SketchEngine Workshop on Wordlists and Concordances

Prof. Fiona Farr, UL

Note 1: If you haven't had a previous introduction to corpus linguistics or SketchEngine, it is advisable that you watch this recorded session before attempting the tasks below: https://media.heanet.ie/page/e83bca1503294580a716dab4c310fc7e

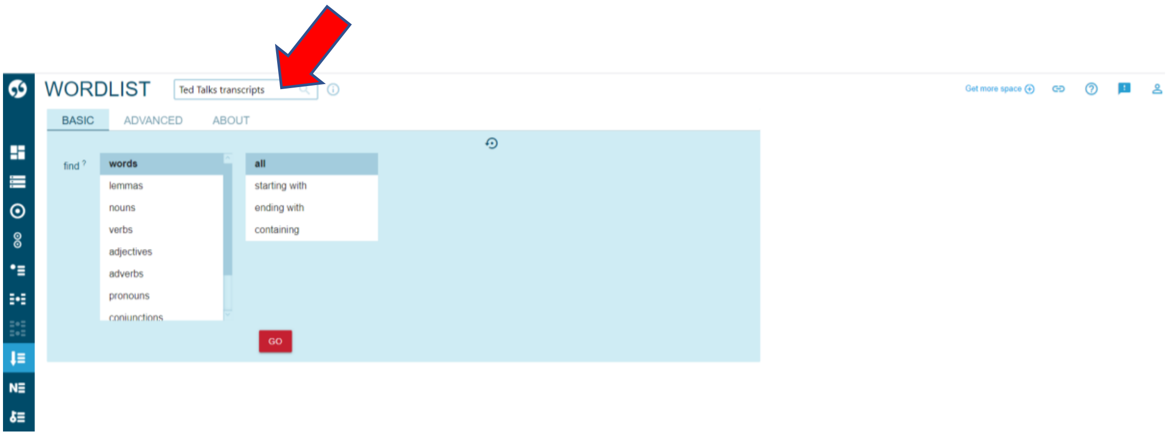
Note 2: In order to complete the tasks below you need to have access to SketchEngine.eu through the institution to which you are affiliated. You can check if your institution has a licence by clicking on the log-in button at SketchEngine.eu, going to institutional log-in, searching for your institution, clicking on it and signing in with your regular institution credentials.

Note 3: The SketchEngine Platform is well supported through integrated video links (click the 'about' tab in the various functions) and through the SketchEngine Channel on YouTube: https://www.youtube.com/channel/UCo2fn2_SNxCikCSAFCBcWBw

Wordlists

Watch this Video: https://www.youtube.com/watch?v=nqpCIICCEdw&feature=emb_logo

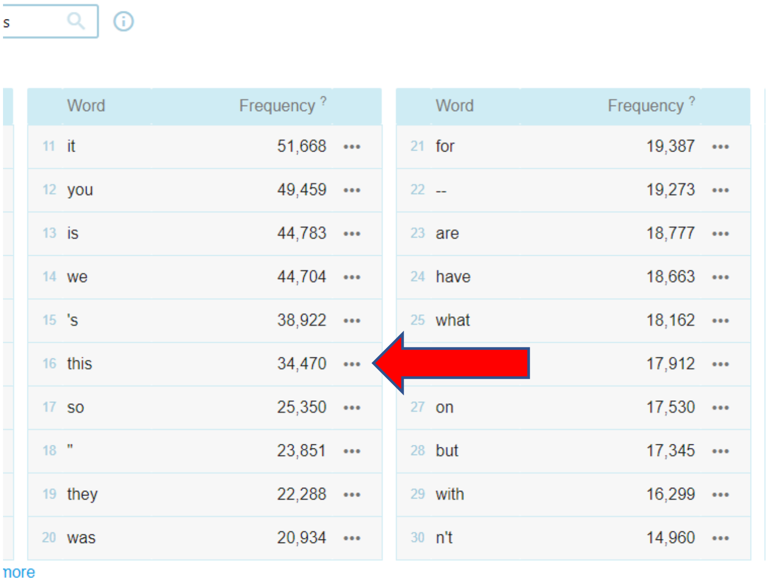
1. Using the TedTalks corpus, run a wordlist for all the **words** (not lemmas – what is the difference?)



What are the top 5 most frequent words?

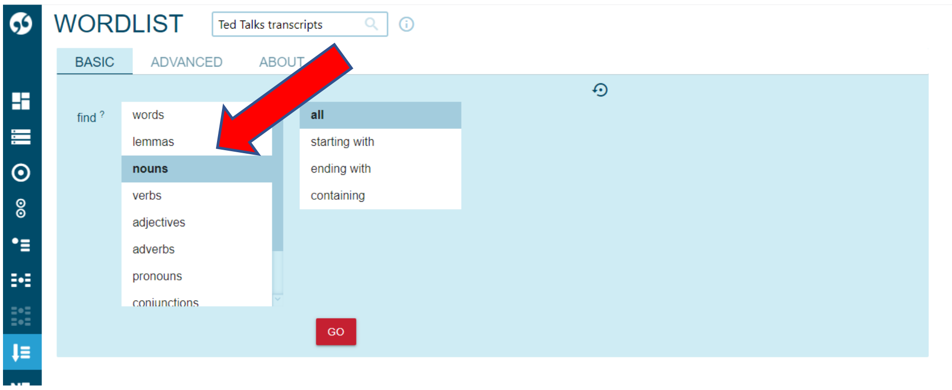
Is there a limit on the number of words returned in a frequency list in SketchEngine? The answer is on the screen.

2. From here, run a concordance for the word ‘this’ by clicking on the three dots next to it on the frequency list.

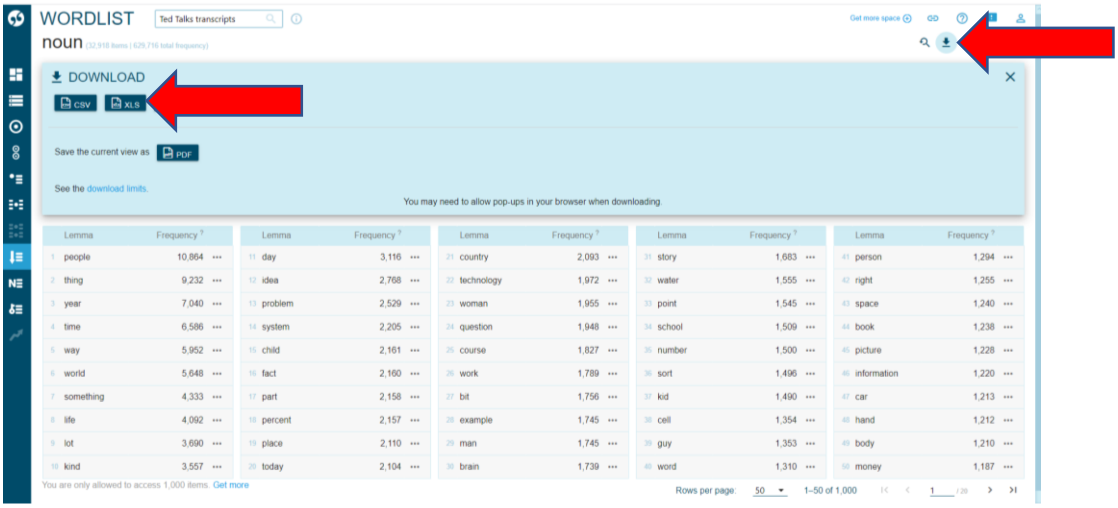


	Word	Frequency ?	
11	it	51,668	•••
12	you	49,459	•••
13	is	44,783	•••
14	we	44,704	•••
15	's	38,922	•••
16	this	34,470	•••
17	so	25,350	•••
18	"	23,851	•••
19	they	22,288	•••
20	was	20,934	•••
21	for	19,387	•••
22	--	19,273	•••
23	are	18,777	•••
24	have	18,663	•••
25	what	18,162	•••
		17,912	•••
27	on	17,530	•••
28	but	17,345	•••
29	with	16,299	•••
30	n't	14,960	•••

3. What are the top 5 nouns in the corpus?



4. Download the noun frequency list to an Excel file and rearrange alphabetically in Excel. What is the first and the last noun on the list?



WORDLIST Ted Talks transcripts

noun (32,918 items | 629,716 total frequency)

DOWNLOAD

CSV XLS

Save the current view as PDF

See the download limits.

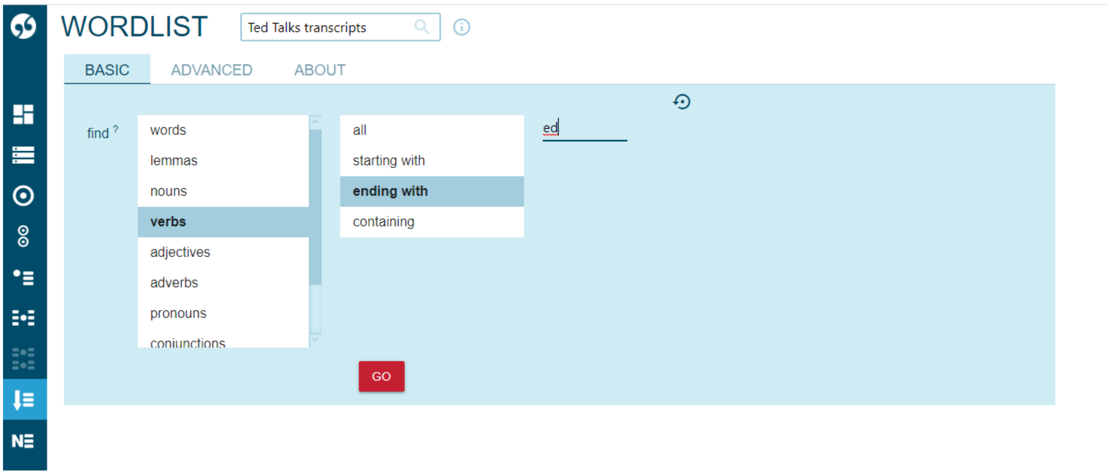
You may need to allow pop-ups in your browser when downloading.

	Lemma	Frequency ?		Lemma	Frequency ?		Lemma	Frequency ?		Lemma	Frequency ?		Lemma	Frequency ?
1	people	10,864	11	day	3,116	21	country	2,093	31	story	1,683	41	person	1,294
2	thing	9,232	12	idea	2,768	22	technology	1,972	32	water	1,555	42	right	1,255
3	year	7,040	13	problem	2,529	23	woman	1,955	33	point	1,545	43	space	1,240
4	time	6,586	14	system	2,205	24	question	1,948	34	school	1,509	44	book	1,238
5	way	5,952	15	child	2,161	25	course	1,827	35	number	1,500	45	picture	1,228
6	world	5,648	16	fact	2,160	26	work	1,789	36	sort	1,496	46	information	1,220
7	something	4,333	17	part	2,158	27	bit	1,756	37	kid	1,490	47	car	1,213
8	life	4,092	18	percent	2,157	28	example	1,745	38	cell	1,354	48	hand	1,212
9	lot	3,690	19	place	2,110	29	man	1,745	39	guy	1,353	49	body	1,210
10	kind	3,557	20	today	2,104	30	brain	1,739	40	word	1,310	50	money	1,187

You are only allowed to access 1,000 items. Get more

Rows per page: 50 1–50 of 1,000 1 / 20

5. Try to get the top 5 regular past tense verbs from the wordlist function. Try this:



WORDLIST Ted Talks transcripts

BASIC ADVANCED ABOUT

find ?: words, lemmas, nouns, **verbs**, adjectives, adverbs, pronouns, conjunctions

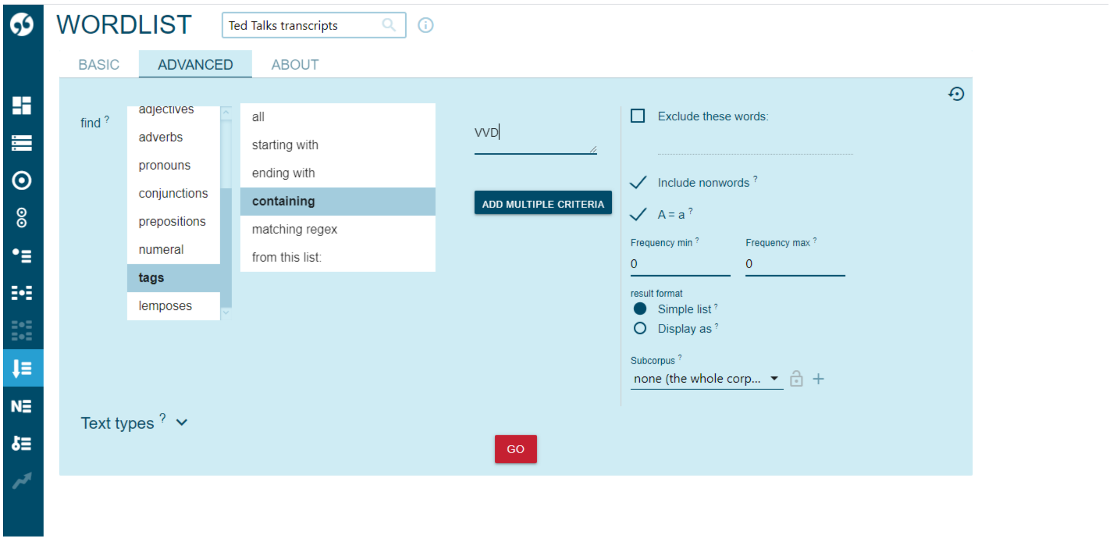
all, starting with, **ending with**, containing

ed

GO

What interference do you find in the results?

Try this (full list of tags is here: https://www.sketchengine.eu/penn-treebank-tagset/):



WORDLIST Ted Talks transcripts

BASIC ADVANCED ABOUT

find ?: adjectives, adverbs, pronouns, conjunctions, prepositions, numeral, **tags**, lemposes

all, starting with, ending with, **containing**, matching regex, from this list:

VVD

ADD MULTIPLE CRITERIA

☐ Exclude these words:

✓ Include nonwords ?

✓ A = a ?

Frequency min ? 0

Frequency max ? 0

result format

● Simple list ?

○ Display as ?

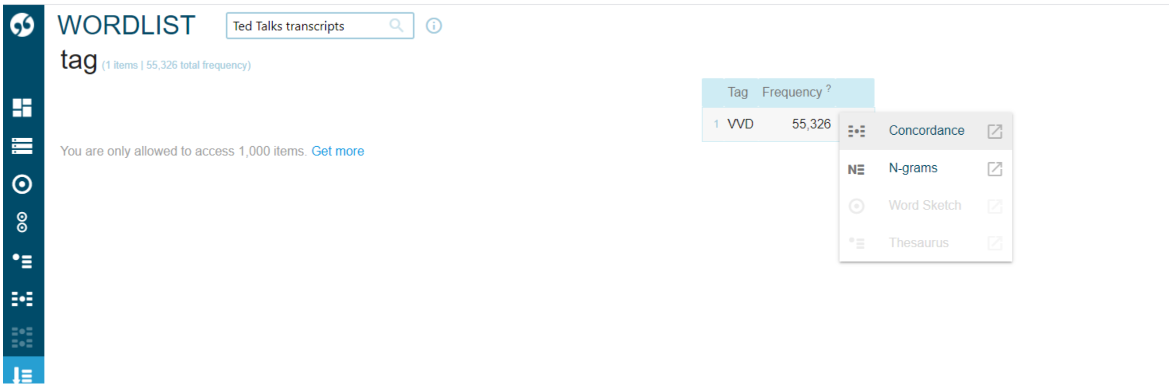
Subcorpus ? none (the whole corp...

Text types ?

GO

What did you get?

After you press ‘go’ in the previous, run a concordance from the next screen (three dots next to the frequency number):



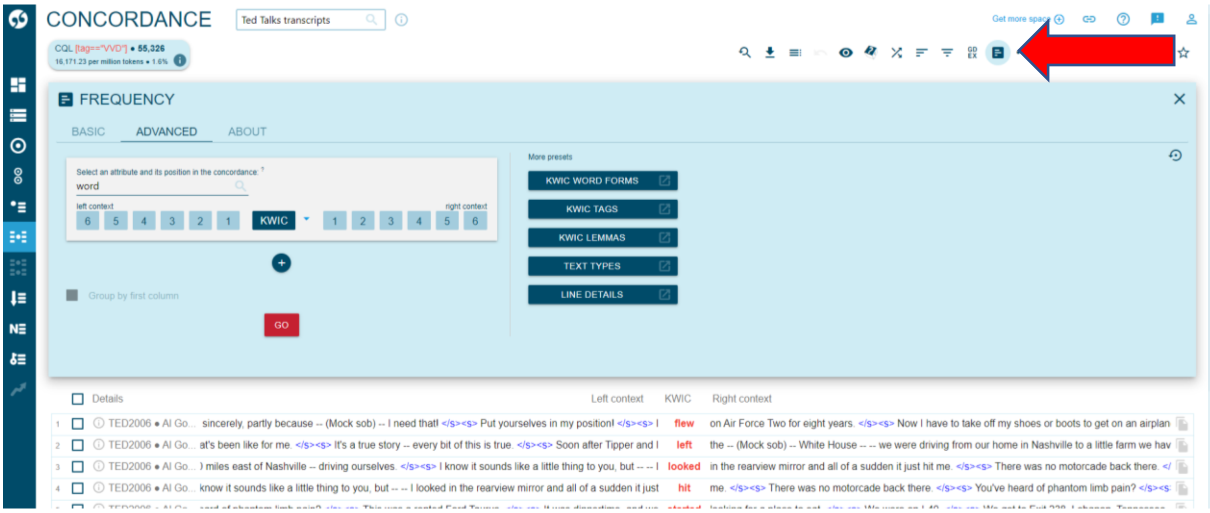
WORDLIST Ted Talks transcripts

tag (1 items | 55,326 total frequency)

You are only allowed to access 1,000 items. Get more

	Tag	Frequency ?
1	VVD	55,326

Do you get what you want? How do you find the most frequent verb from here? Try this:



CONCORDANCE Ted Talks transcripts

CQL [tag=="VVD"] • 55,326
16,171.23 per million tokens • 1.6%

FREQUENCY

BASIC ADVANCED ABOUT

Select an attribute and its position in the concordance: ?
word

left context 6 5 4 3 2 1 KWIC 1 2 3 4 5 6 right context

Group by first column

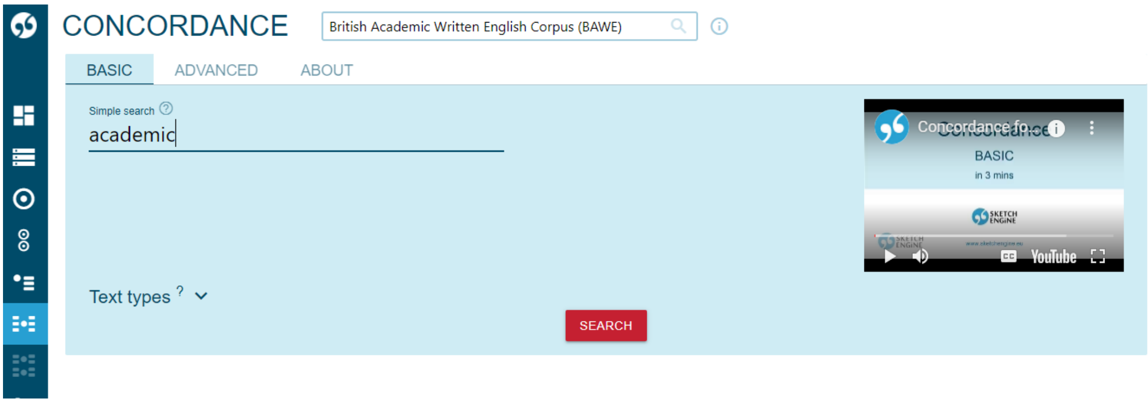
	Details	Left context	KWIC	Right context
1	TED2006 • Al Go...	sincerely, partly because -- (Mock sob) -- I need that! </s><s> Put yourselves in my position! </s><s> I	flew	on Air Force Two for eight years. </s><s> Now I have to take off my shoes or boots to get on an airplan
2	TED2006 • Al Go...	at's been like for me. </s><s> It's a true story -- every bit of this is true. </s><s> Soon after Tipper and I	left	the -- (Mock sob) -- White House -- -- we were driving from our home in Nashville to a little farm we hav
3	TED2006 • Al Go...	) miles east of Nashville -- driving ourselves. </s><s> I know it sounds like a little thing to you, but -- -- I	looked	in the rearview mirror and all of a sudden it just hit me. </s><s> There was no motorcade back there. </
4	TED2006 • Al Go...	know it sounds like a little thing to you, but -- -- I looked in the rearview mirror and all of a sudden it just	hit	me. </s><s> There was no motorcade back there. </s><s> You've heard of phantom limb pain? </s><s
5	TED2006 • Al Go...	eard of phantom limb pain? </s><s> This was a rented Ford Taurus. </s><s> It was dinnertime, and we	started	looking for a place to eat. </s><s> We were on I-40. </s><s> We got to Exit 238, Lebanon, Tennessee

Did you find exactly what you were looking for? If not, what can you now do manually to get exactly what you are looking for?

Concordances

Watch this Video: https://www.youtube.com/watch?v=FzI6tbO5EvQ&feature=emb_logo

1. Using the BAWE corpus, search for the work ‘academic’. How many hits did you get?

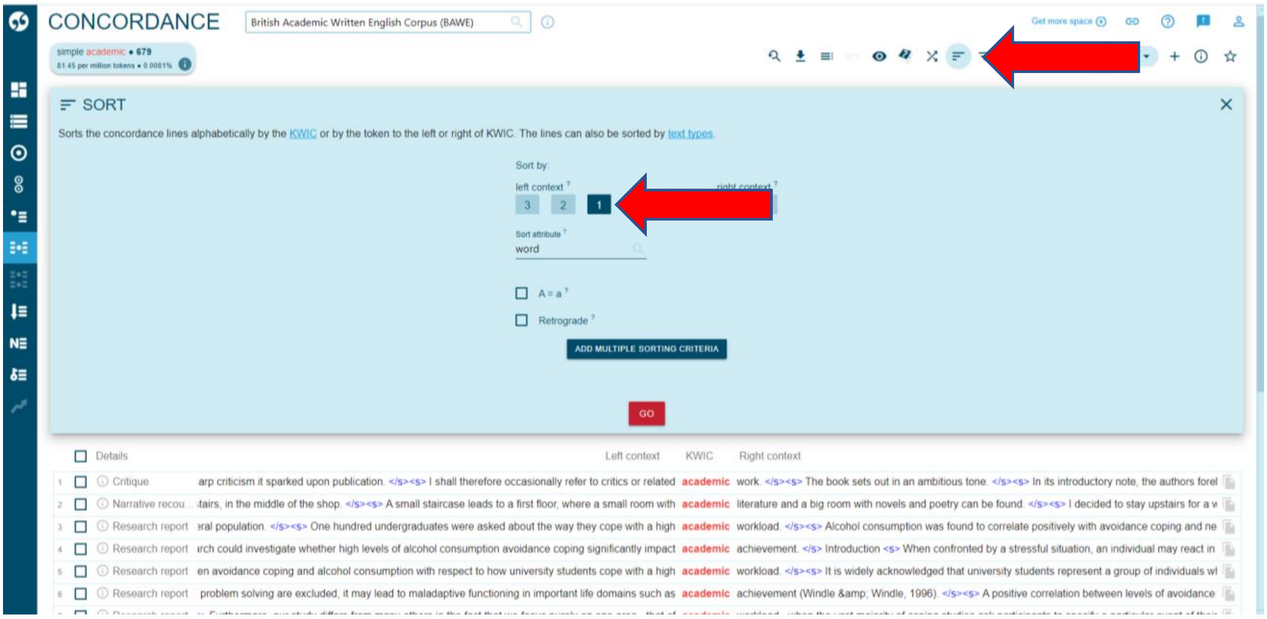


CONCORDANCE

British Academic Written English Corpus (BAWE)

Simple search: academic

2. Which word occurs most frequently one place to the right of ‘academic’? Try this:



CONCORDANCE

British Academic Written English Corpus (BAWE)

simple academic • 679
81.45 per million tokens • 0.0081%

SORT

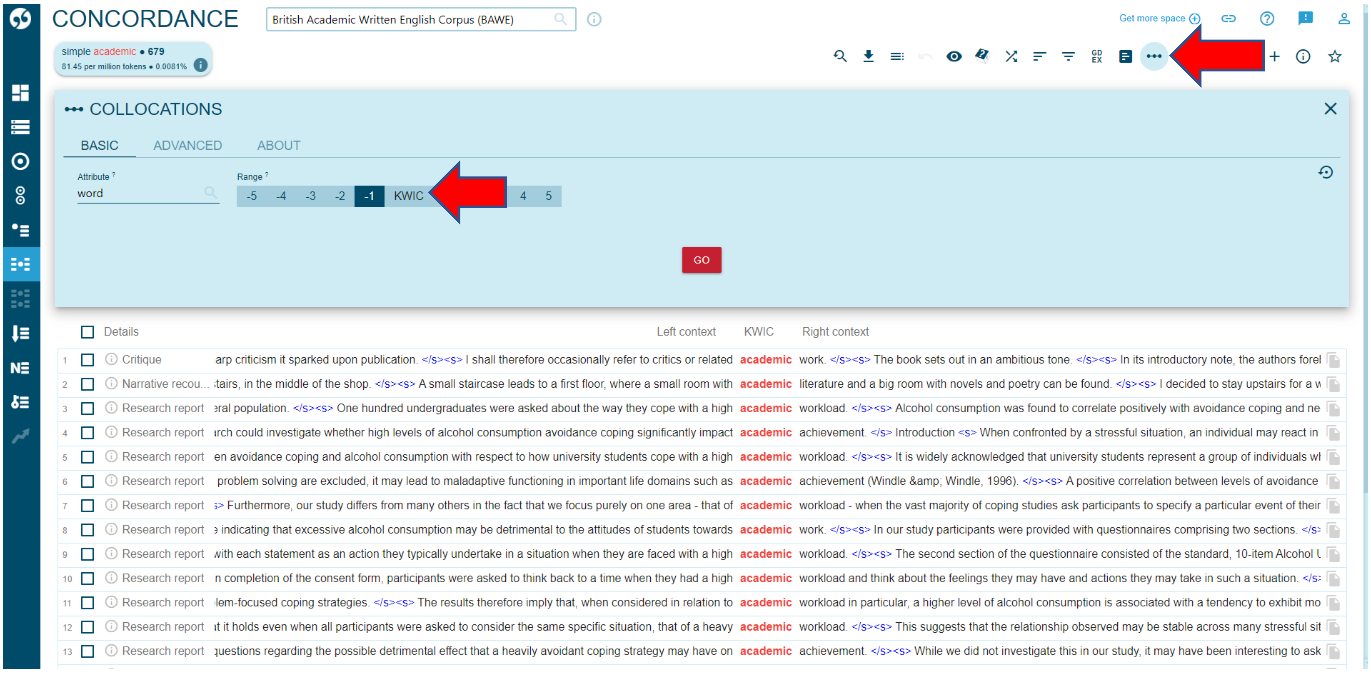
Sorts the concordance lines alphabetically by the KWIC or by the token to the left or right of KWIC. The lines can also be sorted by text types.

Sort by:
left context: 3 2 1
right context
Sort attribute: word
□ A = a
□ Retrograde

	□ Details	Left context	KWIC	Right context
1	□ Critique	arp criticism it sparked upon publication. </s><s> I shall therefore occasionally refer to critics or related	academic	work. </s><s> The book sets out in an ambitious tone. </s><s> In its introductory note, the authors forel
2	□ Narrative recou...	tairs, in the middle of the shop. </s><s> A small staircase leads to a first floor, where a small room with	academic	literature and a big room with novels and poetry can be found. </s><s> I decided to stay upstairs for a w
3	□ Research report	ɔral population. </s><s> One hundred undergraduates were asked about the way they cope with a high	academic	workload. </s><s> Alcohol consumption was found to correlate positively with avoidance coping and ne
4	□ Research report	ırch could investigate whether high levels of alcohol consumption avoidance coping significantly impact	academic	achievement. </s> Introduction <s> When confronted by a stressful situation, an individual may react in
5	□ Research report	en avoidance coping and alcohol consumption with respect to how university students cope with a high	academic	workload. </s><s> It is widely acknowledged that university students represent a group of individuals wl
6	□ Research report	problem solving are excluded, it may lead to maladaptive functioning in important life domains such as	academic	achievement (Windle &amp; Windle, 1996). </s><s> A positive correlation between levels of avoidance

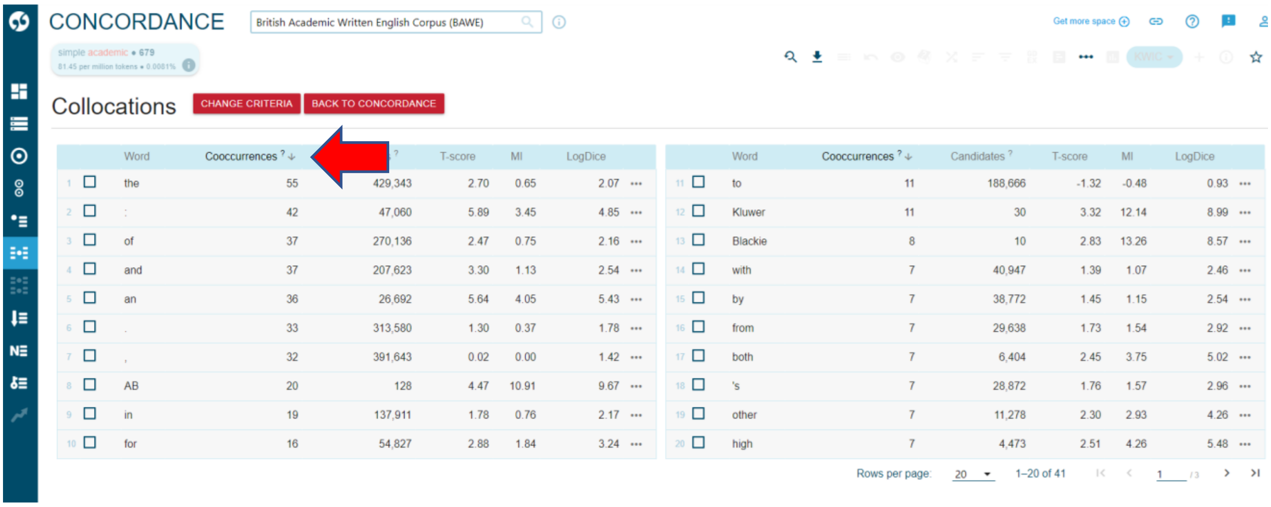
What results did you get? Can you find which is most frequent (manually) from here?

3. Here’s another way to do it more automatically using the collocation function:



		Details	Left context	KWIC	Right context
1	□	Critique	arp criticism it sparked upon publication. </s><s> I shall therefore occasionally refer to critics or related	academic	work. </s><s> The book sets out in an ambitious tone. </s><s> In its introductory note, the authors forel
2	□	Narrative recou...	tairs, in the middle of the shop. </s><s> A small staircase leads to a first floor, where a small room with	academic	literature and a big room with novels and poetry can be found. </s><s> I decided to stay upstairs for a w
3	□	Research report	ral population. </s><s> One hundred undergraduates were asked about the way they cope with a high	academic	workload. </s><s> Alcohol consumption was found to correlate positively with avoidance coping and ne
4	□	Research report	rch could investigate whether high levels of alcohol consumption avoidance coping significantly impact	academic	achievement. </s> Introduction <s> When confronted by a stressful situation, an individual may react in
5	□	Research report	en avoidance coping and alcohol consumption with respect to how university students cope with a high	academic	workload. </s><s> It is widely acknowledged that university students represent a group of individuals wl
6	□	Research report	problem solving are excluded, it may lead to maladaptive functioning in important life domains such as	academic	achievement (Windle &amp; Windle, 1996). </s><s> A positive correlation between levels of avoidance
7	□	Research report	> Furthermore, our study differs from many others in the fact that we focus purely on one area - that of	academic	workload - when the vast majority of coping studies ask participants to specify a particular event of their
8	□	Research report	indicating that excessive alcohol consumption may be detrimental to the attitudes of students towards	academic	work. </s><s> In our study participants were provided with questionnaires comprising two sections. </s
9	□	Research report	vith each statement as an action they typically undertake in a situation when they are faced with a high	academic	workload. </s><s> The second section of the questionnaire consisted of the standard, 10-item Alcohol U
10	□	Research report	n completion of the consent form, participants were asked to think back to a time when they had a high	academic	workload and think about the feelings they may have and actions they may take in such a situation. </s
11	□	Research report	lem-focused coping strategies. </s><s> The results therefore imply that, when considered in relation to	academic	workload in particular, a higher level of alcohol consumption is associated with a tendency to exhibit mo
12	□	Research report	t it holds even when all participants were asked to consider the same specific situation, that of a heavy	academic	workload. </s><s> This suggests that the relationship observed may be stable across many stressful sit
13	□	Research report	uestions regarding the possible detrimental effect that a heavily avoidant coping strategy may have on	academic	achievement. </s><s> While we did not investigate this in our study, it may have been interesting to ask

What do you get? The results appear alphabetically by default. To rearrange the words by frequency, click on the column heading that says ‘cooccurences’:



Collocations

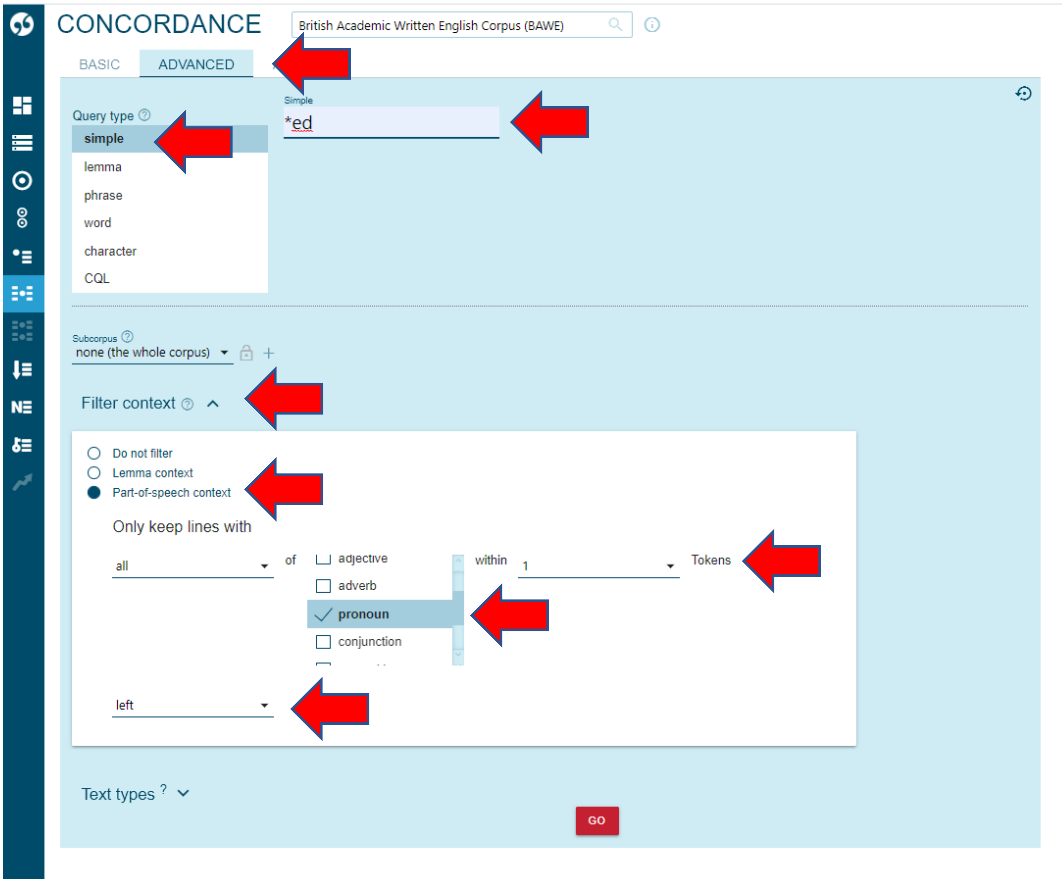
		Word	Cooccurrences ? ↓	?	T-score	MI	LogDice	
1	□	the	55	429,343	2.70	0.65	2.07	•••
2	□	:	42	47,060	5.89	3.45	4.85	•••
3	□	of	37	270,136	2.47	0.75	2.16	•••
4	□	and	37	207,623	3.30	1.13	2.54	•••
5	□	an	36	26,692	5.64	4.05	5.43	•••
6	□	.	33	313,580	1.30	0.37	1.78	•••
7	□	,	32	391,643	0.02	0.00	1.42	•••
8	□	AB	20	128	4.47	10.91	9.67	•••
9	□	in	19	137,911	1.78	0.76	2.17	•••
10	□	for	16	54,827	2.88	1.84	3.24	•••

		Word	Cooccurrences ? ↓	Candidates ?	T-score	MI	LogDice	
11	□	to	11	188,666	-1.32	-0.48	0.93	•••
12	□	Kluwer	11	30	3.32	12.14	8.99	•••
13	□	Blackie	8	10	2.83	13.26	8.57	•••
14	□	with	7	40,947	1.39	1.07	2.46	•••
15	□	by	7	38,772	1.45	1.15	2.54	•••
16	□	from	7	29,638	1.73	1.54	2.92	•••
17	□	both	7	6,404	2.45	3.75	5.02	•••
18	□	's	7	28,872	1.76	1.57	2.96	•••
19	□	other	7	11,278	2.30	2.93	4.26	•••
20	□	high	7	4,473	2.51	4.26	5.48	•••

Rows per page: 20 1–20 of 41 1 / 3

4. Try to find the top 5 regular past tense verbs through the concordance function, using a basic search for *ed. What are the issues with the results?

5. To resolve some of these issues, let’s rerun the concordance with some more specific criteria (filters) from the advanced tab. Try this to tell SE to include only examples of *ed which has pronouns one place to the left of it:



CONCORDANCE British Academic Written English Corpus (BAWE)

Query type: simple

Simple: *ed

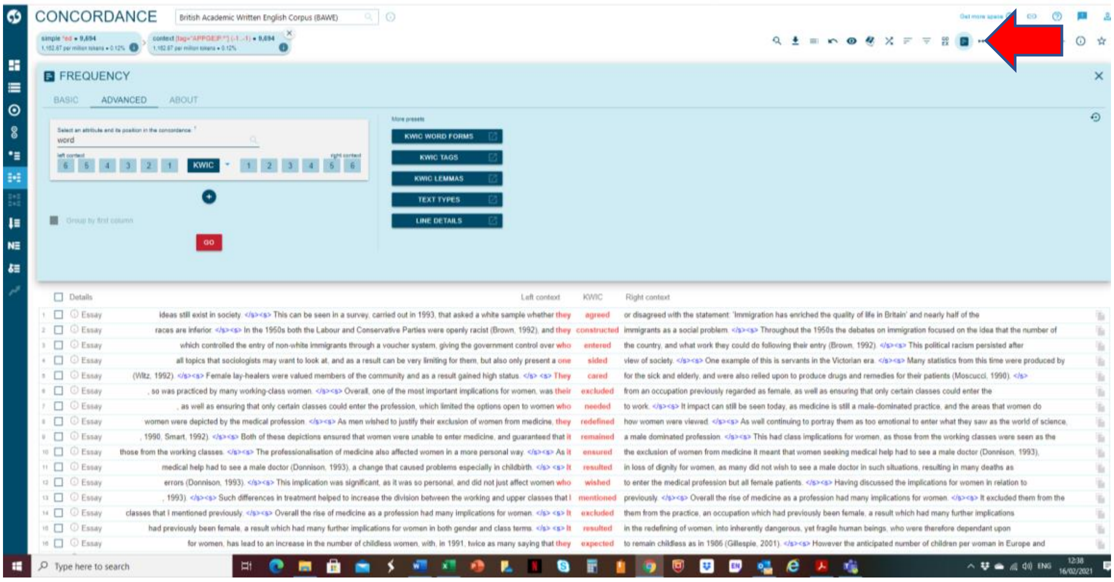
Subcorpus: none (the whole corpus)

Filter context

○ Do not filter
○ Lemma context
● Part-of-speech context

Only keep lines with all of ☑ pronoun within 1 Tokens left

6. The results you get from 5 above, will appear randomly. You can rearrange them alphabetically by clicking on the arrow next to the KWIC column heading, and to the left and right by using the sort function as above. To get a list of which of the past tense verbs is most frequent, click on the 'frequency' tab and go:



CONCORDANCE British Academic Written English Corpus (BAWE)

simple *ed • 9,694 1,152.87 per million tokens • 0.12%

context [tag="APPGE|P.*"] (-1...-1) • 9,694 1,152.87 per million tokens • 0.12%

FREQUENCY

Select an attribute and its position in the concordance: word

left context 6 5 4 3 2 1 KWIC 1 2 3 4 5 6 right context

	Details	Left context	KWIC	Right context
1	Essay	ideas still exist in society. </s><s> This can be seen in a survey, carried out in 1993, that asked a white sample whether they	agreed	or disagreed with the statement: 'Immigration has enriched the quality of life in Britain' and nearly half of the
2	Essay	races are inferior. </s><s> In the 1950s both the Labour and Conservative Parties were openly racist (Brown, 1992), and they	constructed	immigrants as a social problem. </s><s> Throughout the 1950s the debates on immigration focused on the idea that the number of
3	Essay	which controlled the entry of non-white immigrants through a voucher system, giving the government control over who	entered	the country, and what work they could do following their entry (Brown, 1992). </s><s> This political racism persisted after
4	Essay	all topics that sociologists may want to look at, and as a result can be very limiting for them, but also only present a one	sided	view of society. </s><s> One example of this is servants in the Victorian era. </s><s> Many statistics from this time were produced by
5	Essay	(Witz, 1992). </s><s> Female lay-healers were valued members of the community and as a result gained high status. </s> <s> They	cared	for the sick and elderly, and were also relied upon to produce drugs and remedies for their patients (Moscucci, 1990). </s>
6	Essay	, so was practiced by many working-class women. </s><s> Overall, one of the most important implications for women, was their	excluded	from an occupation previously regarded as female, as well as ensuring that only certain classes could enter the
7	Essay	, as well as ensuring that only certain classes could enter the profession, which limited the options open to women who	needed	to work. </s><s> It impact can still be seen today, as medicine is still a male-dominated practice, and the areas that women do
8	Essay	women were depicted by the medical profession. </s><s> As men wished to justify their exclusion of women from medicine, they	redefined	how women were viewed. </s><s> As well continuing to portray them as too emotional to enter what they saw as the world of science,
9	Essay	, 1990, Smart, 1992). </s><s> Both of these depictions ensured that women were unable to enter medicine, and guaranteed that it	remained	a male dominated profession. </s><s> This had class implications for women, as those from the working classes were seen as the
10	Essay	those from the working classes. </s><s> The professionalisation of medicine also affected women in a more personal way. </s><s> As it	ensured	the exclusion of women from medicine it meant that women seeking medical help had to see a male doctor (Donnison, 1993),
11	Essay	medical help had to see a male doctor (Donnison, 1993), a change that caused problems especially in childbirth. </s> <s> It	resulted	in loss of dignity for women, as many did not wish to see a male doctor in such situations, resulting in many deaths as
12	Essay	errors (Donnison, 1993). </s><s> This implication was significant, as it was so personal, and did not just affect women who	wished	to enter the medical profession but all female patients. </s><s> Having discussed the implications for women in relation to
13	Essay	, 1993). </s><s> Such differences in treatment helped to increase the division between the working and upper classes that I	mentioned	previously. </s><s> Overall the rise of medicine as a profession had many implications for women. </s><s> It excluded them from the
14	Essay	classes that I mentioned previously. </s><s> Overall the rise of medicine as a profession had many implications for women. </s> <s> It	excluded	them from the practice, an occupation which had previously been female, a result which had many further implications
15	Essay	had previously been female, a result which had many further implications for women in both gender and class terms. </s> <s> It	resulted	in the redefining of women, into inherently dangerous, yet fragile human beings, who were therefore dependant upon
16	Essay	for women, has lead to an increase in the number of childless women, with, in 1991, twice as many saying that they	expected	to remain childless as in 1986 (Gillespie, 2001). </s><s> However the anticipated number of children per woman in Europe and

Does this give you what you need? Is there any interference in the results that you get?