

# Introducing Sketch Engine for academic prose

## Getting started discussion questions

1. Is the word <sup>1</sup>*whose* used with things as well as people?

2. Is the word *data* used as a singular or plural noun?

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.<sup>2</sup>

3. *Research*<sup>3</sup> is not the object of all of these verbs. Underline the ones that you think are correct.

\_\_\_\_\_ 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 different **subjects**? In other words, are they performed by different types of people or things?

4. Do sentences begin with *First* or *Firstly* in academic prose?

## Registration

MU students go to [ske.fi.muni.cz](http://ske.fi.muni.cz) and click on SSO authentication.

---

<sup>1</sup> [ske.li/bnc\\_whose\\_500](http://ske.li/bnc_whose_500)

<sup>2</sup> [ske.li/bnc\\_sci\\_data\\_500](http://ske.li/bnc_sci_data_500), [ske.li/bnc\\_datum\\_3ps](http://ske.li/bnc_datum_3ps)

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

<sup>3</sup> [ske.li/bnc\\_ws\\_research\\_n](http://ske.li/bnc_ws_research_n)

## Simple Query field


Use the Simple Query field to search for words and phrases. The search will find words in all their forms, e.g. singular and plural, verbs in 3rd person singular, past forms, etc.

### Question 1. Does the *truth lie*?

What forms are these two words in, and what is typically used before and after?

Is it used in academic prose?

#### Text types

Subcorpus:  [info](#) [create new](#) 

### Question 2. Do we say *people who* or *people that*?

After searching for both, click on Text Types under Frequency in the left panel.

What do you observe?

### Question 3. Which of these phrases are used in academic prose?

Use lemmas (base forms) of words. Asterisks are also useful, e.g.,

to \* surprise

to \* \* surprise.

To make a summary list of findings, click on Node forms in the left panel.

What do you learn about their typical contexts?

*to my great surprise, is defined as, can(not) be compared with/to, bear in mind. it revolves around, it turns out that, to the best of my knowledge*

## Phrases

Use the phrase field to study phrases in their exact wordings.

### Question 4. What follows *in the beginning* and *at the beginning*?

Type in one of these phrases, while your colleagues types in the other. Click Sort Right in the left panel. What do you notice?

### Question 5. What is the difference between *Given that* at the beginning of a sentence, with and without a comma?

The use of capital "G" is important. Right sort the concordance page. Does this have a counterpart in spoken language?

### Question 6. Which of these expressions are used in academic prose?

Try your intuition first. Choose several to examine.

*last but not least, to some extent, the fact that, all of a sudden, as a matter of fact, at the expense of, by the way, in this day and age, in turn, in a row, in the scheme of things, one way or another, in the context of, works in mysterious ways, within reason, well within, at the risk of, provided that*

What are they used to express? Are there any patterns in what precedes and follows them?

## Word form

Use the word form field to search for a specific word form, e.g. goings (noun)

### Word focus: *medium*.

Starting with a Simple Query:

**Node tags** tell us that it has two parts of speech, mostly \_\_\_\_\_ as well as \_\_\_\_\_.

What does **Node forms** tells us about its singular and plural noun forms?

Word form:  PoS:

Search for the noun *medium*, in the word form field.

Click Frequency in the left panel and set it like this.

### Observations

What do you think the next words might be?

Click on P to find out. Then Right sort the concordance page.

What do you find by Left sorting the page?

Click the column headings to sort them. Set the Frequency limit to 5. What do you observe as you sort the left and the right columns?

What does the high frequency of punctuation suggest?

What have you learnt about language and about using Sketch Engine through studying *medium*?

Using similar procedures, what can you find out about the words *factor* and *factors* and/or data and *datum*?

## Lemma field

Use the lemma field to search for a word and all its word forms. Use lower case only. Specify part of speech.

**Question 7. Here are some words that commonly appear in academic prose. What can you learn about how they are used in academic texts?**

After searching for a word, you can make a Sample, sort the concordance page left and right, make frequency charts. Some of the features of words we can ask about include:

Are adjectives used attributively, i.e. on the left of the noun e.g. *secure income*, and/or predicatively, i.e. after a verb usually, i.e. after a verb usually, *be*, e.g. *the results are conclusive*.

What prefixes make these words negative? Do the negative adjectives collocate with the same nouns?

Noun	Verb	Adjective	Adverb
problem	support	challenging	potentially
situation	cause	transitional	significantly
attempt	exploit	logical	specifically
strategy	minimise	symbolic	intensely

Lemma:

PoS:

## Character field

The character field can be used to observe punctuation.

**Question 8. How much information occurs between brackets?**

Search for a left bracket using the character field. What types of things occur between the ( and the )?

**Question 9. Are colons followed by capital letters?**

Search for the colon and observe what follows.

**Question 10. How do sentences begin in academic prose?**

Enter a full stop, then use Frequency to make this list<sup>4</sup>.

Do they mostly begin with Organisation/Orientation language?

## Context

This second level search allows us to limit our top level search items by words and parts of speech.

**Question 11. someone ... they – see Q 147.**

Enter *someone* in to Simple Query and *they* into Context Lemma Filter, allowing 3 tokens to the right. In which sentences does the plural *they* refer to the singular *someone*?

<sup>4</sup> [http://ske.li/arc\\_4gram\\_sentence\\_beginnings](http://ske.li/arc_4gram_sentence_beginnings)

**Question 12. Is *not only* always completed with *but also*?**

Search for *not only* in the Simple Query or Phrase field and observe how many hits there are. Then repeat the search adding *but also* in the Context Lemma Filter field.

**Question 13. What's the phrase with ...**

...term ... understood

... term ... embodies

... difficult to define ...

last ... least

**Question 14. When *cause* is used with another verb, is it the to-infinitive form?**

In this search, the verb in Simple Query is followed by *to* up to 3 tokens to the right and a verb up to 4.

The screenshot shows the Sketch Engine search interface. The 'Simple query' field contains the word 'cause'. Below it are links for 'Query types', 'Context', and 'Text types'. The 'Context' section is expanded, showing two filters: 'Lemma filter' and 'PoS filter'. The 'Lemma filter' is set to 'Window: right 3 tokens' and 'Lemma(s): to all of these items'. The 'PoS filter' is set to 'Window: right 4 tokens' and 'PoS: all of these items'. A dropdown menu for the PoS filter is open, showing options: 'numeral', 'particle', 'preposition', 'pronoun', 'verb' (which is checked), and 'full stop'. At the bottom of the context section are buttons for 'Make Concordance' and 'Clear All'.

To search for a verb followed by a verb without *to* in between, select *none of these items*.

This screenshot shows the 'Lemma filter' settings. The 'Window' is set to 'right 3 tokens'. The 'Lemma(s)' field contains 'to' and the dropdown menu is set to 'none of these items'.

In a large corpus, there will be very many hits. To make this manageable, make a sample of the findings, say 500. Then Right Sort the concordance page to see the patterns. What patterns do you observe?

The same question can be asked of many common verbs. For example, *sb makes sb do something* or *sb makes sb to do sth*.

Also, *let, force, require, encourage, enable, know how, expect, manage*. And many more.

When looking at the results, it is useful to see if there are any typical nouns and adverbs occurring with the verbs. These pairs are known as collocation.

## The language of comparison

**An academic writing task:** compare the abstract we studied in class with one that you've read in your academic work.

The language of **comparison** starts with the word *compare*. Create the **word family** of this word using these operators.

Lemma:  PoS: unspecified

Then summarise the results using the Word Forms button<sup>5</sup>.

This data can also be restructured using *lempos*<sup>6</sup>. Clicking on **P** beside the words in these lists, and then sorting the pages shows the **phrases** that are part of, e.g. *Comparisons of X and Y ...* (at the start of a sentence), *As a comparison, ...*, *based on a comparison between ...*

## Word sketch

When determining how the abstract relates to the guidelines for abstract writing, look at the collocations of words such as *advice*, *rules*, *principles*, *guidelines*. For example, using this abbreviated word sketch of *advice*<sup>7</sup>, you can see how the word combines with other words in various grammatical relationships.

**advice** (noun)  
British National Corpus (TreeTagger) freq = 9,689 (85.75 per million)

object_of	subject_of	adj_subject_of	modifier	modifies
3,755 3.40	828 1.00	154 1.80	3,773 1.30	793 0.30
seek + 365 9.04	regard 15 4.75	fatal 3 5.73	legal + 397 9.49	bureau 31 8.67
offer + 330 8.02	please 5 4.18	available 46 4.88	professional + 155 8.40	agency 54 6.67
heed 30 7.88	influence 3 3.26	sensible 3 4.80	sound 73 8.33	guidance 19 6.65
ignore 63 7.43	suggest 7 3.24	excellent 3 3.74	independent + 140 8.21	worker + 107 6.53
give + 746 7.35	seek 5 3.05	free 7 3.26	practical + 112 8.18	centre 98 6.23
obtain 64 6.67	apply 6 2.90	relevant 3 3.06	medical + 114 8.18	session 28 6.14

## Matching activity

object of	what advice does
subject of	types of advice
adj subject of	advice and another noun
modifier	sb or sth <i>verbs</i> advice
modifies	advice is ...

## More grammatical relationships

and/or	pp_on-i	pp_obj_of-i	pp_obj_for-i	pp_of-i
1,909 1.70	807 8.90	699 1.50	509 5.00	478 1.00
assistance + 142 8.93	contraception 5 6.87	giver 4 7.22	pharmacist 7 8.09	Clerk 3 6.27
guidance 77 8.43	suitability 4 6.45	plenty 21 6.38	grateful 13 6.51	psychiatrist 3 5.86
help + 153 8.15	eating 3 5.85	piece 67 6.32	Support 5 6.51	counsel 4 5.79
encouragement 25 7.57	diet 14 5.67	spite 7 6.04	ask 80 5.75	therapist 3 5.61
information + 197 6.70	cultivation 3 5.65	source 63 5.94	request 10 5.21	expert 18 5.61
support + 138 6.66	tariff 3 5.18	recipient 4 5.70	eligible 3 5.19	solicitor 11 5.39

For example, such collocations are the core of clauses:

sb is grateful for your/any advice,

the advice of a \_\_\_ expert,

sb gives/offers advice and encouragement.

<sup>5</sup> [http://ske.li/bnc\\_compare\\_wf](http://ske.li/bnc_compare_wf)

<sup>6</sup> [http://ske.li/bnc\\_compare\\_wf\\_lempos](http://ske.li/bnc_compare_wf_lempos)

<sup>7</sup> <http://ske.li/c4u>

## Now it's time to make and use our own corpora

# "How people use words to make meanings"

### AWL nouns

- ↗ analogy
- ↗ benefit
- ↗ category
- ↗ concept
- ↗ context
- ↗ element
- ↗ entity
- ↗ factor
- ↗ hierarchy
- ↗ ratio
- ↗ role
- ↗ structure

### Field-based words

- ↗ collocation
- ↗ connotation
- ↗ corpus
- ↗ frequency
- ↗ lemma
- ↗ lexicography

### Article-based words

- ↗ double helix
- ↗ introspection
- ↗ mush
- ↗ ontology
- ↗ peacock
- ↗ cognitive profile

Article by Patrick Hanks

What are the key words and phrases in your corpus and how are they used?

- general academic words
- field based words
- article based words
- discourse markers/signposting
- use of hedging
- first person

[http://bit.ly/versatile\\_deske](http://bit.ly/versatile_deske)

