

Table 1. Summary of the 100 most common words in the corpus

Rank	Word	Frequency	Percentage	Category
1	the	10000	10.0%	Article
2	and	8500	8.5%	Conjunction
3	of	7500	7.5%	Preposition
4	is	6500	6.5%	Verb
5	in	5500	5.5%	Preposition
6	to	4500	4.5%	Particle
7	a	3500	3.5%	Article
8	that	3000	3.0%	Relative Pronoun
9	he	2500	2.5%	Pronoun
10	was	2000	2.0%	Verb
11	on	1800	1.8%	Preposition
12	it	1500	1.5%	Pronoun
13	from	1200	1.2%	Preposition
14	with	1000	1.0%	Preposition
15	but	900	0.9%	Conjunction
16	by	800	0.8%	Preposition
17	as	750	0.75%	Conjunction
18	at	700	0.7%	Preposition
19	an	650	0.65%	Article
20	are	600	0.6%	Verb
21	which	550	0.55%	Relative Pronoun
22	so	500	0.5%	Conjunction
23	if	450	0.45%	Conjunction
24	or	400	0.4%	Conjunction
25	not	350	0.35%	Adverb
26	all	300	0.3%	Quantifier
27	very	250	0.25%	Adverb
28	just	200	0.2%	Adverb
29	some	180	0.18%	Quantifier
30	each	150	0.15%	Quantifier
31	both	120	0.12%	Quantifier
32	few	100	0.1%	Quantifier
33	more	90	0.09%	Adverb
34	less	80	0.08%	Adverb
35	most	70	0.07%	Quantifier
36	only	60	0.06%	Adverb
37	too	50	0.05%	Adverb
38	very	40	0.04%	Adverb
39	just	30	0.03%	Adverb
40	some	20	0.02%	Quantifier
41	each	15	0.015%	Quantifier
42	both	10	0.01%	Quantifier
43	few	5	0.005%	Quantifier
44	more	3	0.003%	Adverb
45	less	2	0.002%	Adverb
46	most	1	0.001%	Quantifier
47	only	1	0.001%	Adverb
48	too	1	0.001%	Adverb
49	very	1	0.001%	Adverb
50	just	1	0.001%	Adverb