LITERARY INSIGHTS GROUP

Austen Under the Microscope

Presented by: Youssif Hebesha 10002566
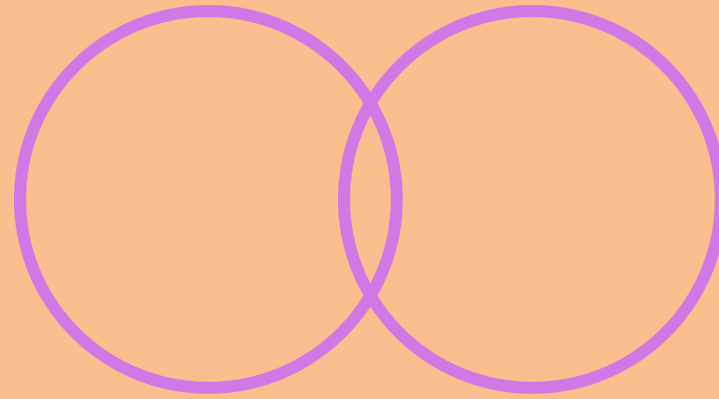Belal Soliman 100001754

Jane Austen's Background

Jane Austen wrote during the **Regency era** in England, a time of social change, highlighting themes of love, marriage, and class through her novels, including Pride and Prejudice and Sense and Sensibility.
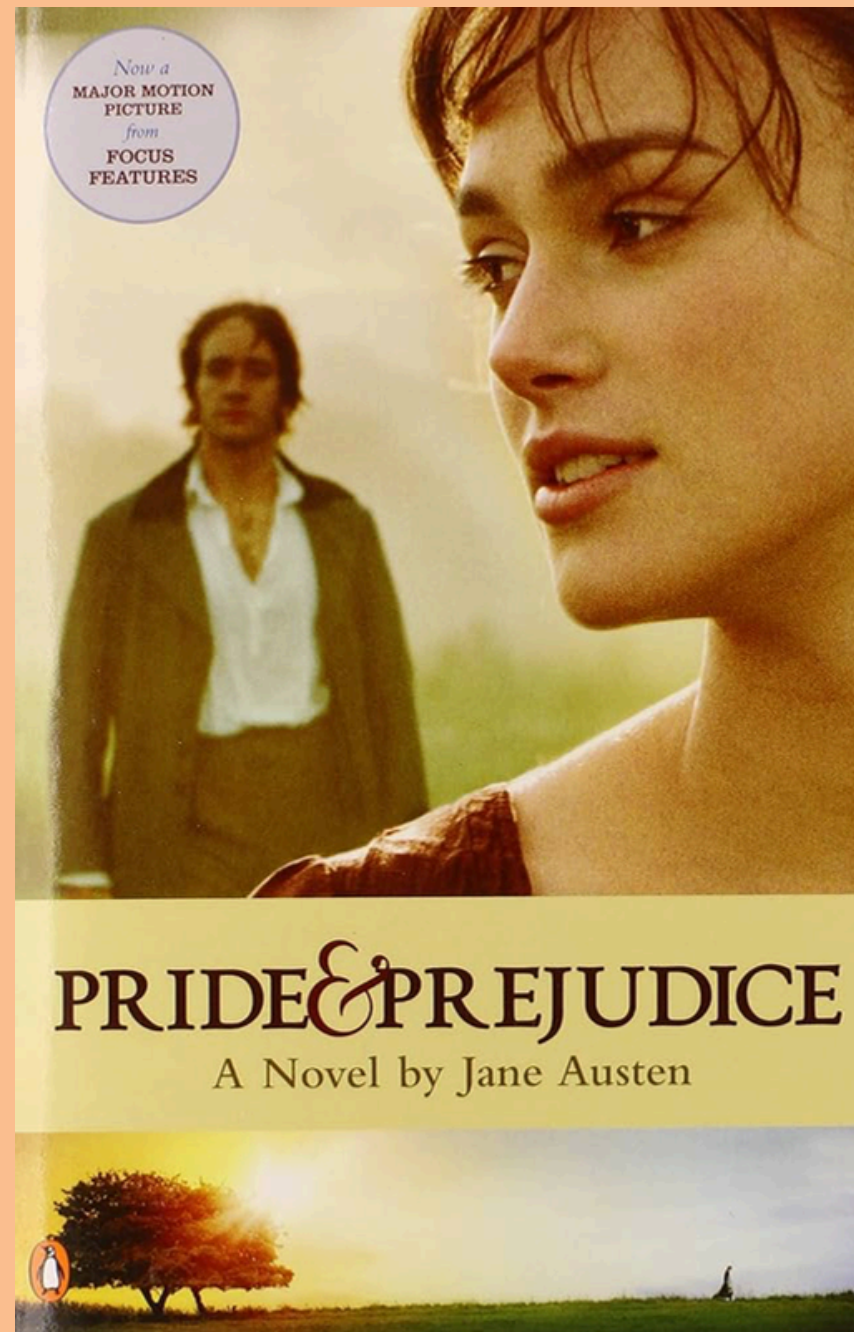
Her work focuses on:
- Social class, relationships, marriage
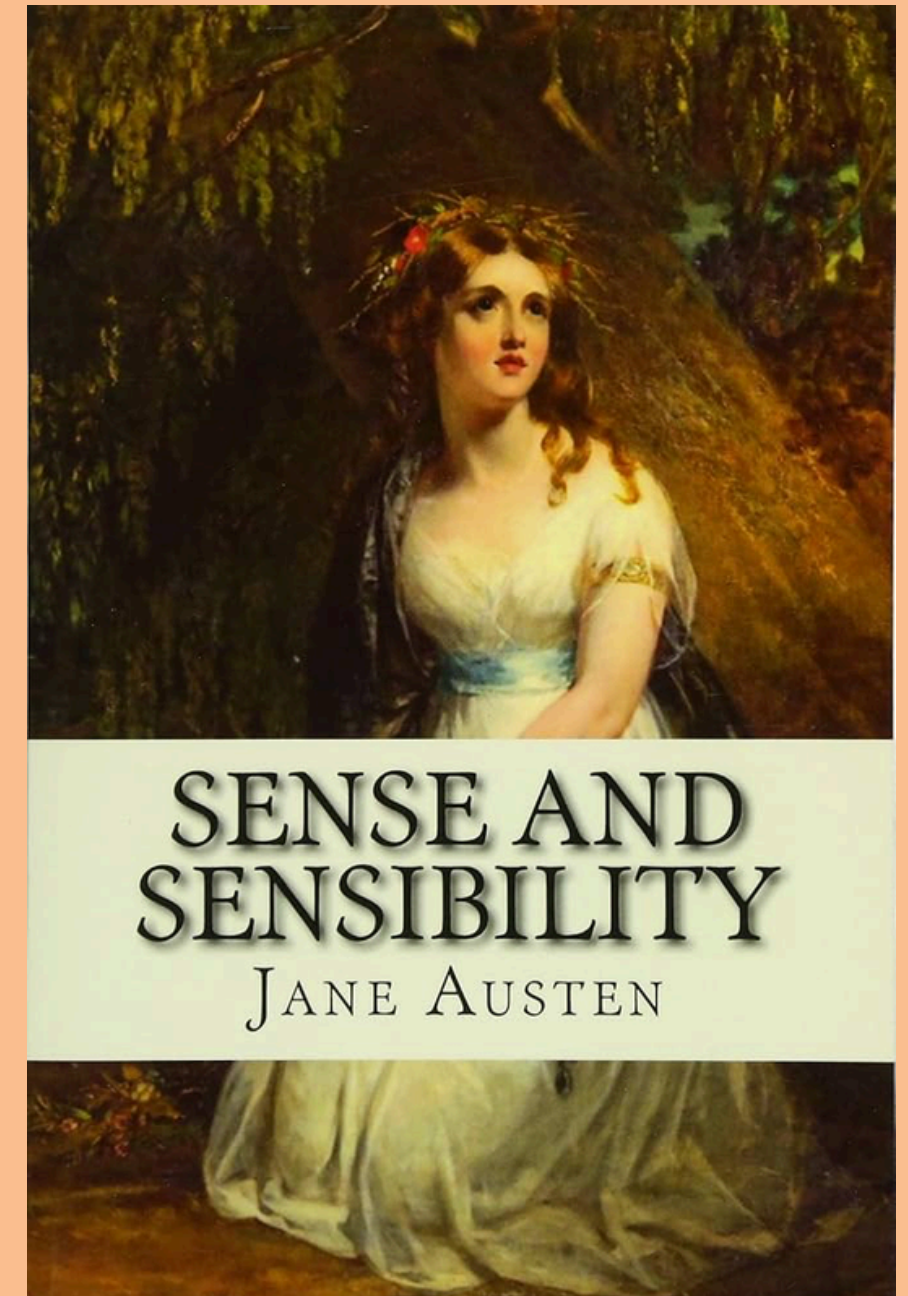- Human behavior and moral development

Novels

Both novels explore **themes of love**, societal norms, and personal growth.

01

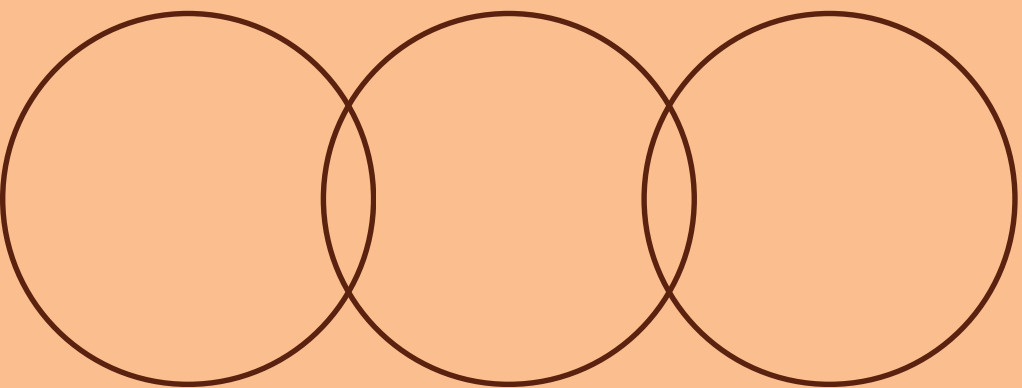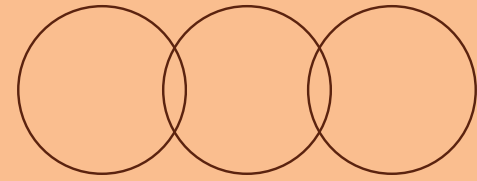**Pride and Prejudice** centers on Elizabeth Bennet's dynamic journey to love.

02

Sense and Sensibility follows sisters Elinor and Marianne Dashwood's contrasting romantic experiences.

Notebooks Overview

Analysing Austen's novels

- Sentence length and distribution

- Readability scores

- Topic drift between paragraphs

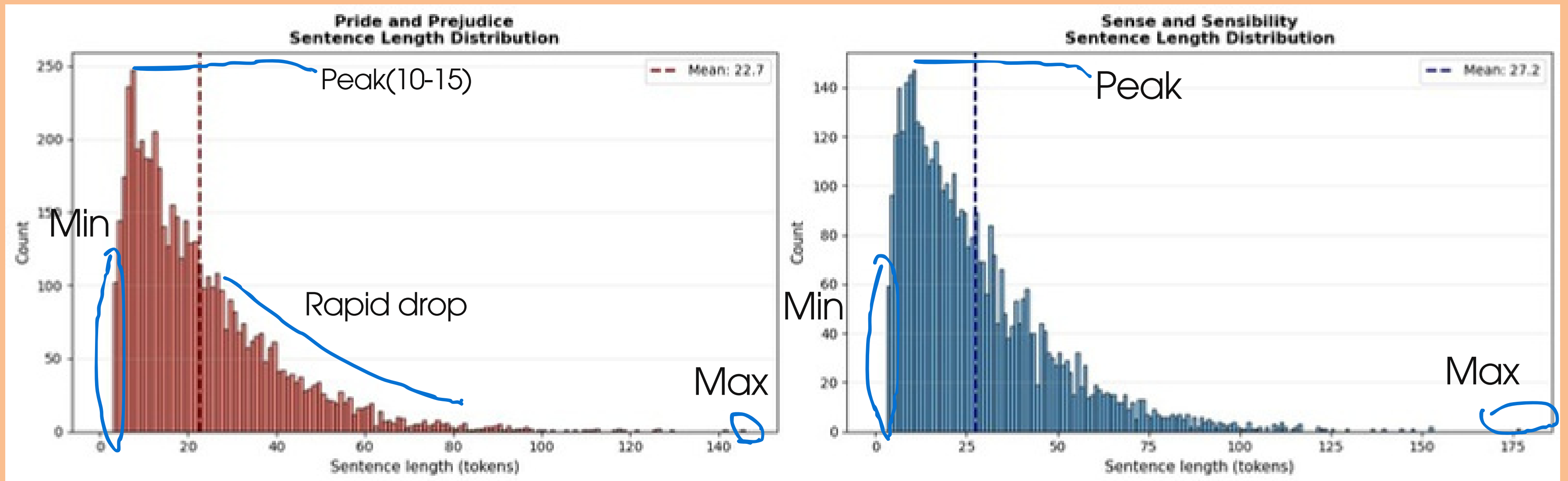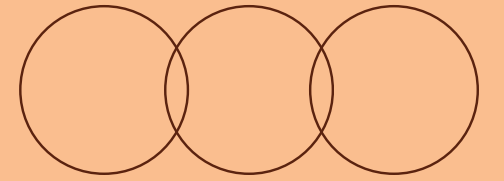- Paragraph function classification

1. Sentence Length and Distribution

Sentence-Level Analysis

This notebook helps you split your text into sentences and compute:

- number of sentences
- average sentence length (in tokens and in characters)
- distribution / histogram of sentence lengths
- comparison across two texts/authors

Graphs

**Pride and Prejudice
Sentence Length Distribution**

-- Mean: 22.7

Peak(10-15)

Min

Rapid drop

Max

Count

Sentence length (tokens)

**Sense and Sensibility
Sentence Length Distribution**

-- Mean: 27.2

Peak

Min

Max

Count

Sentence length (tokens)

Analysis
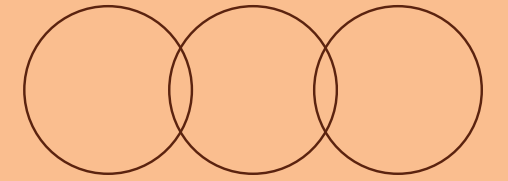
Pride and Prejudice

- Average sentence length is shorter (~22.7 tokens).
- Distribution is tightly clustered around 10–20 tokens.
- Fewer very long sentences → more concise, controlled style.
- Writing feels structured, witty, and direct.

Sense and Sensibility

- Average sentence length is longer (~27.2 tokens).
- Wider variation in sentence lengths, with many sentences over 40 tokens.
- Heavier long-sentence tail → more descriptive and elaborate prose.
- Style shows greater complexity and verbosity.

Comparison

- Pride and Prejudice uses more short sentences.
- Sense and Sensibility has more long, multi-clause sentences.
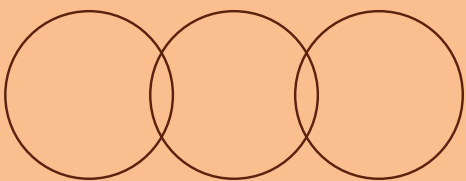- Overall, SS demonstrates broader stylistic range, while PP is tighter and more concise.
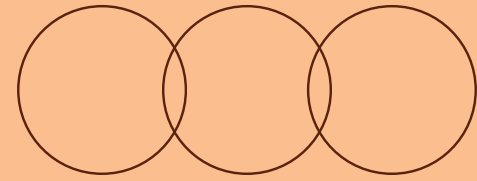
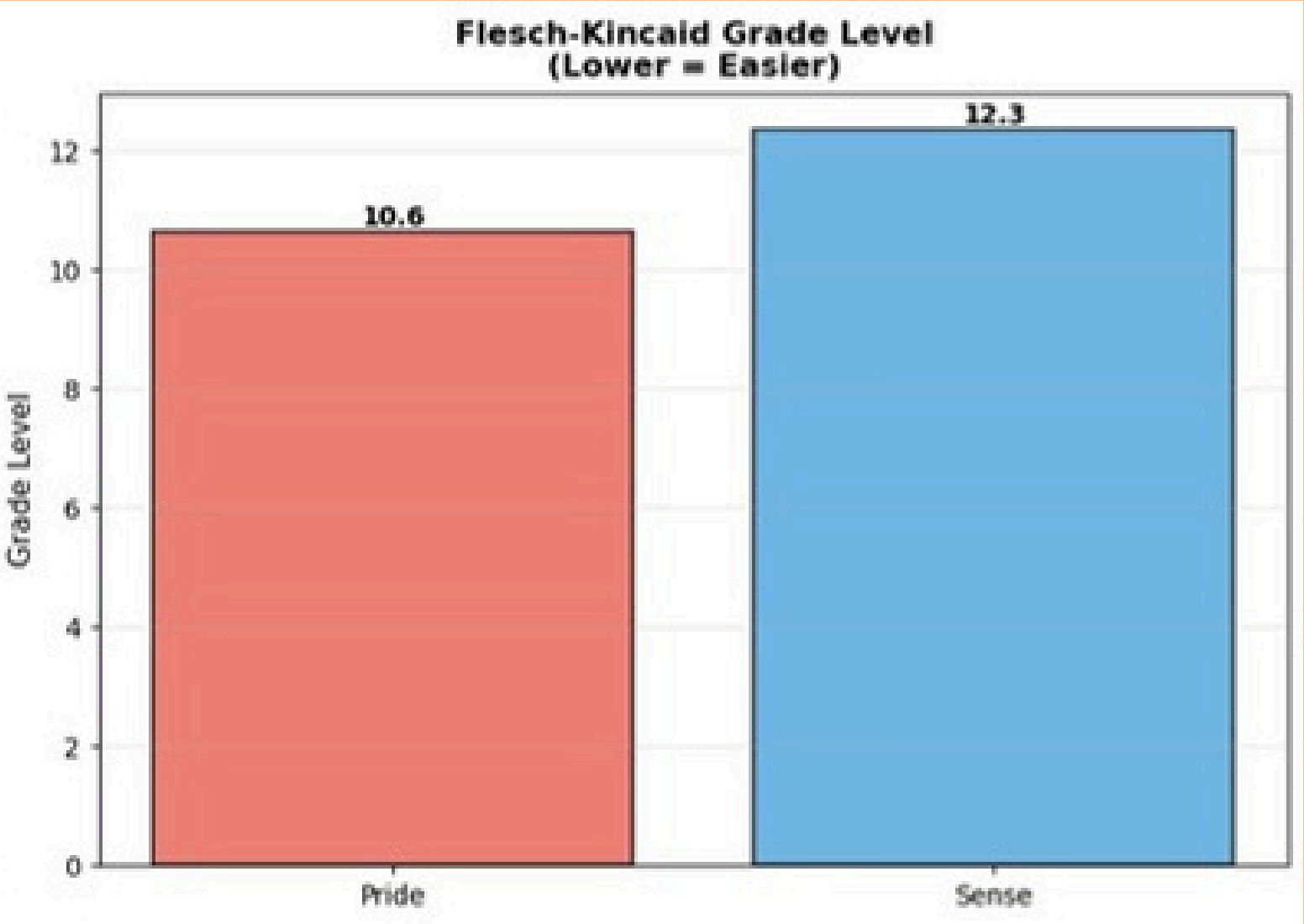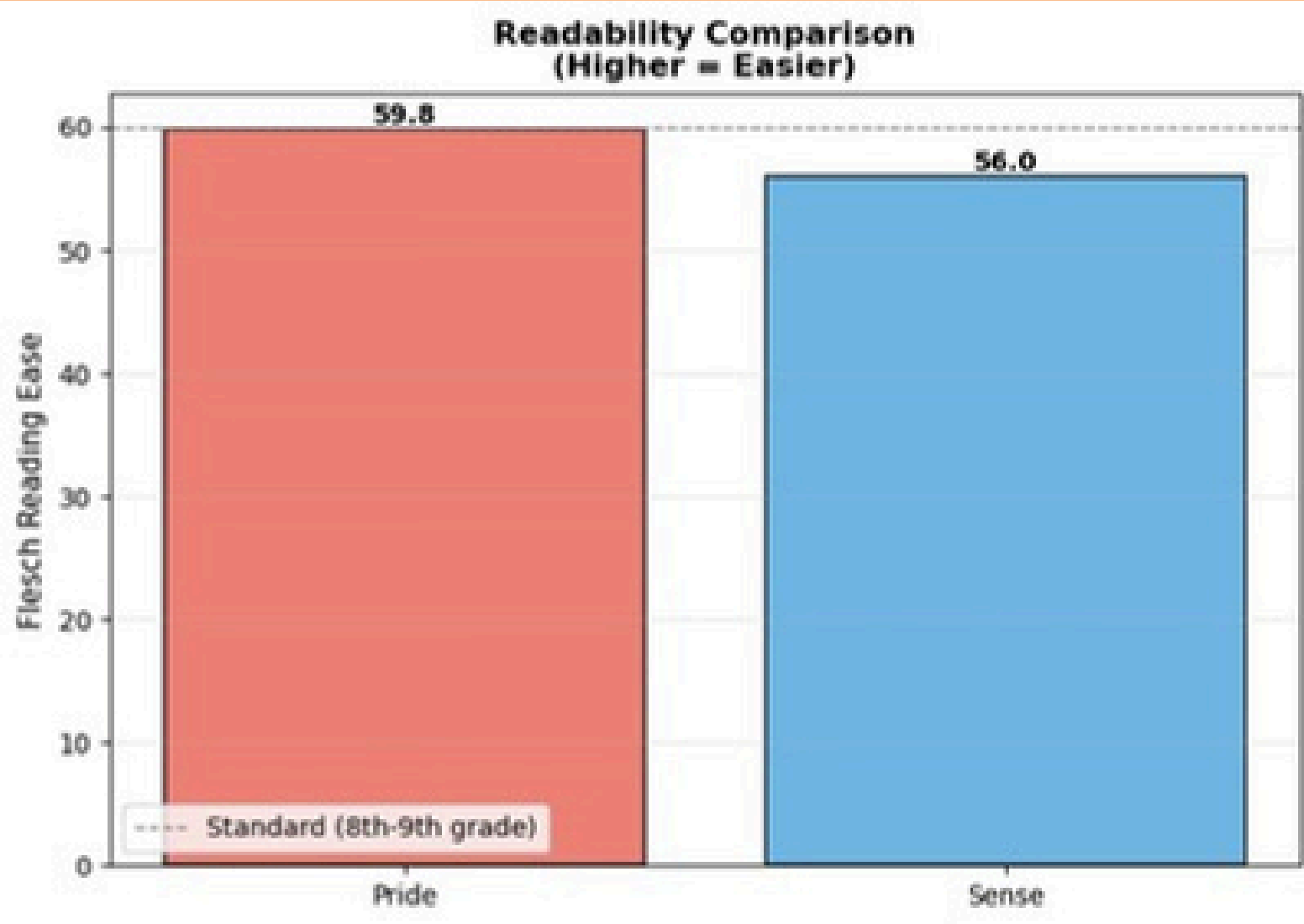# 2. Readability scores

Sentence-Level Analysis

This notebook computes classic readability formulas for:
- Flesch Reading Ease
- Flesch-Kincaid Grade Level

Graphs

Flesch Reading Ease

Score	Grade	Avg. Words Per Sentence	Syllables Per 100 words
90 - 100	5	8	123
80 - 90	6	11	131
70 - 80	7	14	139
60 - 70	8 - 9	17	147
50 - 60	10 - 12	21	155
30 - 50	College	25	167
0 - 30	College Grad	29	192

Readability Comparison (Higher = Easier) — Pride: 59.8, Sense: 56.0; Standard (8th-9th grade)

Flesch-Kincaid Grade Level (Lower = Easier) — Pride: 10.6, Sense: 12.3

Analysis
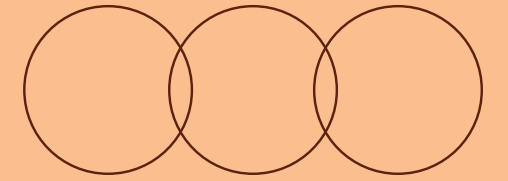
Flesch Reading Ease (Higher = Easier)

- Pride and Prejudice scores 59.8, slightly above the "standard" readability range.
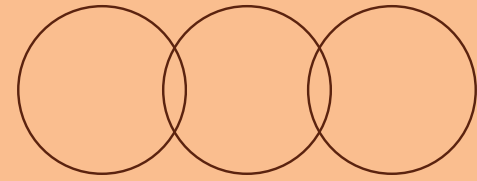- Sense and Sensibility scores 56.0, meaning it is more difficult to read than PP.
- Both novels fall within the 8th–10th grade readability band, but PP is moderately more accessible.

Flesch–Kincaid Grade Level (Lower = Easier)

- Pride and Prejudice: 10.6 grade level.
- Sense and Sensibility: 12.3 grade level, indicating college-level difficulty.
- SS requires more advanced reading skills, consistent with its longer, more complex sentence structures.

Interpretation / What This Means

- Pride and Prejudice is easier to read overall, aligning with its shorter, more consistent sentences.
- Sense and Sensibility is more challenging, reflecting a more elaborate and descriptive narrative style.
- The readability metrics reinforce the patterns seen in the sentence-length analysis:
- PP = concise and accessible, SS = more verbose and sophisticated.

3. Topic drift between analysis

## Paragraph-level analysis

In this notebook, We:
- represent each paragraph as a MiniLM embedding
- compute cosine similarity between consecutive paragraphs
- interpret low similarity as strong topic/scene shifts
- connect this to how LLM systems segment long documents
  into chunks for retrieval and long-context reasoning.

Graphs

Distribution of Topic Continuity (higher = smoother transitions)
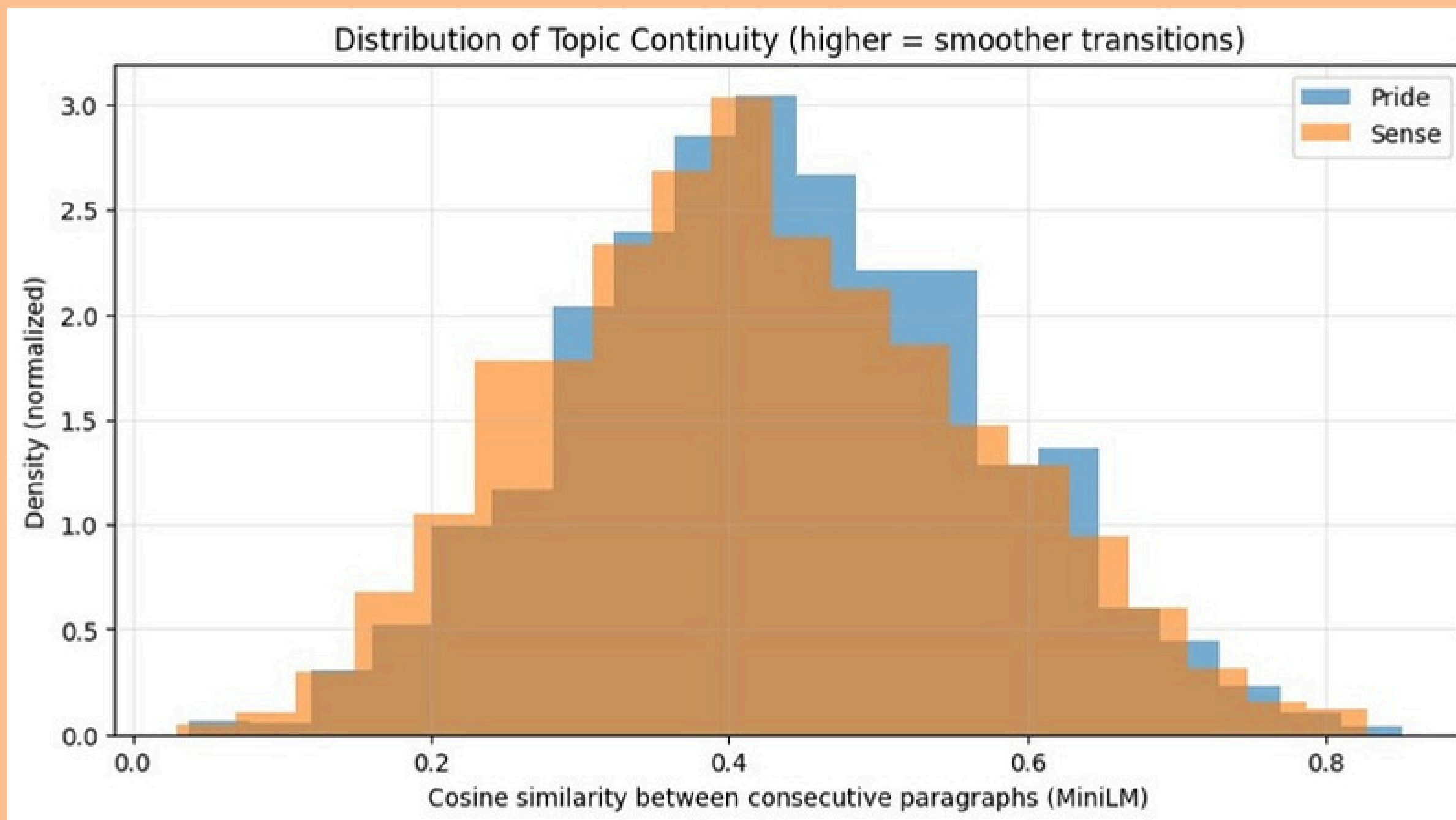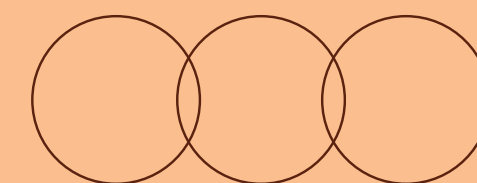
Analysis
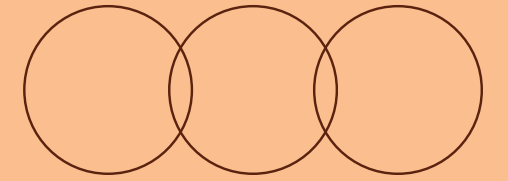
Key Findings

- Both novels have a similar overall distribution of topic continuity, peaking between 0.35–0.45, indicating moderately smooth transitions.
- Pride and Prejudice shows slightly higher similarity scores on average, meaning its paragraphs tend to stay more closely connected in topic.
- Sense and Sensibility has a wider spread and more lower-similarity transitions, suggesting more frequent shifts in focus or scene.
- Sense also shows a small tail toward higher continuity, indicating occasional long stretches of closely related content — consistent with descriptive or reflective passages.

Interpretation

- PP's smoother transitions align with its dialogue-driven, socially interactive scenes where conversations unfold continuously.
- SS's greater variation reflects its more descriptive style and shifts between emotional reflection, narration, and scene changes.
- Overall, PP feels more tightly woven, while SS exhibits more variation in narrative pacing and topical focus.
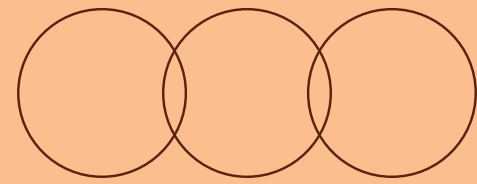
4. Paragraph function classification

Paragraph-level analysis

In this notebook, We:
- represent each paragraph as a MiniLM embedding
- compute cosine similarity between consecutive paragraphs
- interpret low similarity as strong topic/scene shifts
- connect this to how LLM systems segment long documents
into chunks for retrieval and long-context reasoning.

Graphs

Pride paragraph functions (MiniLM) — Sense paragraph functions (MiniLM)
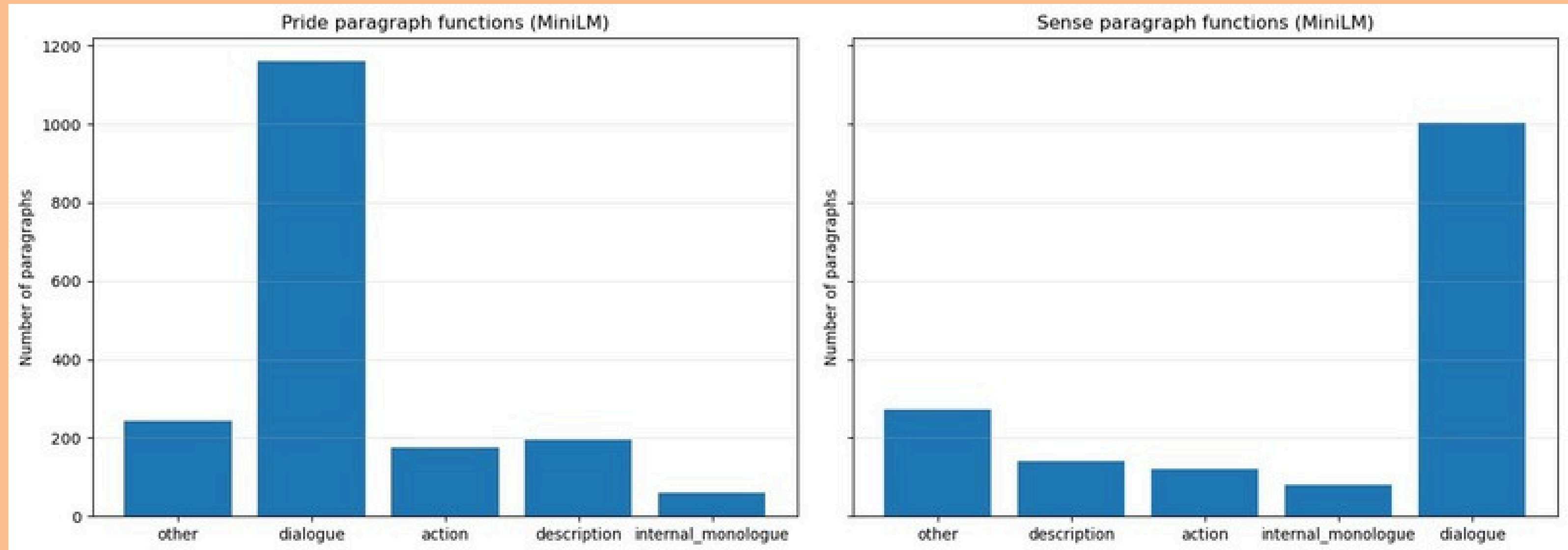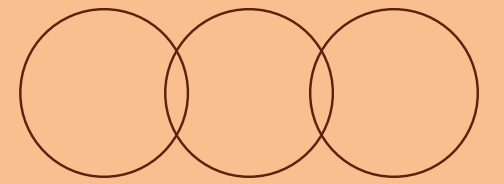
Analysis
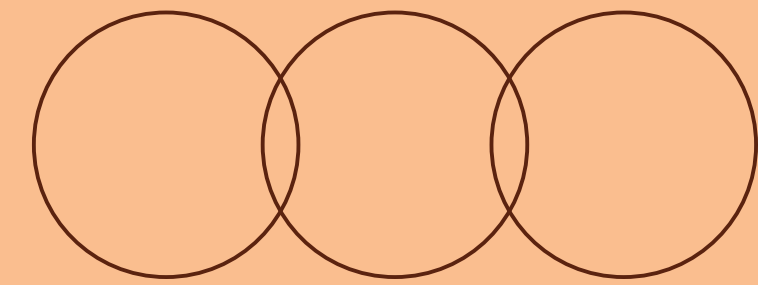
Pride and Prejudice

- Dialogue dominates the novel with ~1150 paragraphs, far more than any other category.
- Heavy use of dialogue supports the novel's social dynamics, character interaction, and wit.
- Description and action appear at moderate levels (~180–200 paragraphs each).
- Internal monologue is relatively low (~70), suggesting Austen reveals characters mostly through interaction rather than introspection.

Sense and Sensibility

- Dialogue remains the largest category, but at ~1000 paragraphs, it's less dominant than in PP.
- Greater balance across paragraph types:
-
    - More description (~150) and action (~120) compared to PP.
    - Slightly more internal monologue, reflecting increased emotional introspection.
-
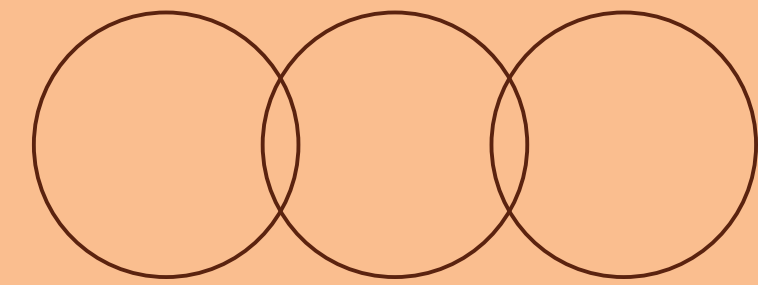- The distribution shows less reliance on dialogue and more narrative exposition and reflection.

Comparative Insights

- Pride and Prejudice is more dialogue-driven, emphasizing social exchange, conversation, and character interplay.
- Sense and Sensibility has a more evenly distributed narrative style, with increased description and internal reflection.
- SS's higher introspective content aligns with its themes of emotion, sensitivity, and internal conflict.
- PP's dominant dialogue aligns with its tone of wit, irony, and social commentary.

Overall Analysis

Pride and Prejudice

-
    - Shorter, more consistent sentences
    - Easier readability
    - Strongly dialogue-driven
    - Smoother topic transitions
    - → A clear, focused, socially dynamic narrative style

-

Sense and Sensibility

-
    - Longer, more variable sentences
    - Higher reading difficulty
    - More description & introspection
    - Greater variation in topic continuity
    - → A richer, more reflective and descriptive style

-

In summary

Austen uses precision and social wit in Pride and Prejudice, while Sense and Sensibility showcases *emotional depth and narrative complexity. The differences highlight her versatility as a writer.

Will You Read Jane Austen?

100002566@stud.srh-university.de
100001754@stud.srh-university.de