Pride and Prejudice

Jane Austen

1.1 Corpus metrics

Linguistic metric	Measured value	Interpretation & reference
Total word count (full text incl. appendix)	121,795 words (687,856 characters)	Full text including the appendix. (~487.2 norm pages à 250 words / ~458.6 norm pages à 1,500 characters).
Main text word count (prose only)	121,608 words (686,949 characters)	Prose only. (61 chapters).
Total sentences (main text)	5,771 sentences	Total sentences of the main text.
Average sentence length (ASL)	20.43 words / sentence	Average sentence length of the main text.
Median sentence length	16 words	Median sentence length.
Sentence-length standard deviation	15.89 words	Spread of sentence lengths.
Lexical diversity (TTR)	0.0538 (V = 6,340 / N = 117,916)	Lexical diversity (type-token ratio).
Guiraud index R (V/√N)	18.46	V/√N; still sensitive to text length and topic.
Yule's Characteristic K	81.11	Vocabulary stability across text length.
MTLD (Textual Lexical Diversity)	102.6	Length-robust lexical diversity; short samples remain unstable.
MATTR (Moving-Average TTR)	0.827	Window-stable type-token ratio (Covington & McFall 2010).
Maas a^2	0.0494	Vocabulary concentration (Maas 1972; lower = richer).
Flesch Reading Ease	64.3	Language-calibrated readability formula of the active profile.
LIX (readability index)	40.0	Readability index (LIX).
Dialogue share (quoted speech)	49,820 words (42.25%)	Share of quoted speech in the prose.
Prose paragraphs	2072 (56.8 words/paragraph)	Paragraph economy of the main text.

1.2 Sentence-length architecture & rhythm

The empirical sentence distribution of the main text:

Sentence-length category	Criterion	Sentences	Share	Interpretation limit
Short sentences (staccato)	≤ 6 words	828	14.3%	Short sentences; pace requires contextual review.
Medium sentences	7–15 words	1,983	34.4%	Sentence-length category, not a narrative judgement.
Long sentences	16–25 words	1,350	23.4%	Sentence-length category, not a narrative judgement.
Very long sentences	> 25 words	1,610	27.9%	Length does not establish syntactic nesting.

1.3 Punctuation as a stylistic seismograph

Punctuation mark	Frequency	Density (per 1,000 words)	Interpretation limit
Periods (.)	5,961	49.0	Basic sentence rhythm.
Commas (,)	9,178	75.5	Enumerations and appositions.
Dashes (–/—)	0	0.0	Afterthought, self-correction, insertion.
Colons (:)	194	1.6	Announcing lists, quotes, insights.
Semicolons (;)	1,614	13.3	Rare; avoids a lecturing tone.
Question marks (?)	473	3.9	Self-questioning and dialogue.
Exclamation marks (!)	422	3.5	Used sparingly.
Ellipses (…/...)	0	0.0	Breaking off a thought.

1.4 Chapter comparison matrix (linguistic deep profiles)

Ch.	Title	Words	Sentences	ASL	Dialogue %	TTR	Filter verbs	Dominant tense

Ch.	Title	Words	Sentences	ASL	Dialogue %	TTR	Filter verbs	Dominant tense
01	Chapter I	849	58	14.6	78.8%	0.378	2	Present
02	Chapter II	798	52	15.3	71.2%	0.387	1	Mixed
03	Chapter III	1,698	84	20.2	34.0%	0.335	3	Past
04	Chapter IV	1,062	53	20.0	32.5%	0.391	7	Past
05	Chapter V	958	51	18.8	68.4%	0.411	6	Past
06	Chapter VI	2,347	114	20.6	59.2%	0.310	6	Mixed
07	Chapter VII	1,845	91	20.3	36.9%	0.336	8	Past
08	Chapter VIII	1,929	106	18.2	56.8%	0.342	10	Past
09	Chapter IX	1,731	103	16.8	58.8%	0.343	8	Mixed
10	Chapter X	2,212	122	18.1	68.3%	0.325	10	Mixed
11	Chapter XI	1,602	95	16.9	53.6%	0.365	2	Mixed
12	Chapter XII	676	22	30.7	0.0%	0.466	4	Past
13	Chapter XIII	1,313	75	17.5	57.8%	0.354	7	Mixed
14	Chapter XIV	1,117	52	21.5	48.0%	0.433	2	Past
15	Chapter XV	1,699	47	36.1	3.0%	0.367	6	Past
16	Chapter XVI	3,377	150	22.5	61.6%	0.289	16	Past
17	Chapter XVII	1,279	47	27.2	31.4%	0.403	9	Past
18	Chapter XVIII	5,178	210	24.7	40.5%	0.246	21	Past
19	Chapter XIX	1,916	78	24.6	80.5%	0.341	4	Mixed
20	Chapter XX	1,646	87	18.9	58.5%	0.341	2	Mixed
21	Chapter XXI	2,001	77	26.0	59.4%	0.342	9	Mixed
22	Chapter XXII	1,731	64	27.0	21.5%	0.362	10	Past
23	Chapter XXIII	1,612	58	27.8	11.9%	0.366	6	Past

Ch.	Title	Words	Sentences	ASL	Dialogue %	TTR	Filter verbs	Dominant tense
24	Chapter XXIV	1,928	104	18.5	60.2%	0.333	19	Mixed
25	Chapter XXV	1,515	75	20.2	43.7%	0.374	10	Past
26	Chapter XXVI	1,845	95	19.4	49.3%	0.331	10	Mixed
27	Chapter XXVII	1,259	65	19.4	39.4%	0.419	8	Past
28	Chapter XXVIII	1,448	66	21.9	23.1%	0.391	3	Past
29	Chapter XXIX	2,384	124	19.2	33.5%	0.327	11	Past
30	Chapter XXX	1,228	35	35.1	3.3%	0.392	3	Past
31	Chapter XXXI	1,557	82	19.0	58.6%	0.349	4	Mixed
32	Chapter XXXII	1,497	66	22.7	39.1%	0.352	4	Past
33	Chapter XXXIII	1,862	97	19.2	53.2%	0.338	9	Past
34	Chapter XXXIV	2,101	102	20.6	44.3%	0.340	16	Past
35	Chapter XXXV	3,001	109	27.5	0.9%	0.317	8	Past
36	Chapter XXXVI	2,011	73	27.5	6.3%	0.351	10	Past
37	Chapter XXXVII	1,371	69	19.9	38.6%	0.389	9	Past
38	Chapter XXXVIII	1,038	44	23.6	45.0%	0.425	5	Past
39	Chapter XXXIX	1,630	95	17.2	60.2%	0.347	15	Past
40	Chapter XL	1,679	107	15.7	70.6%	0.329	4	Mixed
41	Chapter XLI	2,269	110	20.6	40.1%	0.351	9	Past
42	Chapter XLII	1,831	63	29.1	9.0%	0.370	8	Past
43	Chapter XLIII	4,858	222	21.9	22.7%	0.235	28	Past
44	Chapter XLIV	2,336	72	32.4	1.5%	0.316	6	Past
45	Chapter XLV	1,730	55	31.5	15.4%	0.371	4	Past
46	Chapter XLVI	3,022	156	19.4	51.1%	0.299	10	Past

Ch.	Title	Words	Sentences	ASL	Dialogue %	TTR	Filter verbs	Dominant tense
47	Chapter XLVII	3,863	196	19.7	75.4%	0.251	18	Past
48	Chapter XLVIII	1,901	85	22.4	26.7%	0.334	6	Past
49	Chapter XLIX	1,841	137	13.4	62.4%	0.300	6	Mixed
50	Chapter L	2,201	87	25.3	5.1%	0.334	6	Past
51	Chapter LI	2,071	111	18.7	29.3%	0.310	14	Past
52	Chapter LII	1,238	86	14.4	46.4%	0.389	5	Past
53	Chapter LIII	2,941	159	18.5	42.2%	0.277	12	Past
54	Chapter LIV	1,594	93	17.1	30.7%	0.354	11	Past
55	Chapter LV	2,325	136	17.1	30.8%	0.302	4	Past
56	Chapter LVI	2,732	195	14.0	74.7%	0.288	7	Mixed
57	Chapter LVII	1,653	77	21.5	54.5%	0.363	9	Past
58	Chapter LVIII	2,432	135	18.0	66.5%	0.311	22	Past
59	Chapter LIX	2,473	169	14.6	51.1%	0.288	11	Mixed
60	Chapter LX	1,544	84	18.4	63.6%	0.363	4	Past
61	Chapter LXI	1,129	38	29.7	0.1%	0.414	1	Past