

# Concordancing: An ugly word, but the key to beautiful sentences!

Here is a link to a web-based concordancer:

[http://www.lex tutor.ca/concordancers/concord\\_e.html](http://www.lex tutor.ca/concordancers/concord_e.html)

It has several corpora\* (text databases) to select for your searches. Choose wisely. The *BNC Written Corpus* is a good sample of written British English to begin with.

Here is a ‘screen shot’ of the Concordancer, with some instructions:

The screenshot shows the 'Corpus Concordance English (v.6.5)' web interface. It features a search form with the following elements:

- Keyword(s):** A text input field containing 'equals'.
- Choose a Corpus:** A dropdown menu.
- Sort By:** A dropdown menu set to 'word (s) | Left'.
- Line Width (chars.):** A dropdown menu set to '100'.
- Number of Lines:** A dropdown menu set to '1,000'.
- Gapped?:** A dropdown menu set to 'No'.
- OPTION:** A section with 'With associated word' and 'within 4 words to Left side'.
- Scan for any recurring word (potential colloc.) within 5 words occurring <= 4 times**
- Get concordance:** A yellow button.
- Link extractor:** A section with buttons for 'Show Full URL', 'Test URL', 'Select URL', and 'See URL as Link'.

Four numbered instructions are overlaid on the screenshot:

1. Type your word
2. Select a corpus\*
3. Choose the sort direction
4. Click “Get concordances”

At the bottom of the screenshot, it says: 'Concordance concept by Chris Greaves, HK Polytechnic; V.6.6 code in PERL by Tom Cobb, UQAM Montreal'.

These simple steps will get you started, but you will need to incorporate the use of a dictionary, and possibly a thesaurus, before you can confidently use new words in your writing.

English has one of the most complex preposition systems in the world, making their correct selection a difficult task. Concordancing is ideal for this kind of search — try searching for words that require prepositions alongside them and see what you can discover from the results you get.

Try more than one corpus if your searches are not successful. A search engine such as Google will tell you if a word or phrase is really used – if the only ‘hits’ you get are dictionaries, the word is not really used by English speakers, it only appears in dictionaries. Do not use such words; they are obscure. A Concordancer can then show you how to use some of the ‘real’ words you find.

\* What is a corpus? What are corpora? See page 2 for explanations of the terms used here.

**DO NOT USE THE STUDENT WRITING CORPUS — IT HAS ERRORS IN IT!**

Questions and comments about the use of the Concordancer are welcome:

**Ray Adams:**

E-mail: [rayadams@internode.on.net](mailto:rayadams@internode.on.net)

Web: <http://users.on.net/~rayadams/>  
(find ‘Student Links’ for more language resources)

# An introduction to corpus linguistics for writing

**Terms:** *corpus, corpora, collocates, collocation, concordance, and concordancer.*

- A **corpus** is a collection of texts. These texts can be written (i.e., books) or spoken (i.e. transcribed speech).
- Because **corpora** (plural of corpus) contain large amounts of language, they can be examined for patterns of language use. The patterns of language use most commonly examined are related to **grammar** and **vocabulary**.
- When we look at **grammar** in a corpus, we can notice how structures are put together to make sentences.
- When we look at **vocabulary** in a corpus, we can see that certain words sometimes co-occur with other words. Words that co-occur frequently are called **collocations**.
- We can use a tool called a **concordancer** to sort words into a way that is easy to read.

## Here is an example:

terms, they can expect a much more **ferocious** assault by air in Baghdad and footwork, timeless balances and **ferocious** attack are gradually gaining like a surprised frog, but its **ferocious** blasts of cool air meant that is regarded as the world's most **ferocious** breed. The seven-strong side hard but rarely troubled Wigan's **ferocious** defence. Only Leeds' Alan Sabin iron will contributed to a **ferocious** feud he had with Jonas Salk in the past Mr Clarke was the most **ferocious** opponent of the use of a precision, And Dave Brubeck's **ferocious** recasting of the blues in of shale, beds of it tilted up to **ferocious** slopes, and the lines of its muscle problems will not affect his **ferocious** tackling against the Italians nurse, with Jo storming out in a **ferocious** temper and me shouting invented more rapids and avoided the most **ferocious** waves and water I've ever met. they really need them: to cope with **ferocious** winter driving conditions or

The examples found in concordances can tell us about patterns of word use. For example, we see that the **indefinite article** *a*, **possessive forms**, such as *its*, *Wigan's*, *Brubeck's* and *his*, and **the superlative** *most* sometimes occur before *ferocious* in the concordance above; we also see that words associated with **discord** and **violence** (e.g., *assault*, *attack*, *defence*, *feud*, *opponent*, *tackling*, *temper*) or the **natural world** (e.g., *blasts of cool air*, *breed*, *slopes*, *-waves*, *water*, *winter*) often appear after the word *ferocious* in the examples above<sup>1</sup>.

What does this tell us about the meaning and the usage of *ferocious*?

What does this tell us about how we sometimes think about nature?

Could you use *ferocious* correctly in a sentence after reading this?

---

<sup>1</sup> The description of terms, the concordance and explanation are quoted from:  
Project MORE, CMS Technology In-Service (17-6-03)  
<[http://education.uncc.edu/more/workshop/Introduction\\_to\\_Corpora\\_and\\_Concordancing.pdf](http://education.uncc.edu/more/workshop/Introduction_to_Corpora_and_Concordancing.pdf)>

## What else can we discover from a corpus?

Look at the two lists in the table below. They show something called word frequency in two different corpora – how many times each word appears in each corpus. The lists are not complete, but show the 51 most common words in each. Can you tell what topics the texts in those corpora are related to?

Freq.	Word	Occurrences	Percentage of text	Freq.	Word	Occurrences	Percentage of text
1	the	1453	6.9913 %	1	THE	4515	5.7159 %
2	of	930	4.4748 %	2	and	3470	4.3930 %
3	and	715	3.4403 %	3	TO	2053	2.5991 %
4	in	516	2.4828 %	4	we	1672	2.1167 %
5	to	445	2.1412 %	5	a	1601	2.0268 %
6	a	390	1.8765 %	6	it	1566	1.9825 %
7	is	205	0.9864 %	7	OF	1519	1.9230 %
8	that	169	0.8132 %	8	was	1257	1.5913 %
9	as	164	0.7891 %	9	he	1122	1.4204 %
10	are	151	0.7266 %	10	in	1016	1.2862 %
11	on	150	0.7217 %	11	said	1009	1.2774 %
12	for	143	0.6881 %	12	I	780	0.9875 %
13	by	138	0.6640 %	13	but	760	0.9621 %
14	which	130	0.6255 %	14	not	760	0.9621 %
15	be	128	0.6159 %	15	you	760	0.9621 %
16	with	127	0.6111 %	16	had	729	0.9229 %
17	have	110	0.5293 %	17	that	712	0.9014 %
18	It	104	0.5004 %	18	on	581	0.7355 %
19	This	98	0.4715 %	19	with	572	0.7241 %
20	p	97	0.4667 %	20	Oswald	568	0.7191 %
21	more	96	0.4619 %	21	is	557	0.7052 %
22	their	91	0.4379 %	22	as	520	0.6583 %
23	Australian	90	0.4330 %	23	all	511	0.6469 %
24	or	83	0.3994 %	24	for	492	0.6229 %
25	an	81	0.3897 %	25	were	471	0.5963 %
26	pp	74	0.3561 %	26	so	469	0.5937 %
27	Not	73	0.3512 %	27	his	461	0.5836 %
28	some	72	0.3464 %	28	They	454	0.5748 %
29	from	71	0.3416 %	29	at	445	0.5634 %
30	Australia	65	0.3128 %	30	be	422	0.5342 %
31	they	65	0.3128 %	31	when	417	0.5279 %
32	would	60	0.2887 %	32	up	382	0.4836 %
33	at	58	0.2791 %	33	she	380	0.4811 %
34	Democrats	58	0.2791 %	34	have	365	0.4621 %
35	people	56	0.2695 %	35	then	354	0.4482 %
36	but	55	0.2646 %	36	us	332	0.4203 %
37	power	55	0.2646 %	37	like	323	0.4089 %
38	op	54	0.2598 %	38	did	313	0.3963 %
39	cit	53	0.2550 %	39	if	312	0.3950 %
40	government	52	0.2502 %	40	this	302	0.3823 %
41	policies	50	0.2406 %	41	our	295	0.3735 %
42	one	49	0.2358 %	42	out	292	0.3697 %
43	its	48	0.2310 %	43	about	287	0.3633 %
44	other	48	0.2310 %	44	him	281	0.3557 %
45	was	48	0.2310 %	45	Alice	275	0.3481 %
46	has	47	0.2261 %	46	there	272	0.3443 %
47	J	47	0.2261 %	47	what	272	0.3443 %
48	Party	47	0.2261 %	48	very	260	0.3292 %
49	many	45	0.2165 %	49	do	249	0.3152 %
50	political	45	0.2165 %	50	would	249	0.3152 %
51	were	45	0.2165 %	51	her	243	0.3076 %

This highlights the point that different kinds of words are used to talk and write about different topics, and to talk and write in different ways about those topics. It is therefore important to use a corpus of appropriate language when you are trying to research grammar and vocabulary issues.