

Sample Research Essay 1

The Iron Heel: Keywords and Collocations

Introduction

The Iron Heel is a dystopian novel by Jack London published in 1908. The text of the novel was analysed to find the keywords. The ten most key words were then studied to find their five most frequent collocations. It was hypothesised that these key words and collocations would be used negatively or be negative in meaning as the novel is about the violent beginnings of a dystopia.

Theoretical Background

Corpora are large electronic collections of texts which can be analysed to see how the words and grammar are used in these texts and how frequent particular words and combinations of words occur (Bowker and Pearson, 2002). Key words are words which appear in a text more frequently than they do in larger corpora (Scott and Tribble, 2006). Key words are thus indicative of the topic of the text. To find key words you have to create frequency word lists of the text(s) of interest and compare them with the frequency word lists of the reference corpus (Bowker and Pearson, 2002). This will give you a list of key words ranked by keyness. Collocations are words which occur together more frequently than by chance (Halliday, 1966).

Method

The text file of The Iron Heel taken from The Gutenberg Project [<https://www.gutenberg.org/ebooks/1164>] was topped and tailed before analysis. The contents page was removed from the beginning, and the Gutenberg licence text from the end was cut. This file was uploaded into the WordSmith Tools program [<https://lexically.net/wordsmith/>]. A frequency word list was created of this file.

As the novel is American English an American reference corpus word list was needed a sample of the Corpus of Contemporary American English (COCA) was downloaded from <https://www.corpusdata.org/formats.asp>. The Fiction file from the sample was used to create the reference corpus wordlist. The Key Word function on WordSmith Tools was then used to create a key word list for The Iron Heel (Key Word List 1).

This list was then edited to remove proper nouns, verbs, function words and prepositions from the list, to produce Key Word List 1 (edited). These words were removed as they tell us very little of interest about the novel. Obviously, the name of the main character(s) will occur more frequently in the novel than in the reference corpus.

The text file of The Iron Heel was then analysed with WordSmith Tools to produce a concordance of the novel. The five most common collocations for the five most key words from Key Word List 1 (edited) were then identified from the concordance list. These combinations were then evaluated as being either positive or negative in direct meaning or in connotation.

Results

The Word List for The Iron Heel (see Figure 1) shows the conventional frequencies of such texts; *the* is the most common, most frequent word, as expected.

Sample Research Essay 1

the iron heel book for analysis edited.lst

N	Word	Freq.	%	Texts	% emmas	Set
1	THE	6 512	7,30	1	100,00	
2	AND	3 405	3,82	1	100,00	
3	OF	3 314	3,71	1	100,00	
4	TO	1 927	2,16	1	100,00	
5	IN	1 777	1,99	1	100,00	
6	WAS	1 749	1,96	1	100,00	
7	A	1 598	1,79	1	100,00	
8	I	1 366	1,53	1	100,00	
9	THAT	1 364	1,53	1	100,00	
10	IT	1 200	1,34	1	100,00	
11	HE	1 080	1,21	1	100,00	
12	YOU	768	0,86	1	100,00	
13	HIS	740	0,83	1	100,00	
14	IS	687	0,77	1	100,00	
15	FOR	686	0,77	1	100,00	
16	HAD	674	0,76	1	100,00	
17	WERE	674	0,76	1	100,00	
18	NOT	592	0,66	1	100,00	
19	WITH	577	0,65	1	100,00	
20	AS	543	0,61	1	100,00	

Figure 1: The Iron Heel Frequency List: Top 20

Figure 2 shows the thirty most key words of the Key Word list created by comparing the Iron Heel Word List with the reference (COCO) Corpus Word List. The most common key word, the most key word, is Ernest, the first name of the protagonist Ernest Everhard.

the iron heel book for analysis edited.kws

N	Key word	Freq.	%	. Freq.	RC. %	synsyn	P
1	ERNEST	318	0,36	0		657,21	000000
2	OF	3 314	3,71	21 718	1,94	053,96	000000
3	LABOR	163	0,18	27		698,00	000000
4	THE	6 512	7,30	57 439	5,14	694,38	000000
5	HEEL	105	0,12	22		433,25	000000
6	BISHOP	87	0,10	4		420,97	000000
7	OLIGARCHY	77	0,09	0		401,08	000000
8	CLASS	161	0,18	180	0,02	394,72	000000
9	COMRADES	74	0,08	2		367,26	000000
10	AND	3 405	3,82	29 882	2,68	362,18	000000
11	WERE	674	0,76	3 635	0,33	333,03	000000
12	IRON	113	0,13	93		319,29	000000
13	EVERHARD	58	0,06	0		302,10	000000
14	STATES	86	0,10	44		288,32	000000
15	UPON	123	0,14	159	0,01	278,83	000000
16	WAS	1 749	1,96	14 069	1,26	275,50	000000
17	UNITED	78	0,09	42		257,35	000000
18	WILL	285	0,32	1 072	0,10	254,52	000000
19	OLIGARCHS	48	0,05	0		250,01	000000
20	TRUSTS	47	0,05	0		244,80	000000
21	CANNOT	47	0,05	1		235,23	000000
22	MERCENARIES	45	0,05	0		234,38	000000
23	SOCIALISTS	46	0,05	1		230,07	000000
24	ILL	44	0,05	0		229,17	000000
25	UNIONS	44	0,05	0		229,17	000000
26	GREAT	166	0,19	438	0,04	221,72	000000
27	REVOLUTION	56	0,06	15		220,77	000000
28	PLUTOCRACY	42	0,05	0		218,76	000000
29	SURPLUS	46	0,05	4		212,33	000000
30	REVOLT	43	0,05	3		202,25	000000

Figure 2: The Iron Heel KeyWord List

Sample Research Essay 1

This list was then edited with the removal of nouns, verbs, function words and prepositions – see Table 1. **Annex 1** explains why each word was removed from the list.

Raw Key Word List	Edited key Word List
1. ERNEST	1. LABOR
2. OF	2. OLIGARCHY
3. LABOR	3. CLASS
4. THE	4. COMRADES
5. HEEL	5. OLIGARCHS
6. BISHOP	6. TRUSTS
7. OLIGARCHY	7. MERCENARIES
8. CLASS	8. SOCIALISTS
9. COMRADES	9. UNIONS
10. AND	10. GREAT
11. WERE	11. REVOLUTION
12. IRON	12. PLUTOCRACY
13. EVERHARD	13. SURPLUS
14. STATES	14. REVOLT
15. UPON	15. CAPITALIST
16. WAS	16. PROLETARIAT
17. UNITED	17. SOCIETY
18. WILL	18. STRIKE
19. OLIGARCHS	19. REVOLUTIONISTS
20. TRUSTS	20. PROFITS
21. CANNOT	
22. MERCENARIES	
23. SOCIALISTS	
24. I'LL	
25. UNIONS	
26. GREAT	
27. REVOLUTION	
28. PLUTOCRACY	
29. SURPLUS	
30. REVOLT	
31. CAPITALIST	
32. PROLETARIAT	
33. IN	
34. SOCIETY	
35. OUR	
36. STRIKE	
37. REVOLUTIONISTS	
38. PROFITS	

TABLE 1: The Iron Heel 20 Most Key Words (edited)

At this stage the hypothesis that the key words and collocations would be used negatively or be negative in meaning as the novel is about the violent beginnings of a dystopia as stated in the introduction was revised. The five most key words were identified as shown in Table 2 and coded as positive, neutral or negative prior to an examination of the collocations for each word.

1. LABOR	Positive
2. OLIGARCHY	Negative
3. CLASS	Positive/Negative?
4. COMRADES	Positive
5. OLIGARCHS	Negative

Table 2: Top 5 Key Words coded as positive, neutral or negative

Sample Research Essay 1

The most common collocations (one word to the left or right of the key word) for these five words are shown in Table 3.

Key Word	Collocations
LABOR	labor castes organized labor labor leaders great labour labor troubles
OLIGARCHY	the Oligarchy
CLASS	working class middle class capitalist class your class my class class struggle
COMRADES	the comrades our comrades his comrades other comrades own comrades
OLIGARCHS	the Oligarchs

Table 3: Most Common Collocations of the Top 5 Key Words

Discussion

The collocations of *labor* and *class* are the most interesting, but it is difficult to identify them as positive, neutral or negative without a further study of the wider context. The key collocation for *oligarchy* and *oligarchs* is the same, *the*, the definite article. Interestingly the enemy of the people in the book is not qualified by any negative adjectives; they are not described as bad but are known as bad through their actions and the commentary of Meredith Everhard. Similarly, *comrades* are not described as good, or bad, or skilled (and so on) in any way in the book, but just identified by determiners.

Conclusion

Using WordSmith Tools it was possible to create a word list for *The Iron Heel* and then to produce a Key Word list for the text. After editing this list to remove certain types of words, the collocates for the top five key words were examined to see if their positive, neutral or negative aspect could be predicted with any degree of certainty. This study clearly shows that this was not possible, despite having read the novel. A reader's intuitions regarding this are not reliable and the data needs to be accessed directly and studied instead.

Word Count (excluding figures, tables and bibliography): 841 words

Bibliography

Bowker, L. and Pearson, J. (2002) *Working with Specialized language A practical guide to using corpora*, Routledge

Halliday, M.A.K. (1966) 'Lexis as a Linguistic Level', *Journal of Linguistics* 2(1) 1966: 57–67

Scott, M. & Tribble, C. (2006) *Textual Patterns: keyword and corpus analysis in language education*, Benjamins

Annex 1

Sample Research Essay 1

Raw Key Word List	Reason for Removal
1. ERNEST	name
2. OF	preposition
3. LABOR	
4. THE	determiner
5. HEEL	name
6. BISHOP	name
7. OLIGARCHY	
8. CLASS	
9. COMRADES	
10. AND	conjunction
11. WERE	verb
12. IRON	name
13. EVERHARD	name
14. STATES	name
15. UPON	preposition
16. WAS	verb
17. UNITED	name
18. WILL	verb
19. OLIGARCHS	
20. TRUSTS	
21. CANNOT	verb
22. MERCENARIES	
23. SOCIALISTS	
24. I'LL	pronoun + verb
25. UNIONS	name
26. GREAT	
27. REVOLUTION	
28. PLUTOCRACY	
29. SURPLUS	
30. REVOLT	
31. CAPITALIST	
32. PROLETARIAT	
33. IN	preposition
34. SOCIETY	
35. OUR	adjective
36. STRIKE	name
37. REVOLUTIONISTS	name
38. PROFITS	name