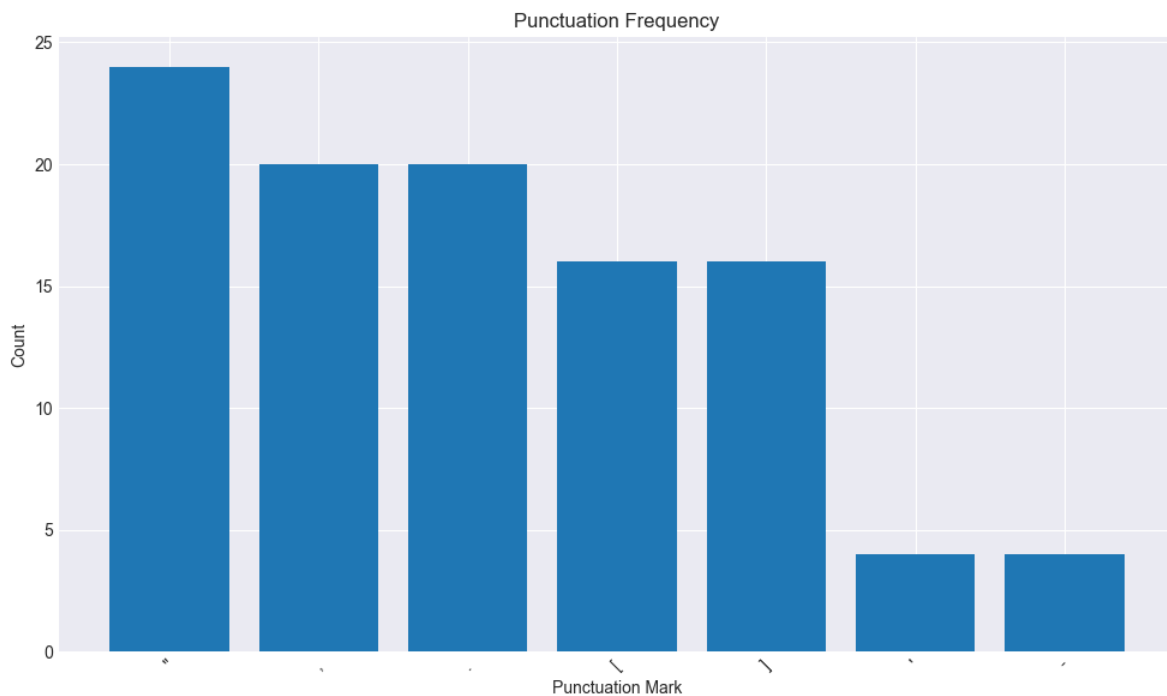
Part-of-Speech Distribution (Top 25)



Sentiment Distribution (VADER)



Punctuation Frequency



5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

the (40), and (32), of (24), a (20), modernist (12), in (12), by (12), conscious (12), writing (12), with (12), literary (12), to (12), literature (8), 20th (8), traditional (8)

Most Common Punctuation:

' (24), , (20), . (20), [(16),] (16), - (4), " (4)

6. N-gram Analysis

Common 2-grams (Top 10):

'modernist literature' (8), 'by a' (8), 'a conscious' (8), 'of the' (8), 'with the' (8), 'literature originated' (4), 'originated in' (4), 'in the' (4), 'the late' (4), 'late 19th' (4)
2-gram Repetition Rate: 0.7581

Common 3-grams (Top 10):

'modernist literature originated' (4), 'literature originated in' (4), 'originated in the' (4), 'in the late' (4), 'the late 19th' (4), 'late 19th and' (4), '19th and early' (4), 'and early 20th' (4), 'early 20th centuries' (4), '20th centuries and' (4)
3-gram Repetition Rate: 0.7491

Common 4-grams (Top 10):

'modernist literature originated in' (4), 'literature originated in the' (4), 'originated in the late' (4), 'in the late 19th' (4), 'the late 19th and' (4), 'late 19th and early' (4), '19th and early 20th' (4), 'and early 20th centuries' (4), 'early 20th centuries and' (4), '20th centuries and is' (4)
4-gram Repetition Rate: 0.7487

Character N-gram Analysis:

Common Character 3-grams (Top 10):

'_th' (60), 'the' (48), 'es_' (44), 'he_' (40), '_an' (36), 'and' (32), 'nd_' (32), 'lit' (28), 'th_' (28), 'ent' (28)

Common Character 4-grams (Top 10):

'_the' (48), 'the_' (40), '_and' (32), 'and_' (32), 'mode' (24), 'tion' (24), '_of_' (24), 'ing_' (24), 'oder' (20), 'dern' (20)

Common Character 5-grams (Top 10):

'_the_' (40), '_and_' (32), 'moder' (20), 'odern' (20), 'derni' (20), '_lite' (20), 'liter' (20), 'itera' (20), 's_of_' (20), '_mode' (20)