Text Analysis Report: crispr_gene_editing_wikipedia

1. Basic Statistics

Metric	Value
Word Count	217
Sentence Count	10
Character Count	1367
Unique Word Count	143
Avg Word Length (chars)	5.13
Avg Sentence Length (words)	21.70
Avg Sentence Length (chars)	136.70
Sentence Length Std Dev (words)	9.90
Lexical Diversity (TTR)	0.6590
MATTR (Window=50)	0.8124

2. Entropy & Complexity

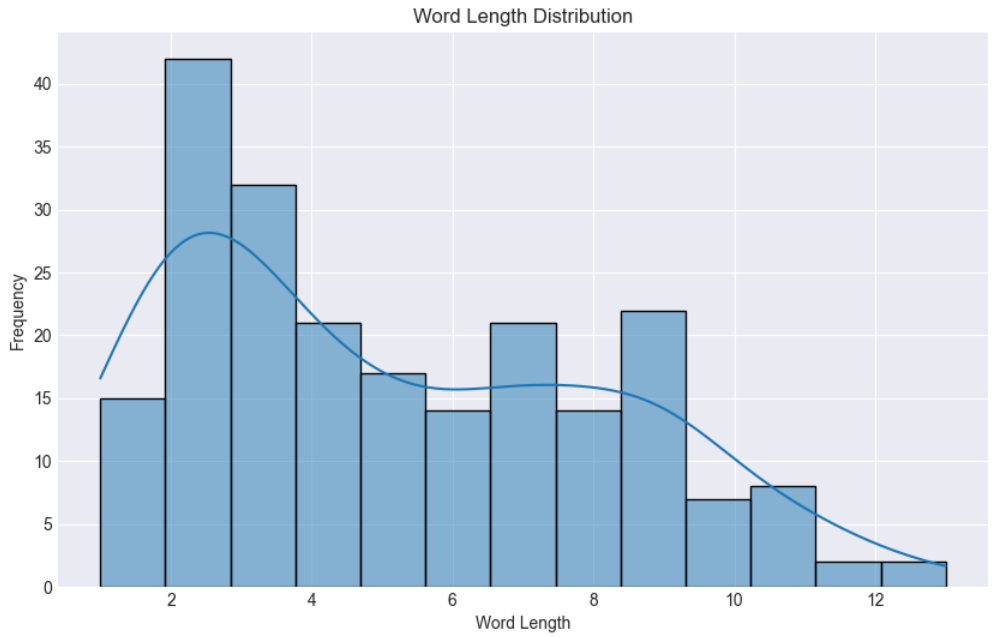
Metric	Value
Shannon Entropy (Words)	6.7556 bits
Shannon Entropy (Chars)	4.5550 bits
Function Word Ratio	0.3779

3. Readability Scores

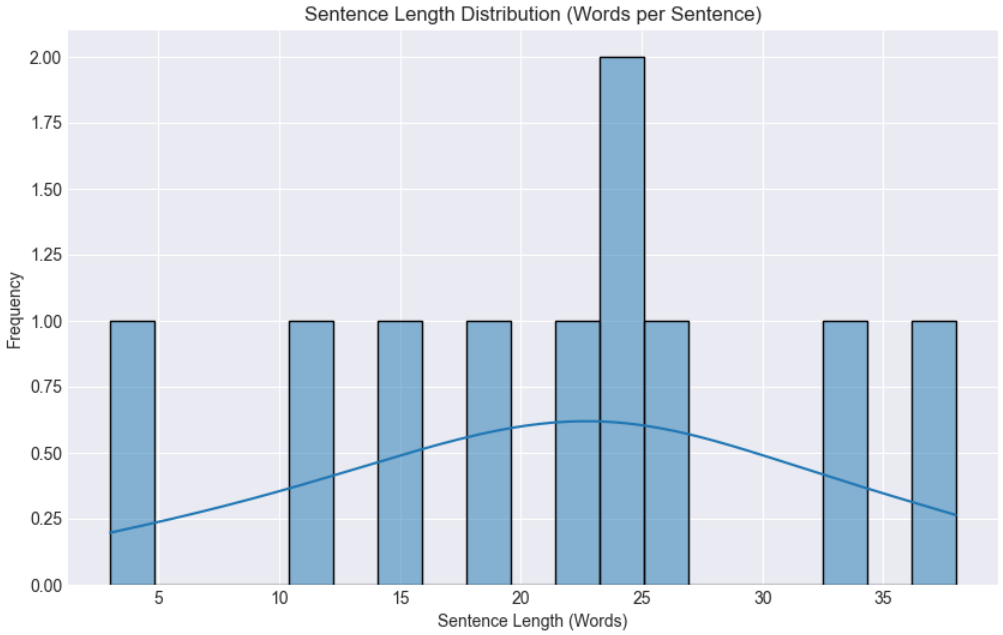
Index	Score
Flesch Reading Ease	31.21
Flesch-Kincaid Grade Level	14.60
Gunning Fog Index	18.67
SMOG Index	17.50
Coleman-Liau Index	14.22
Automated Readability Index	16.40

4. Visualizations

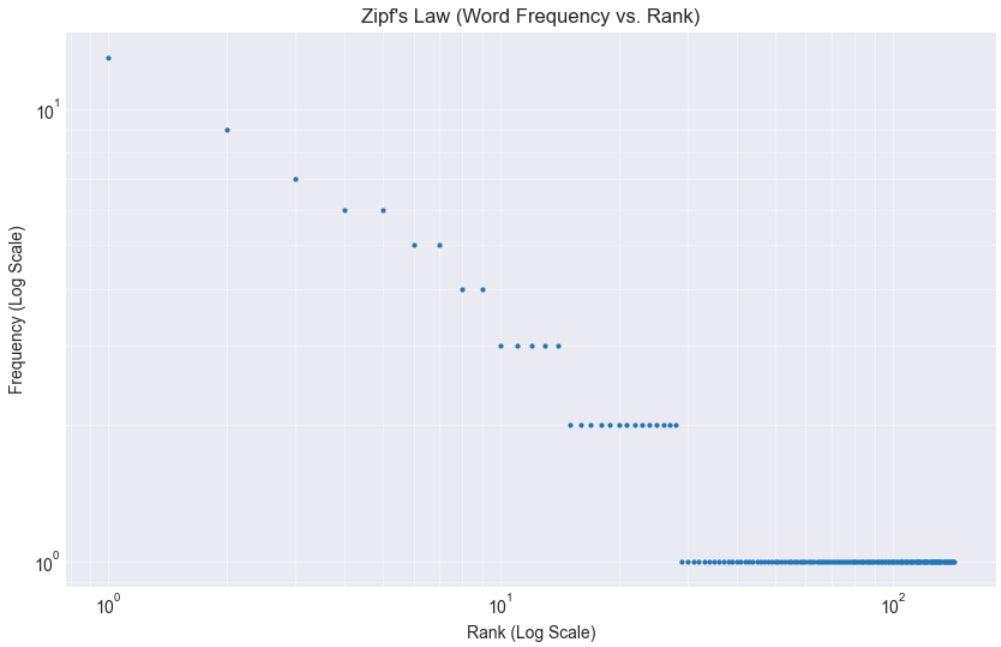
Word Length Distribution



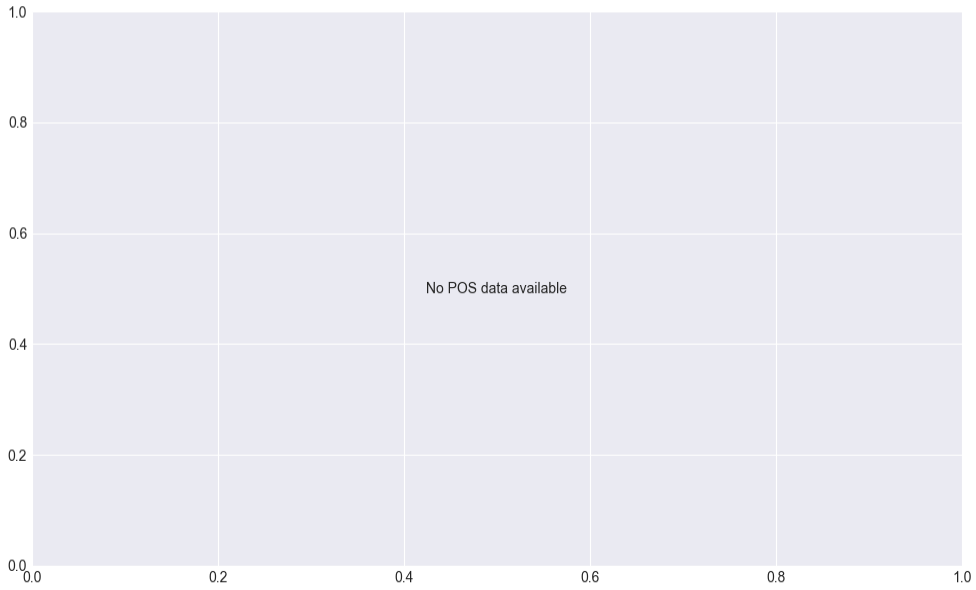
Sentence Length Distribution



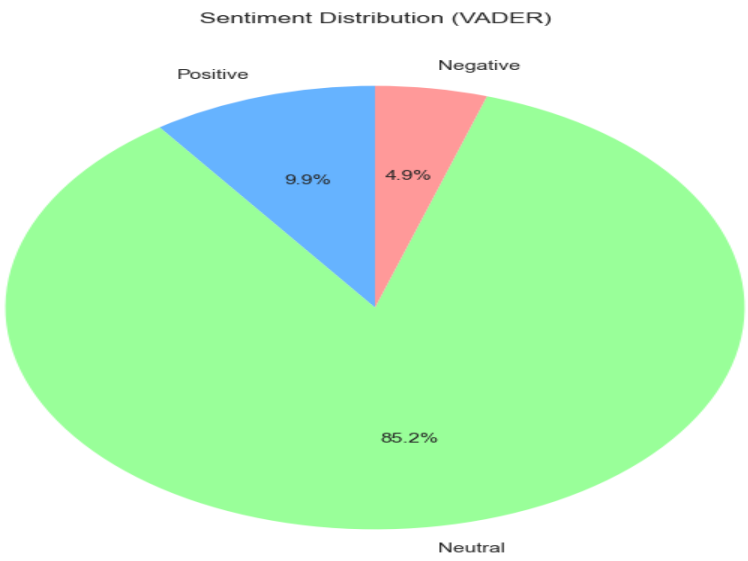
Zipf's Law Analysis (Word Frequency vs Rank)



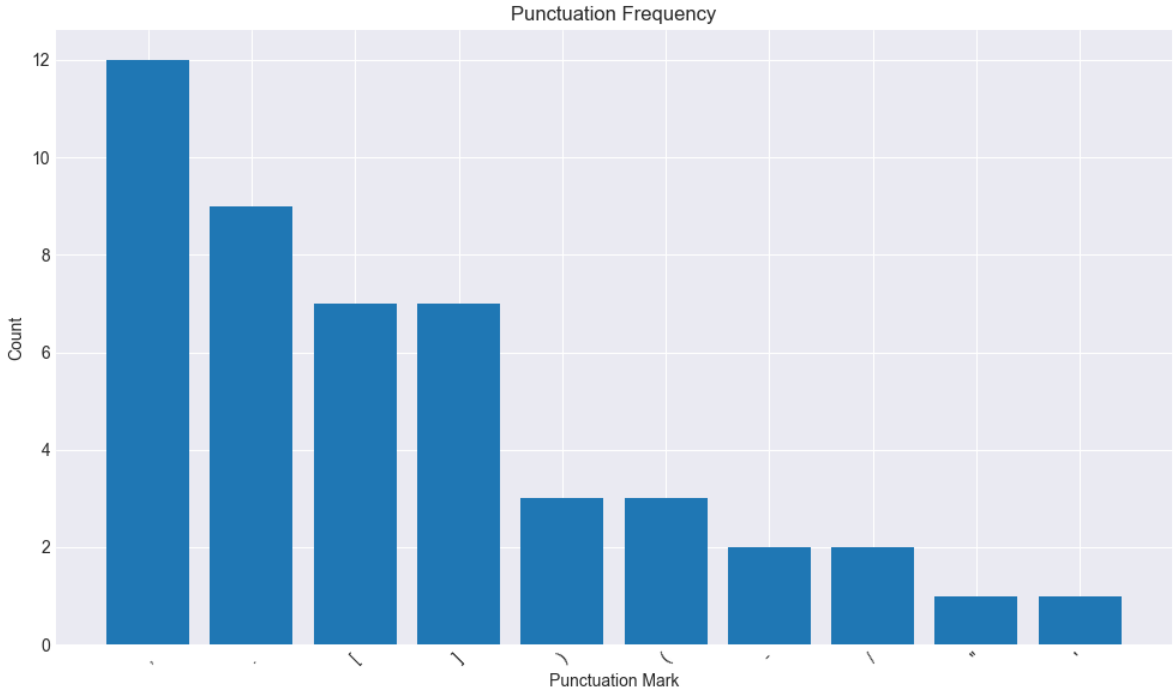
Part-of-Speech Distribution (Top 25)



Sentiment Distribution (VADER)



Punctuation Frequency



5. Frequency Analysis

Most Common Words (Top 15, with stopwords):

the (13), in (9), a (7), of (6), and (6), is (5), as (5), be (4), it (4), crispr (3), genetic (3), technique (3), by (3), prize (3), editing (2)

Most Common Punctuation:

',' (12), '.' (9), '[' (7), ']' (7), '(' (3), ')' (3), '/' (2), '-' (2), '"' (1), "'" (1)

6. N-gram Analysis

Common 2-grams (Top 10):

'can be' (2), 'in vivo' (2), 'in the' (2), 'the nobel' (2), 'nobel prize' (2), 'crispr gene' (1), 'gene editing' (1), 'editing crispr' (1), 'crispr pronounced' (1), 'pronounced ■kr■sp■r' (1)
2-gram Repetition Rate: 0.0231

Common 3-grams (Top 10):

'the nobel prize' (2), 'crispr gene editing' (1), 'gene editing crispr' (1), 'editing crispr pronounced' (1), 'crispr pronounced ■kr■sp■r' (1), 'pronounced ■kr■sp■r crisper' (1), '■kr■sp■r crisper refers' (1), 'crisper refers to' (1), 'refers to a' (1), 'to a clustered' (1)
3-gram Repetition Rate: 0.0047

Common 4-grams (Top 10):

'crispr gene editing crispr' (1), 'gene editing crispr pronounced' (1), 'editing crispr pronounced ■kr■sp■r' (1), 'crispr pronounced ■kr■sp■r crisper' (1), 'pronounced ■kr■sp■r crisper refers' (1), '■kr■sp■r crisper refers to' (1), 'crisper refers to a' (1), 'refers to a clustered' (1), 'to a clustered regularly' (1), 'a clustered regularly interspaced' (1)
4-gram Repetition Rate: 0.0000

Character N-gram Analysis:

Common Character 3-grams (Top 10):

```
'ed_' (18), '_th' (15), 'the' (14), 'he_' (13), '_in' (12), '_ge' (10), 'gen' (10), 'ing' (9),
'ng_' (9), 'in_' (9)
```

Common Character 4-grams (Top 10):

```
'the_' (13), '_the' (12), '_gen' (9), 'ing_' (9), '_in_' (9), 'gene' (6), '_of_' (6), '_and'
(6), 'and_' (6), '_is_' (5)
```

Common Character 5-grams (Top 10):

```
'_the_' (12), '_gene' (6), '_and_' (6), 'crisp' (4), 'genet' (4), 'eneti' (4), 'netic' (4),
'etic_' (4), 'techn' (4), 'ation' (4)
```