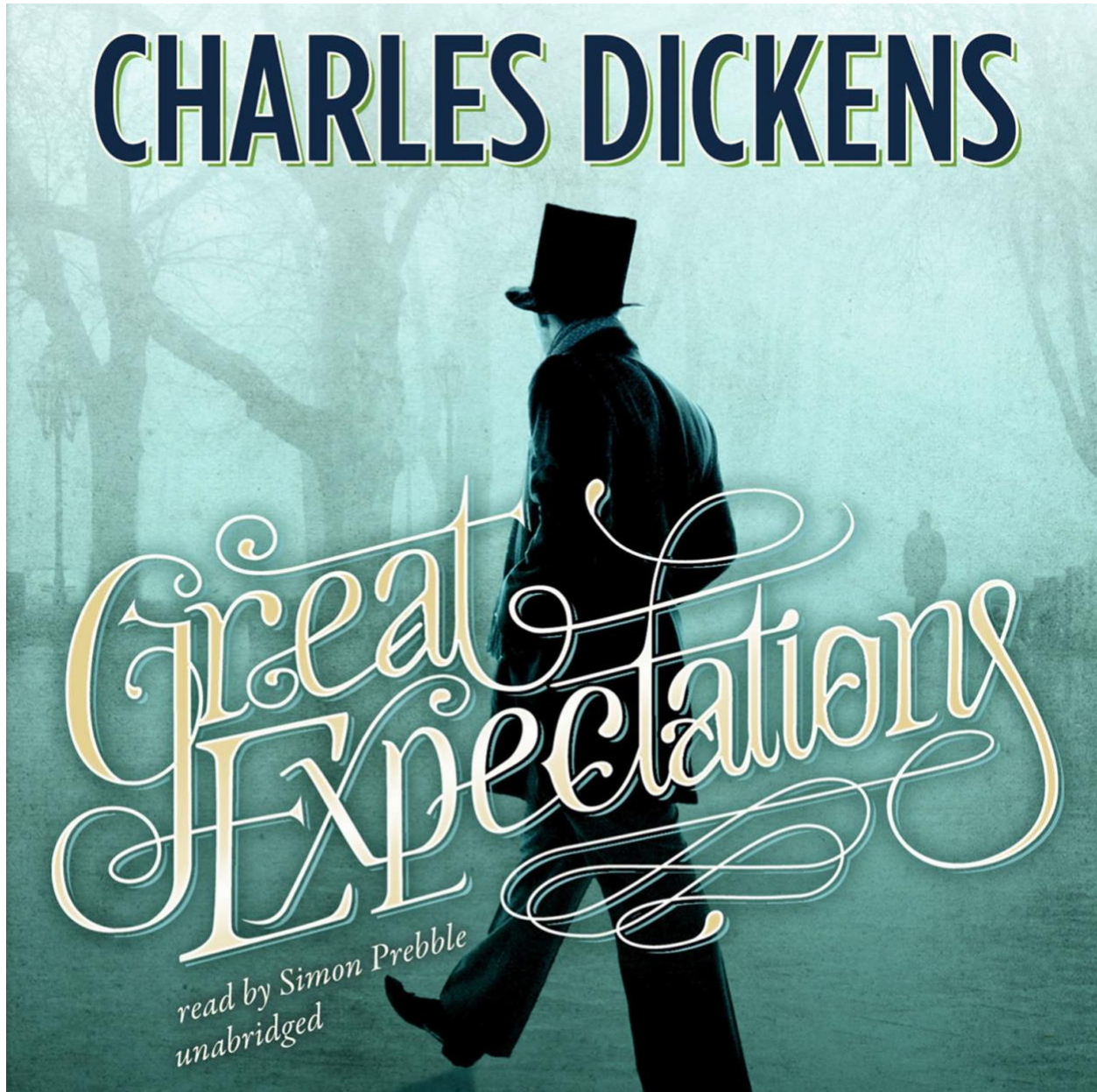# PROJECT GUTENBERG ANALYSIS

[NOVEL ANALYZED: GREAT EXPECTATIONS BY CHARLES DICKENS]

I.   **Total words in novel:** 184,377

II.   **Total unique words in novel:** 12,214

III.   **20 Most Frequent words:**

1. The 8121
2. And 6954
3. I 6418
4. To 5051
5. Of 4421
6. A 4007
7. In 2989
8. That 2958
9. Was 2822
10. It 2632
11. He 2196
12. You 2167
13. Had 2089
14. My 2064
15. Me 1978
16. His 1855
17. With 1750
18. As 1746
19. At 1633
20. On 1403

IV.   **20 Most Interesting Frequent words:**
1. Mr 711
2. Joe 684
3. Into 494
4. Upon 363
5. Looked 323
6. Pip 317
7. Herbert 289
8. Dont 282
9. Being 261
10. Wemmick 254
11. Havisham 240
12. Going 237
13. Myself 232
14. Estella 231
15. Biddy 223

16. Jaggers 208
17. Hands 180
18. Eyes 177
19. Himself 177
20. Looking 177

V. **20 Least Frequent Words:**
   1. Aback 1
   2. Abased 1
   3. Abashed 1
   4. Abbey 1
   5. Abed 1
   6. Aberdeen 1
   7. Aberration 1
   8. Abet 1
   9. Abeyance 1
   10. Abhorring 1
   11. Abided 1
   12. Ability 1
   13. Ablazing 1
   14. Ablutions 1
   15. Abode 1
   16. Abominate 1
   17. Aboutas 1
   18. Abraham 1
   19. Abrupt 1
   20. Abruptness 1

VI. **Interesting word usage pattern:**
The word "Havisham" which refers to the character Mrs. Havisham has an interesting pattern where we can see that the name is references more frequently towards the middle of the book than the beginning and the end, and references extra frequently in certain chapters – probably because those chapters had a lot of character buildup for Mrs. Havisham.

[0, 0, 0, 0, 0, 0, 5, 13, 6, 0, 24, 12, 13, 1, 13, 0, 5, 2, 12, 0, 0, 18, 0, 0, 0, 0, 4, 0, 28, 1, 0, 1, 4, 0, 0, 1, 1, 23, 0, 1, 0, 0, 1, 28, 0, 0, 0, 1, 12, 0, 5, 2, 0, 0, 0, 0, 3, 0, 0]

VII. **Favorite Quote:**

"You acted noble, my boy" – Chapter 39

VIII.   Generated Sentence:

"the said and to a he in that of it was you had my at me his on with as"

IX.   Finding closest matching quote:

"Acted Noble Boy" -> "You acted noble, my boy"