

Discovering English with the Sketch Engine

Learning language from language

Getting started: comparing our intuition with data



1. Can the word *whose* be used with things as well as people? Permalink¹ ske.li/bnc_whose_500

Yes No

2. What is the difference between *blond* and *blonde*? ske.li/bnc_blond_pos, ske.li/bnc_blonde_pos_500

- One of them is a noun, the other is an adjective
- One of them refers to males, the other to females
- One of them is for hair, the other is for describing people
- Only one of them is used in proper English
- There is no difference. You can use either of them in any situation

3. Is the word *data* singular or plural?

Singular: The data is ...

Plural: The data are ...

Both - it depends on the regional variety of English. e.g., US, UK

Both - it depends on the context, e.g., popular use, scientific use

ske.li/bnc_sci_data_500, ske.li/bnc_datum_3ps²

4. Which of these verbs have *research* as their object? ske.li/bnc_ws_research_n

- _____ conduct research
- _____ make research
- _____ design research
- _____ do research
- _____ create research
- _____ cause research
- _____ publish research
- _____ determine research
- _____ support research

Do these verb collocates of *research* have very different subjects?

In other words, are they performed by different types of people or things?

5. Can you find any "issues" in this sentence? Try the Hoey Procedure as illustrated in the ppt.

* *Chapter 6 is highly advisable for anxious parents, for it clarifies the way how do children learn language.*

¹ The permalinks on this page open the Sketch Engine pages with the relevant data:

² The CQL search that yields these findings: [lemma = "datum"][tag = "V.Z"]

Registering for the Sketch Engine

A Sketch Engine account allows you to customise various aspects of the site, to create your own corpora, and to save Permalinks (as seen in footnotes here). Registered users can share their own corpora with each other. To use the full gamut of the Sketch Engine's resources, it is necessary to register.

Non-MU users get a 30 day trial account. To continue using the full gamut, there is a small annual fee. You receive an email as the 30 day trial is coming to an end.

Register here

Public: <https://the.sketchengine.co.uk/register/>

MU: <https://ske.fi.muni.cz/register>

Note:

- *User Name* must be unique in The Sketch Engine system.
- *Email*: make sure you can open this email account immediately so that you can retrieve your password.
- *Organisation*: where you work – university, type of school, etc.
- *Antibot*: More commonly known as Captcha (see Wikipedia), it ensures that humans are registering, not (ro)bots.

Like most webpages created in the 21st Century, the help text provided is helpful. Don't hesitate to read it.

Passwords

After you click on Register, read the message. Read the message. Your confirmation email contains a randomly generated password. Change it to something memorable. It goes without saying that when you change your password to something memorable, it should be memorable. In the worst case, you can click on Lost Password and get another randomly generated one that you will need to change to something even more memorable.

Open Corpora

It is possible to use a small set of corpora without registering or paying – see Open Corpora.

- <https://the.sketchengine.co.uk/open/>

Some language questions that can be asked of the Sketch Engine

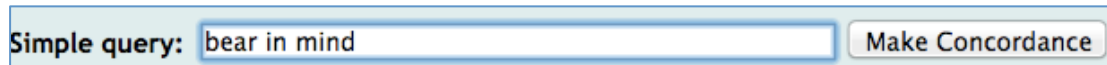
The following questions guide use through most of the main modes of querying corpora.

Simple Query field

The words here are treated as lemmas, i.e. searches include conjugated and declined forms.

Question 1 What precedes and follows *bear in mind*?

Firstly, perform the search as shown. Click Node Forms in the left panel (pictured) to see how the words in the phrase might conjugate and decline.



Click on Sort left and right to see what patterns emerge either side of it.

Question 2 Is *handsome woman* bad English?

Lemma field

This field searches for all forms of a word, e.g. *drive, drives, driving, drove, driven*.

Question 3 Is *thatcher* a common noun just as *Thatcher* is a proper noun?

Given that most of the texts in the BNC are from the late 1980s, it is not surprising that the Iron Lady is well-represented.

Question 4 If you are not familiar with the common noun, *thatcher*, can you infer its meaning from its contexts?

Can you deduce the verb from the noun *thatcher*? Search for *thatch** and right sort and you'll be well on the way to answering the questions.

Inferring meanings of unknown words in a text is much discussed in ELT literature. It is worth comparing what can be discovered about a word which occurs once in a text with multiple examples in short extracts from thousands of native speakers.

Phrase field

This field searches for a string of words in the exact form that you enter.

Question 5 How does Margaret Thatcher's nickname appear in the BNC?

Is it used as frequently? Do *iron lady* and *Iron Lady* return the same results?

Before you look, try to predict the answers. Check it in the Phrase field, then using Sentence instead of KWIC view (in the left panel), compare what is said about the Lady with what is said about the nickname itself.

So, be aware of the differences that capitalisation makes when creating searches in the Simple Query field. Try another example: General Grant vs. general grant.

Question 6 Is the 'way how to do something', a structure in English?

Before you search for *way how* give some thought to the frequencies of these two words separately. Perform individual searches for *way* and *how*. Write their frequencies here.

BNC	way:	how:
-----	------	------

Test your intuition: would you predict *that allay fears, infinite wisdom, worst year, rich colour, stripped naked* and *unconditional surrender* occur more or less frequently than *way how*?

Type *way how* into the Phrase Field and see how often it occurs. Then see how often it is followed by a to-infinitive. Does this look like a pattern of normal usage?

It is worth noting that other 'way-like' words do not occur in front of *how* either. The BNC returns no *strategy, system, approach, method, plan, technique + how*. The nouns that do precede *how* can be seen here: ske.li/bnc_1_n_how.

PoS filter
 Window: left 1 tokens.
 PoS: determiner noun noun singular noun plural all

Word form field

This field searches for a specific word form. For example, searching for *go*, without the rest of the lemma. Or searching for a specific word form, e.g. *going* (as a noun if you select this POS).

Question 7 Are both *though* and *although* used to start sentences? Equally?

Perhaps you already know the answer to this through intuition, or through being taught it. **Match Case** is useful for identifying words at the start of a sentence.

Word Form: Though|Although PoS: unspecified match case

You don't have to process tens of thousands of lines. Rather, click on **Node Forms** in the left panel.

Character field

Searches for a single character. In English, this is particularly useful for punctuation.

Question 8 Are capital letters used after colons and semicolons?

Type these punctuation marks into the character field, perform searches and then use Frequency in the right panel to make a list of at 1R.

Question 9 How much information typically appears in brackets?

Type a left bracket into the Character Field. Make a sizeable Sample and Sort to the right.

What appears in brackets? How many words? Are there any clauses, i.e. containing a finite verb? Give some thought to the reasons for putting such things in brackets. And observe the text types of your findings.

Operators

Question 10 What are some examples of the suffix *-ity* indicating the 'state or quality of being the adjective'?

Try *.*ity*. This gives all the words that end in *-ity*. Those which are the abstractions of adjectives are typically singular nouns. Nevertheless, performing this search in the **Lemma field** returns plural forms as well, e.g. *opportunities, eccentricities*. Are these also abstract forms of adjectives?

Question 11 Prefixes work similarly? Try *mega.** in the Word Form field.

In the BNC, click on Node Forms – the list has *Megan* at the top! And the following occur only once in the BNC:

megafirm, megafabulously, megamoney, megamum, megamouth

You might also try *hyper.*, over.**

Context

Question 12 How do we *elaborate*?

To find the adverbs that describe how we *elaborate*, use a Context POS Filter: adverb 1L, then Left sort. It is unfortunate that English grammar does not distinguish lexical and functional adverbs more overtly.

Question 13 Compare the structures of *photograph with take a photo(graph)*, *guess with make a guess*, *smile with give a smile*, etc.

These are called delexical verb structures. Put the verb in Simple Query and the noun in the Lemma Filter.

Text Types

Question 14 Are *moreover* and *whereas* used in speech, or do they belong to the written language?

After searching for the words, click on Text types in the left panel.

Is this the right question? Perhaps it is more to do with the level of formality, regardless of the medium. What can you discover?

Collocation

Question 15 What are the collocates of the phrasal verb, *blow up*?

Type this into Simple Query field, and then click on collocation at the bottom of the left panel. Accept the default settings. What do the collocates indicate?

Question 16 It would be reasonable to expect the words that follow the adverb *positively* to be positive, would it not?

Create lists of collocates of this word and investigate.

Question 17 Does *flying colours* refer to flags?

Search for *fly colour* in Simple Query. Make a list of left collocates (Range -4 to 0) and see that the answer is a resounding *no*! What phrases does *flying colours* occur in? Can you therefore infer its meaning?

Thesaurus

Question 18 What are the relationships between the search word and the words in the Thesaurus list when you search for a specific colour, job or animal, for example?

Click on Thesaurus in the left panel. Enter your search word, e.g. *blue*, and choose the part of speech. Click the button. Synonyms of *blue* are not listed, rather the list contains words which occur in the same collocation and grammatical contexts.

Question 19 What words are listed when you search for the hypernyms, *colour*, *job* and *animal*?

Before performing this search, can you predict the content of the lists?

Word Sketch

Question 20 How many things can you learn about *sympathetic* and about English in general from studying the word in this format?

sympathetic (*adjective*)
British National Corpus (TreeTagger) freq = 1,478 (13.1 per million)

adj. subject	190	16.5	modifier	228	0.6	modifies	736	1.9	and/or	262	1.7	pp. to-i	278	15.5
tone	6	4.31	excessively	3	7.34	detonation	3	6.86	supportive	11	7.92	plight	4	6.38
historian	3	4.11	suitably	3	6.56	listener	8	6.73	tolerant	5	7.52	cause	16	4.8
someone	5	2.65	politically	4	6.36	onlooker	3	6.51	cardiac	3	6.99	ideal	3	4.47
official	3	2.29	broadly	4	6.28	affiliation	5	6.48	nervous	8	6.0	aim	7	4.19
council	3	1.93	genuinely	3	5.8	magic	7	6.46	patient	3	5.75	regime	3	3.44
paper	3	1.48	however	6	5.55	ear	27	6.38	hostile	4	5.59	movement	7	2.75
officer	3	1.47	generally	10	4.92	nerve	11	6.26	friendly	5	4.94	idea	10	2.69
government	5	0.67	largely	6	4.65	biographer	3	6.23	sensitive	4	4.41	view	9	2.49
group	5	0.67	deeply	3	4.57	handling	6	5.76	professional	4	3.5	concern	4	2.36
			especially	3	4.57	hearing	10	5.48	independent	3	3.03	proposal	3	2.33
			highly	6	4.31	adjustment	6	5.44	local	4	1.41	demand	3	1.84
			extremely	4	4.28	correspondent	3	5.21	little	3	0.77	need	4	1.5
			particularly	7	4.24	noise	9	4.99	first	3	0.08	problem	7	1.28
			very	51	4.17	understanding	14	4.87	many	3	0.04	party	3	0.48
			apparently	3	4.13	observer	4	4.67				woman	3	0.02
			fairly	3	4.1	glance	4	4.66						
			obviously	3	3.75	coverage	4	4.48	adj. comp. of	396	16.3			
			clearly	3	2.94	reception	3	4.45	feel	11	2.69	np. adj. comp. of	36	14.0
			quite	5	2.59	response	18	4.43	become	16	2.53	find	5	0.66
			too	6	2.25	smile	8	4.38	seem	5	1.75	make	11	0.39
			really	3	1.31	audience	7	4.31	appear	3	1.69			
			as	4	0.95	manner	7	4.27	be	345	1.49			
			not	15	0.22	imagination	3	4.22	find	7	1.14			
			so	3	0.19	attitude	12	4.19	look	4	0.16			
						supporter	5	4.01						

A **Word Sketch** reveals a wonderful amount of information about a word.

Col 1: who and what is sympathetic

Col 2: how are people sympathetic

Col 3: what things are said to be sympathetic

Col 4: when joined by *and* or *or*, is the other adjective lending support to the notion of *sympathetic* or is it adding something new?

Col 5: someone or something is sympathetic to your

Col 6: we *feel* sympathetic, we *seem* sympathetic, we ...

Col 7: we can *make* someone sympathetic to ...

What groups do the nouns in the Modifies column (col. 3) form?

Look at the list and see which things you would group under the same headings. Even if you can't think of a heading, a semantic notion is likely to form in your mind.

In the process of answering these questions for each column of the *sympathetic* Word Sketch, it is possible that your idea of this word has changed, especially if your first language has a cognate form with a different usage, e.g. *sympatisch*, *simpatico*, *sympathique*, *sympatický*. In any case, answering the questions forms connections between words you already know and new words: it strengthens connections in your brain.

These notes have been extracted from *Discovering English with the Sketch Engine* (Thomas - forthcoming). See http://bit.ly/lulu_deske