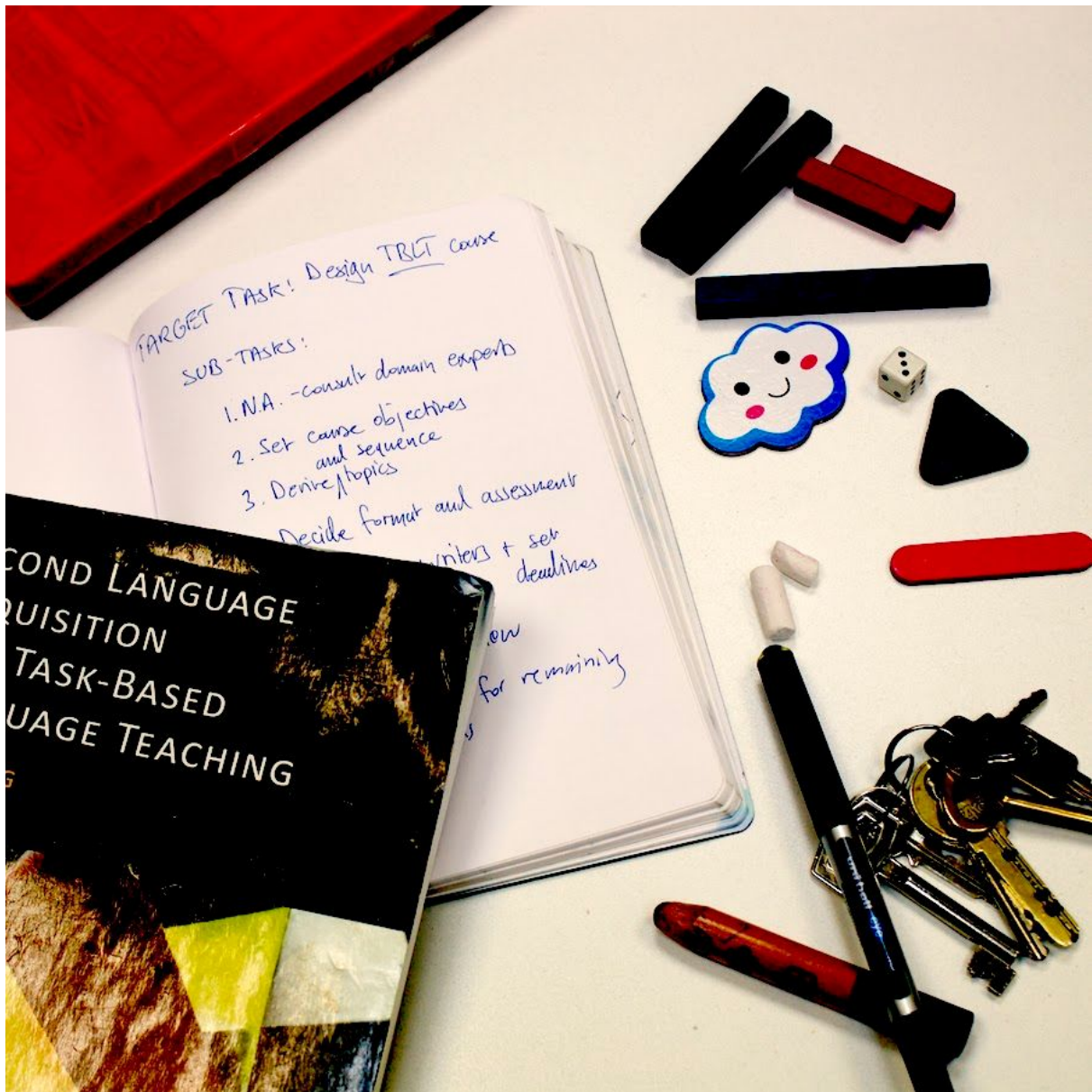


Task-Based Language Teaching (TBLT): From Theory to Practice

Using Corpora for Analysis of Discourse



Marc Jones

Using Corpora for Analysis of Discourse

Contents

Introduction	3
Scraping	3
BootCaT	3
SketchEngine	8
Compiling your own corpus (using SketchEngine)	10
Corpus Analysis	11
SketchEngine	12
AntConc	14
BYU	20
Analysing corpora for TBLT	22

Introduction

Corpora are simply collections of texts, though usually they are sometimes very large, such as the 14-million-word iWeb corpus (available through Brigham Young University's corpus interface, which will be discussed below). Other small, specialized corpora can be around tens of thousands of words.

Corpora are not new. Many people know the word corpus from the Latin, and the name Corpus Christi, or 'the body of Christ'. The term comes from monastic collections of texts on certain subjects. Those monks required a lot of reading and purposeful human searching which can now be accomplished in a matter of seconds, using computer software called a **concordancer**, or in some cases just very advanced text-editing software. While this may make things sound easy, it might not always be. There are times when corpus results may go against one's expectations. However, this is not a negative; one of the bonuses of using corpora is to avoid wholly relying upon intuition and instead draw upon authentic language use when designing tasks.

Scraping

When using corpora for discourse analysis (or analysis of discourse) in task-based language teaching, there are two main things we can do: gathering discourse -- often by scraping the web -- and analysing. Scraping is a process of automatically pulling text from websites and is one of the most convenient ways to build a corpus. If you want a written corpus this is especially useful. Even if you want a spoken corpus, it might be difficult to get access to real speech in the context you want to analyse discourse in (especially medical contexts), but samples of written discourse can provide the course designer and teachers with lexis and grammar that has a high probability of coming up.

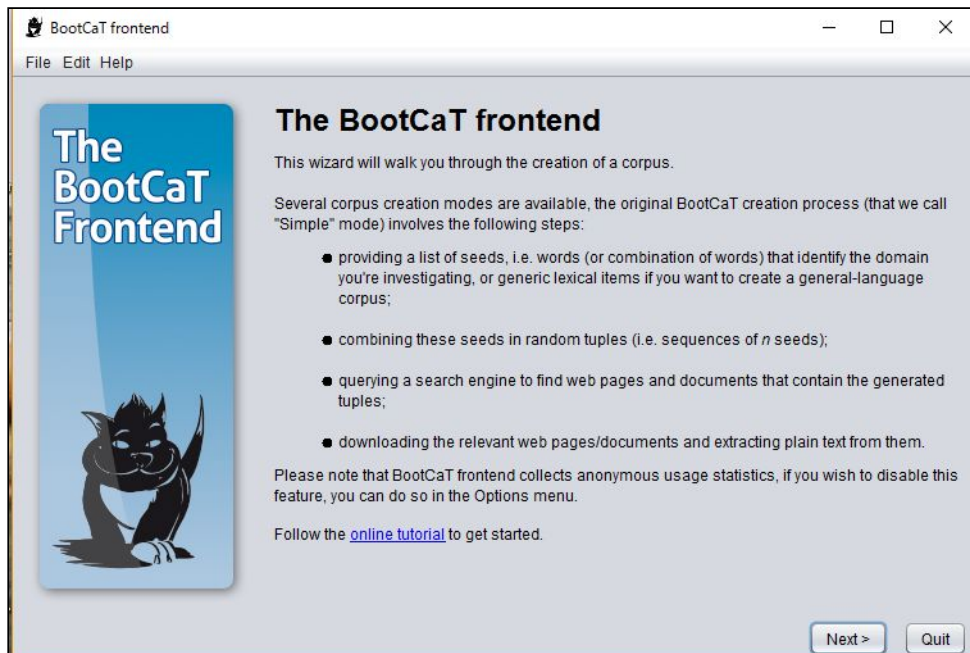
For beginners, you can scrape the web in two ways: one with the **BootCaT** software installed on your computer (PC/Mac/Linux, free), and another with **SketchEngine**, a web-based tool (subscription, free trial available).

BootCaT

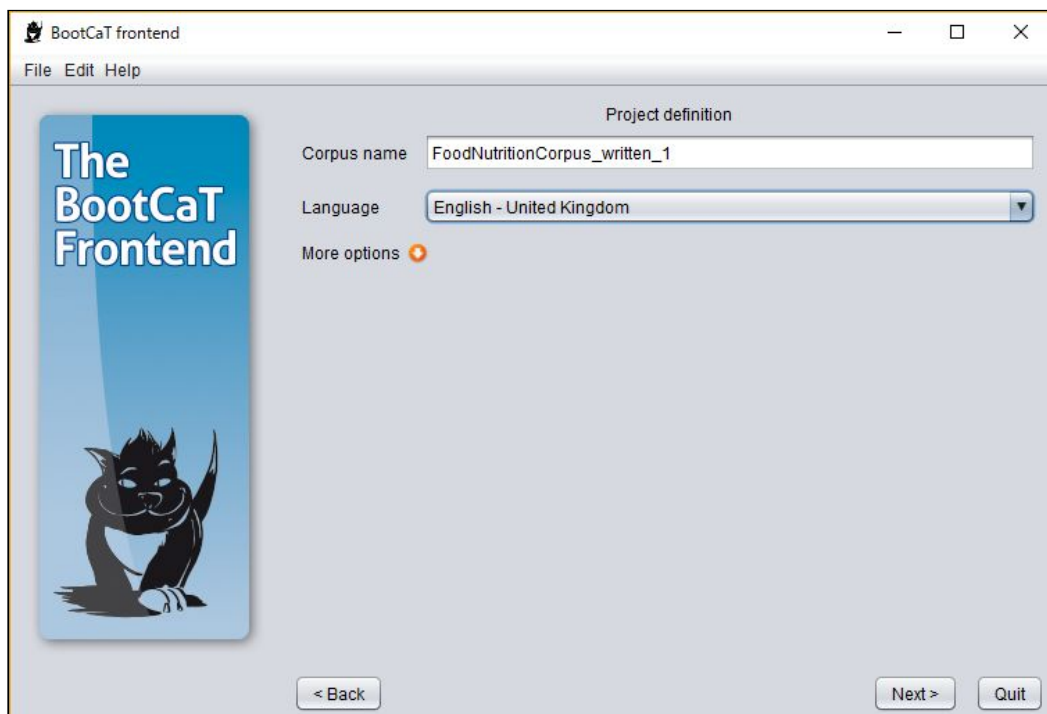
In this example we are going to scrape the web of cooking-related language.

1. Check you have the current version of Java by downloading it from <https://www.java.com/en/download/>. Install it. It should take about five minutes. Download the BootCaT software from the website, <http://bootcat.dipintra.it/>, then install it.

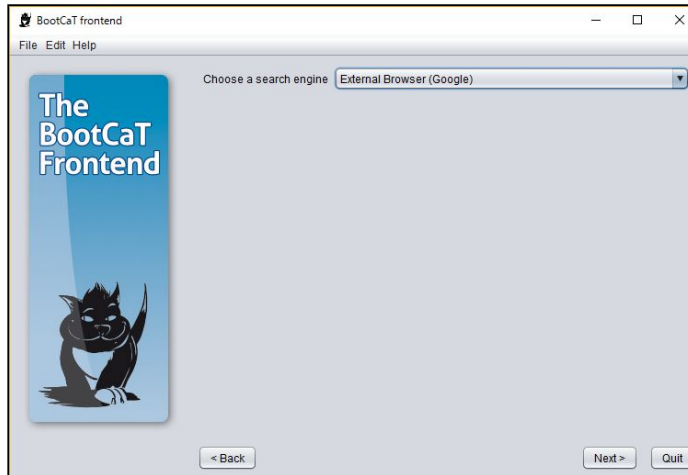
2. Start the software.



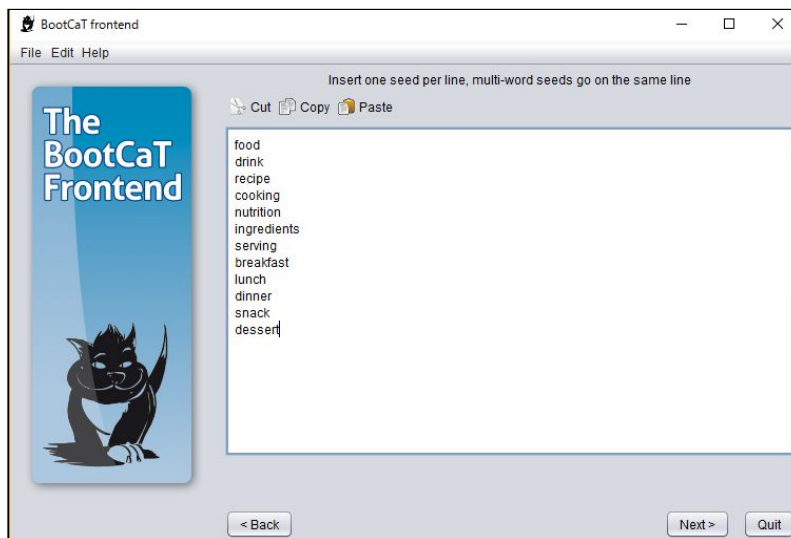
3. Choose a name for your corpus. I would also recommend giving it a number in case you want to edit it or redo it afterwards.



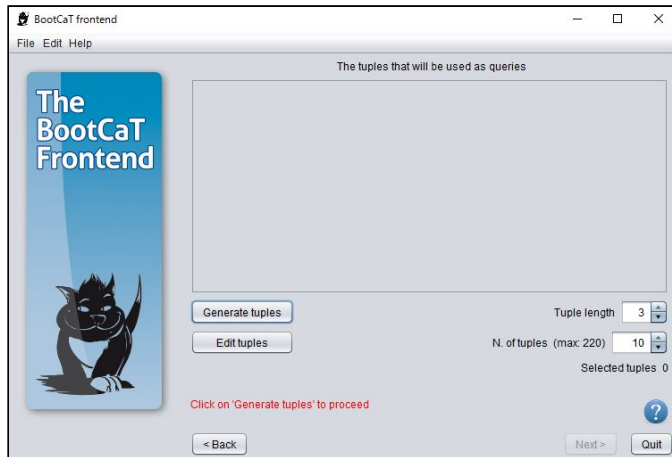
- Click Simple Mode then choose your search engine to be Google. It is not worth the hassle of using Bing.



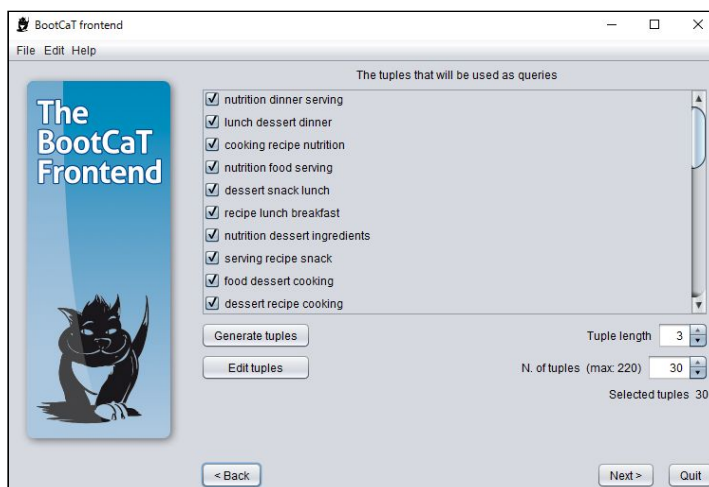
Choose your seed keywords. If you have multi-word terms these should go into a line by themselves.



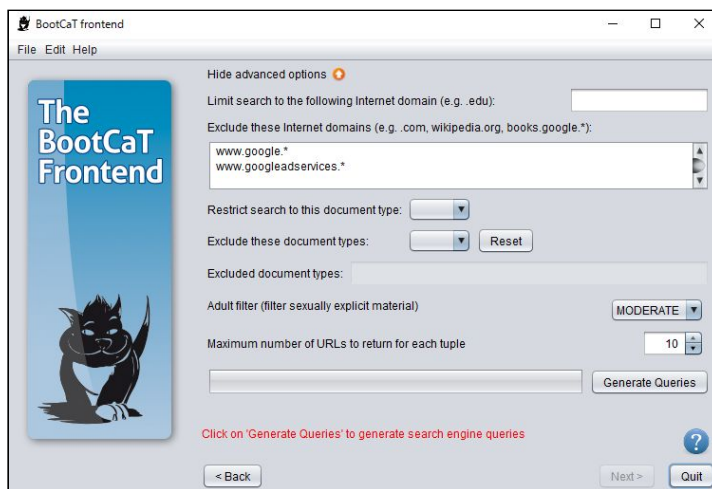
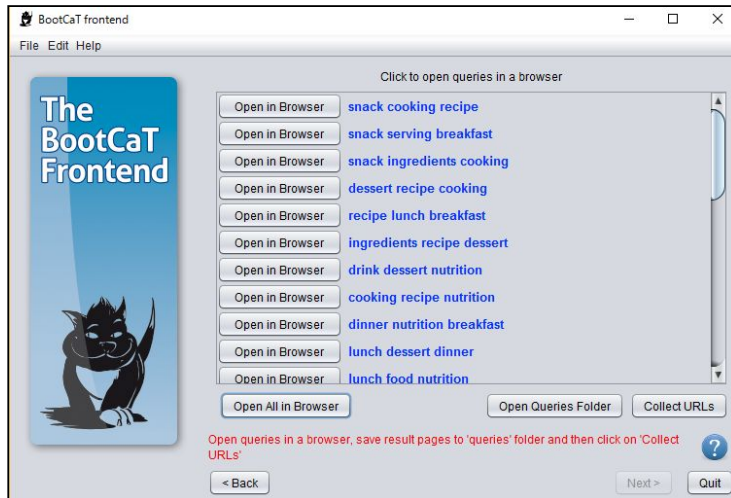
- Choose your tuples (combinations of seeds). How many seeds should combine and how many tuples do you want? In this case we will just use the default settings.



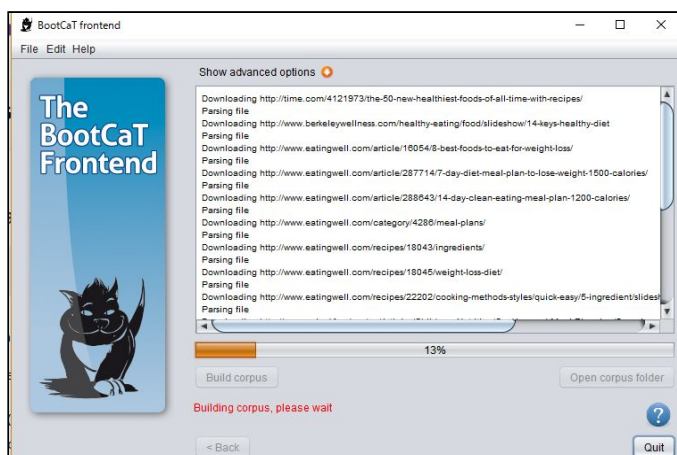
This will then generate the list and you can disable any tuples that you don't want but that is pointless in most cases because BootCaT deletes duplicate data.



6. Generate the queries by clicking Open All in Browser, (unless you have an extremely large number of queries, over about 50, otherwise Google might lock your account if you are signed in, or block your IP address temporarily) and saving the Google search results. If Chrome is your default browser you cannot do this from the menu button. Right click on the page and (Save Page As) and choose HTML only. You need to save it into your BootCaT Corpus Directory, which is at My Documents/BootCaT Corpora/Whatever you called your corpus/queries/ .



7. Make the corpus by waiting for BootCaT to pull the websites and clean them.

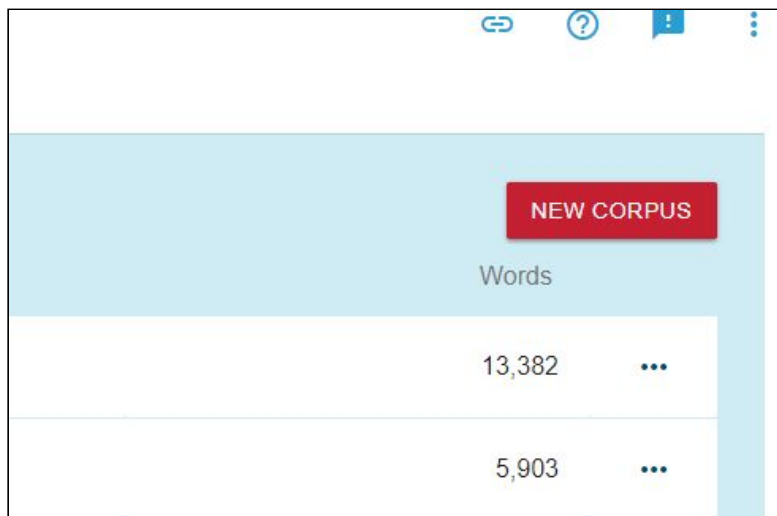


8. You are now ready to choose your concordancer to do the analysis: see the Corpus Analysis section below.

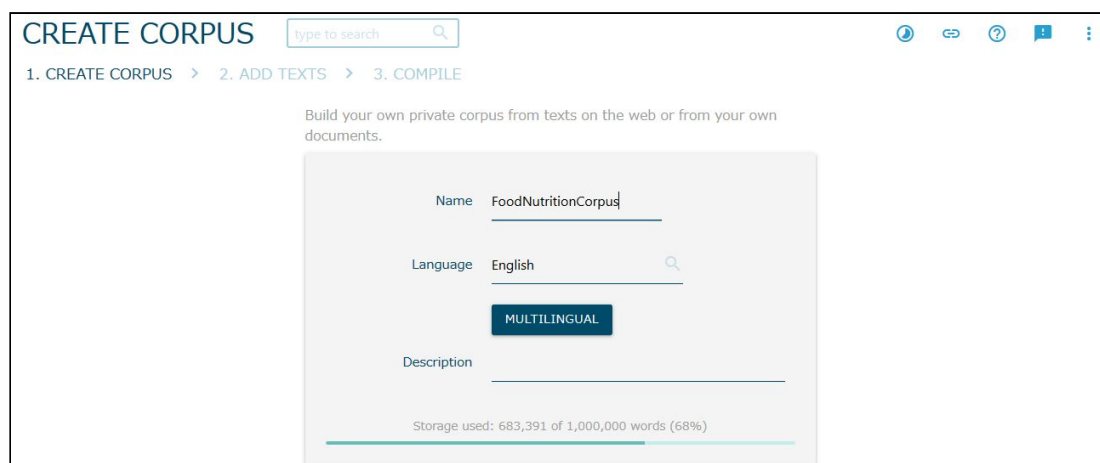
An alternative to BootCaT, which can be temperamental after Java updates, is SketchEngine. Although this requires a paid subscription (after an initial free trial), it's a powerful tool.

SketchEngine

1. Go to <http://app.sketchengine.eu/> and register for a free trial.
2. You can build your own corpus or upload one. First we'll show you how to use SketchEngine in the same way as we've just demonstrated with BootCaT, i.e. by scraping the web to build a corpus.
3. Log in to SketchEngine, select the **My Corpora** tag and click **New Corpus** on the right.

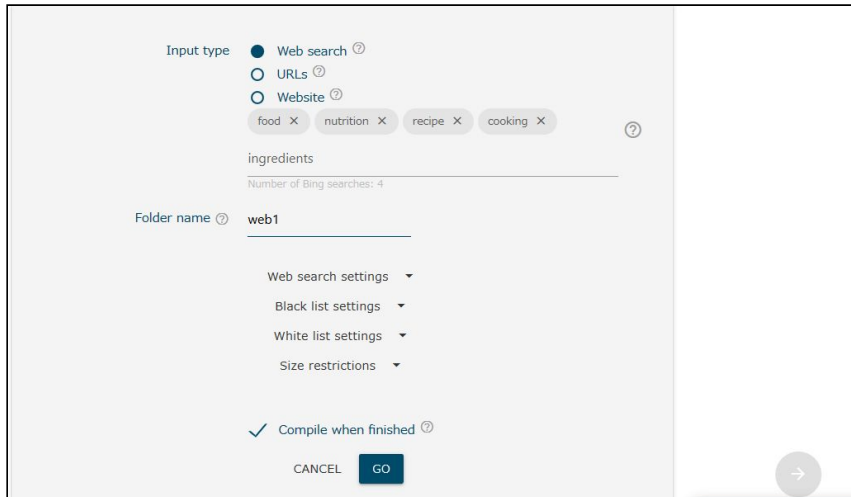


4. Give your corpus a title, select the language and click **next**.

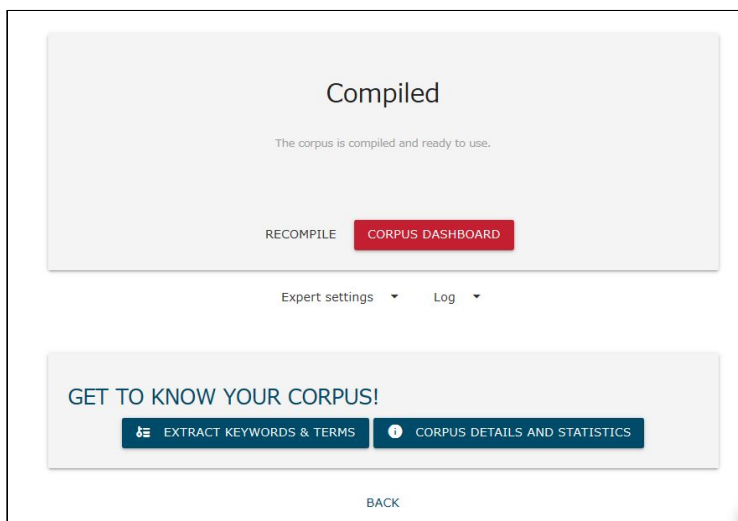


5. Select **Find texts on the web**. SketchEngine gives you the option of choosing search words or websites. It is advisable to choose words for a wider variety of discourse. Choose your seed keywords. Separate them by commas. If you have

multi-word terms these should be spaced as normal with a comma after the multi-word terms themselves.



6. You can choose how big the files should be. It is probably best not to fiddle with these numbers. SketchEngine works well with PDF files so big files should be included. Remember that a webpage is usually only a matter of kilobytes so the small size is useful. SketchEngine will tell you when it is finished.



Compiling your own corpus

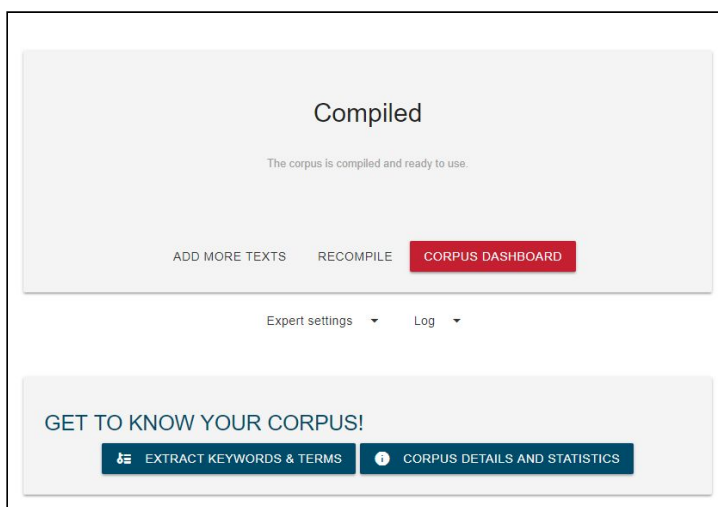
In this instance we will look only at SketchEngine, which is by far the easiest as it accepts a wide range of document formats and does all the tagging automatically.

The advantage of compiling your own corpus over scraping is that by collecting examples of specific target discourse. e.g. medical abstracts or transcriptions of Scrum demo dialogues, we can analyse language as it is used to complete the target task, rather than the wider language of the domain.

1. Collect your examples of target discourse e.g. in .txt, .doc or .pdf format and place them in a folder on your computer. They could be examples of texts written for a specific purpose, or transcriptions of spoken English that you have typed up yourself or used YouTube to create. See the Session 3 Extension for more information on how to do this.
2. Follow the above steps on using SketchEngine to scrape a corpus, until **Step 5**. This time, however, select: **I have my own texts**.
3. Select and drag files from the folder your created into the upload box.



4. Wait until your texts have been processed. You can add more texts or folders in the meantime.
5. Click **Compile** and shortly your corpus should be ready to analyse.



Corpus Analysis

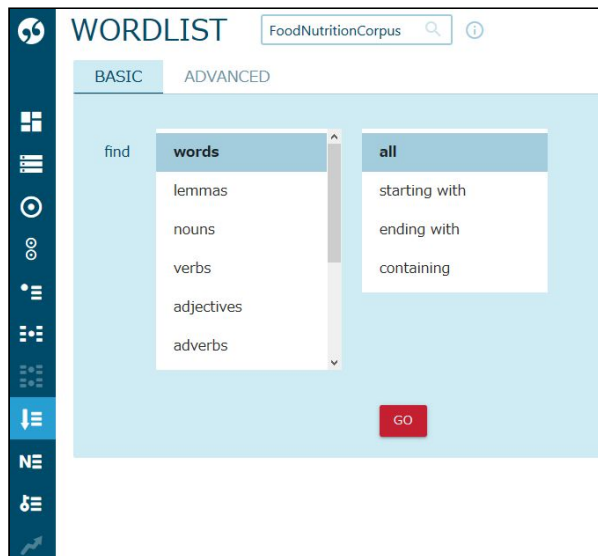
The main way language teachers have their first encounter with corpus linguistics is through dictionaries or vocabulary lists in coursebooks. While dictionaries and word lists both have their uses, most coursebooks contain very little in the way of real-life tasks. One way that we can bring authentic language into authentic situations for our learners is by analysing corpora. In this section, we'll use SketchEngine first, then some free software called AntConc made by Laurence Anthony, an academic in Japan who also teaches English for Specific Purposes. Finally, we'll take a quick look at the Brigham Young University corpus interface (BYU Online corpora).

Some output we are likely to want from the corpora we are investigating:

- **Concordances** - samples of your corpus, with your search term in the middle. Your concordancer usually shows between 10-20 lines of examples. You won't usually get full sentences but you can click to get the full context. The concordance functions are extremely similar in SketchEngine, BYU Online Corpus and AntConc. You can search for single words, phrases and you can also search for parts of words using wild card functions like * and ?. The wildcard functions let you leave a gap so you can search for grammatical patterns in AntConc. Additionally, if you have programming experience, you can also search using Regex, which can save you time, but is not essential for our purposes.
- **Word list** - the words in the corpus ordered by frequency;
- **Keyword list** - the words in the corpus that occur at a statistically higher frequency than those in a reference corpus. This is often referred to as 'keyness' which is positive for higher frequency words, and negative for words that appear statistically more frequently in the reference corpus.
- **N-Grams** - these are essentially chunks of language ordered by frequency. It's usually a good idea to check the frequency of N-Grams against high frequency words because some of your N-Grams are likely to be more high frequency than your key words. The name comes from the length being N words long, or the user chooses how long they are. Mostly you'll choose up to five words long.
- **Collocates** - these are words that appear within a certain distance of your search term. The default settings in most tools are 5L 5R, which simply means within five words to the left and five words to the right.
- **Clusters** - these are lists of words that are frequently to the left or to the right of your search term.

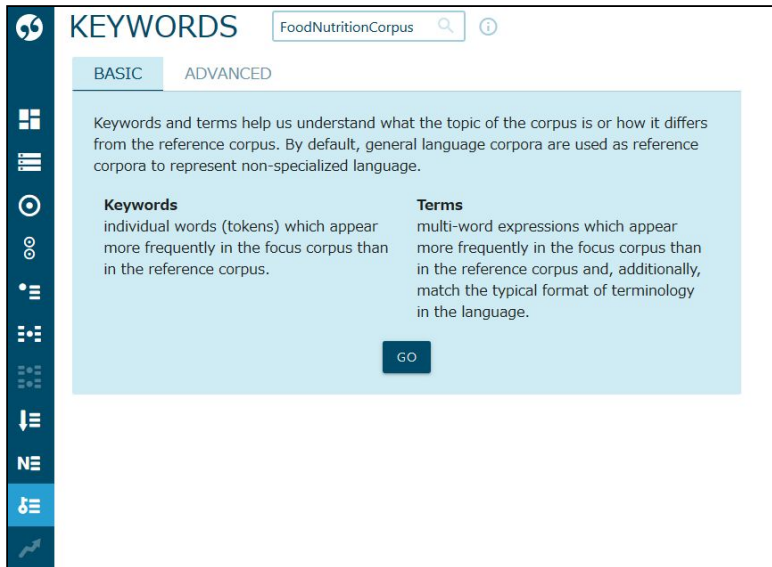
SketchEngine

In the sections above, you made a corpus either by scraping the web or uploading your own folder of texts. You're ready to analyse that corpus now. It's a good idea to generate your word list first (the most frequently occurring words in your corpus). This will give you an idea of how frequent your search terms are (in the case of a scraped corpus) and if any other specific terms are high in the list. It might mean going back a stage and remaking your corpus (and possibly deleting the one you just made due to lack of storage space).

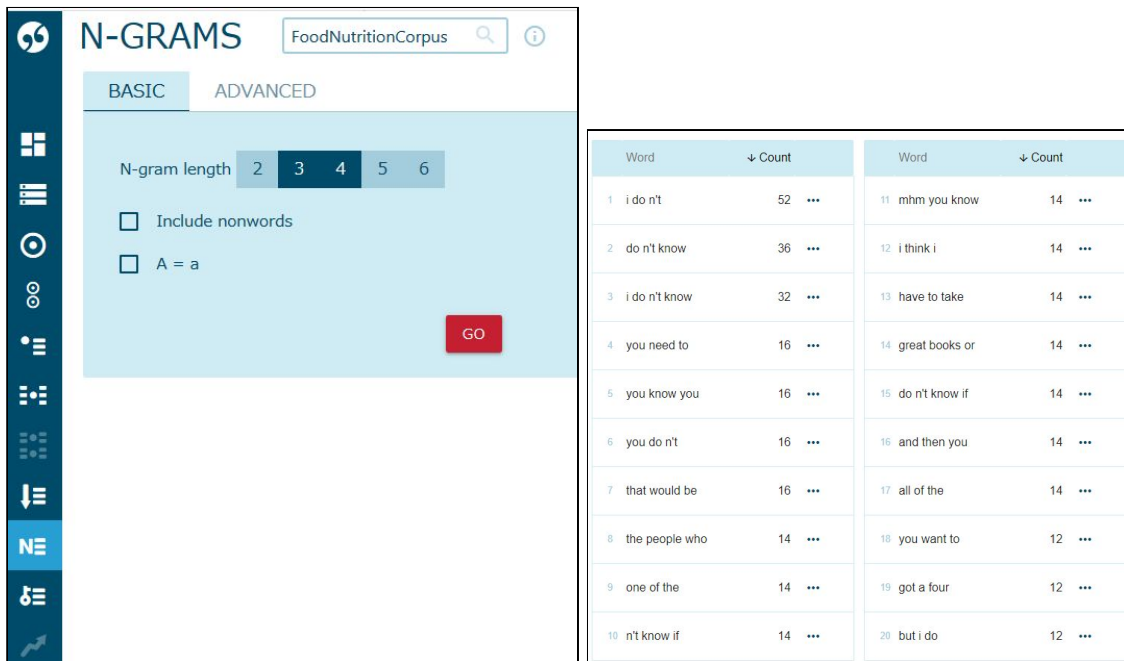


Your **keyword list** is also incredibly useful. It is a list of the most probable words to occur in your corpus as compared to a reference corpus. The reference corpus is just another corpus, usually much larger, that you compare your own corpus to. The default reference corpus in SketchEngine is their own EnTenTen15 web corpus (which is around a billion words in size). You can change this to other corpora such as BNC15 (British National Corpus, 2015) and COCA (Corpus of Contemporary American English). For specialised corpora, such as healthcare, it gives you an idea of how much jargon you might expect in typical written communication, and therefore how much of this would need to be known for spoken communication.

Alongside the keyword list, the new version of SketchEngine will also generate a list of key **N-Grams** or **terms**. There's more on this below. It's worth noting that all lists on SketchEngine can be downloaded in PDF, CSV, XLS and XML formats.



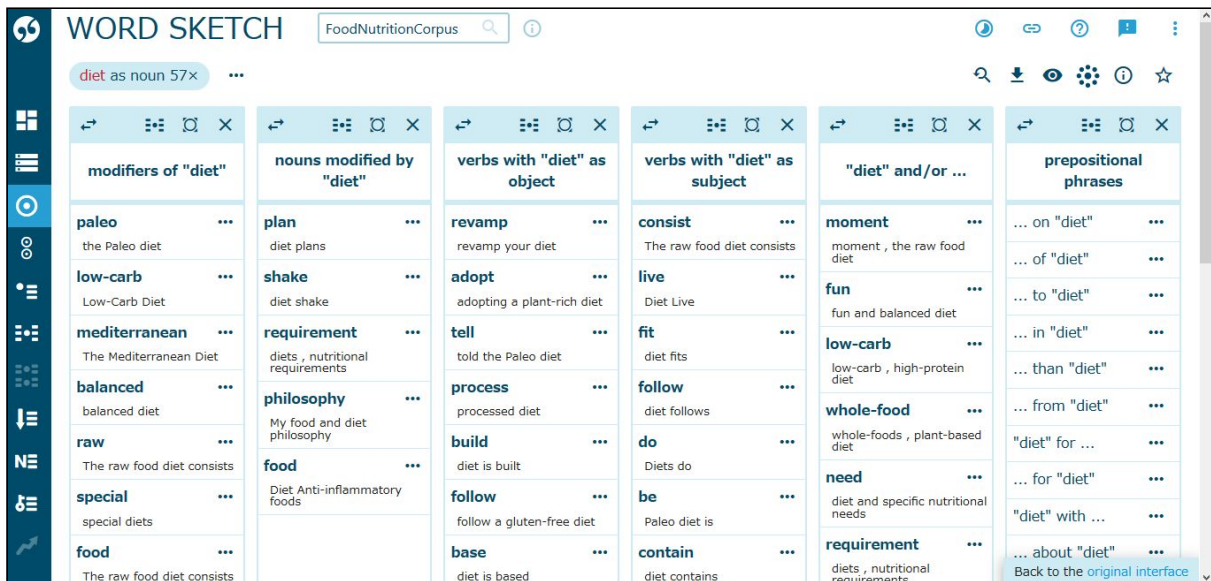
You can also generate a separate **N-Gram list**. If you choose your N-Gram to be 3 or 4 words long (the default) you get three-to-four word chunks ranked by frequency in the corpus. You can also specify the minimum range (how many texts each N-Gram should appear in). Don't be afraid to fiddle with these settings until you get something you can work with. It's really useful for finding idioms and discourse markers, and especially for pragmatic functions like hedging.



Word	↓ Count	Word	↓ Count
1 i do n't	52 ...	11 mhm you know	14 ...
2 do n't know	36 ...	12 i think i	14 ...
3 i do n't know	32 ...	13 have to take	14 ...
4 you need to	16 ...	14 great books or	14 ...
5 you know you	16 ...	15 do n't know if	14 ...
6 you do n't	16 ...	16 and then you	14 ...
7 that would be	16 ...	17 all of the	14 ...
8 the people who	14 ...	18 you want to	12 ...
9 one of the	14 ...	19 got a four	12 ...
10 n't know if	14 ...	20 but i do	12 ...

Most frequent N-Grams in spoken academic advice sessions: from R. C. Simpson, S. L. Briggs, J. Ovens, and J. M. Swales. (1999) [The Michigan Corpus of Academic Spoken English](#). Ann Arbor, MI: The Regents of the University of Michigan

Finally, the **Word Sketch** function is the major advantage of SketchEngine. You can type in a lemma (root form of a word. e.g. 'ride' is a lemma, 'rides', 'riding' and 'rider' are not lemmas). It will give you a comprehensive display of how the word collocates with different words and different parts of speech. It's a good idea to run the Word Sketch on at least the top 20 or so keywords. This should give you an idea of any obvious 'marked' (somewhat unusual) language use, such as nominal modification (nouns used to modify nouns as opposed to adjectival modification), which is often a feature of technical/scientific language.



modifiers of "diet"	nouns modified by "diet"	verbs with "diet" as object	verbs with "diet" as subject	"diet" and/or ...	prepositional phrases
paleo the Paleo diet	plan diet plans	revamp revamp your diet	consist The raw food diet consists	moment moment , the raw food diet	... on "diet"
low-carb Low-Carb Diet	shake diet shake	adopt adopting a plant-rich diet	live Diet Live	fun fun and balanced diet	... of "diet"
mediterranean The Mediterranean Diet	requirement diets , nutritional requirements	tell told the Paleo diet	fit diet fits	low-carb low-carb , high-protein diet	... to "diet"
balanced balanced diet	philosophy My food and diet philosophy	process processed diet	follow diet follows	whole-food whole-foods , plant-based diet	... in "diet"
raw The raw food diet consists	food Diet Anti-inflammatory foods	build diet is built	do Diets do	need diet and specific nutritional needs	... than "diet"
special special diets		follow follow a gluten-free diet	be Paleo diet is	requirement diets , nutritional requirements	... from "diet"
food The raw food diet consists		base diet is based	contain diet contains		"diet" for ...

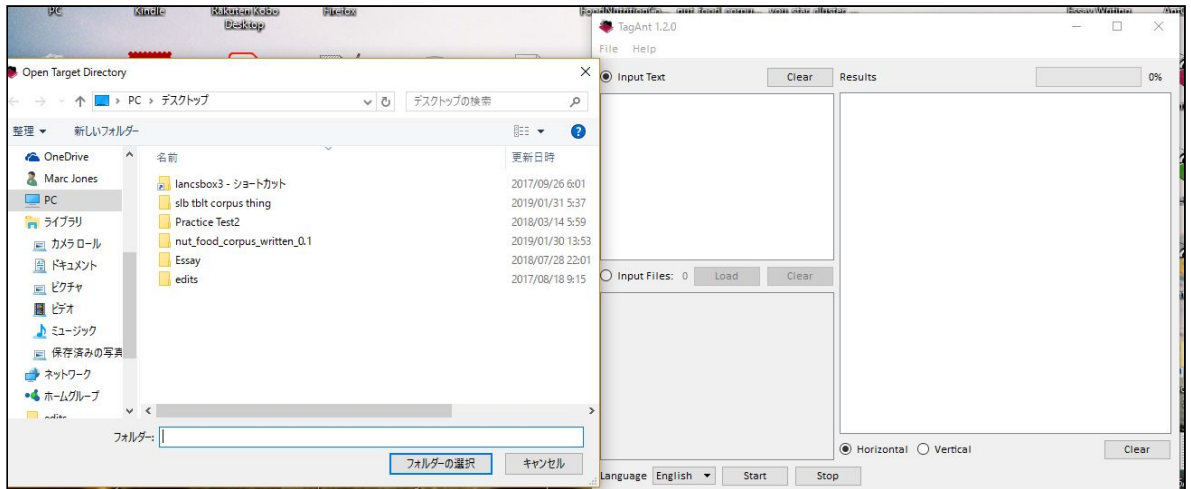
AntConc

This is another way to analyse corpora you have compiled yourself, and is free. However, it involves more work on your part.

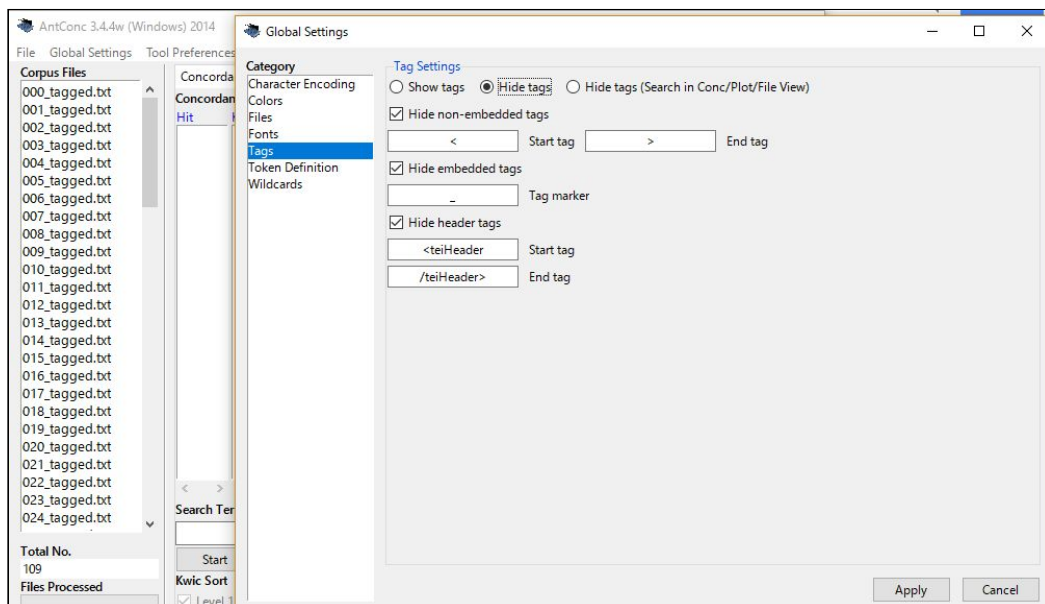
1. Go to Laurence Anthony's webpage <http://www.laurenceanthony.net/software.html> . Download **AntConc** if you haven't already. You might also download **TagAnt** and the tag list, too. You should also download some useful word lists, such as the BNC British English word list and BNC American English word list in AntConc format from Paul Baker's website: <https://www.lancaster.ac.uk/linguistics/about-us/people/paul-baker> . You'll need to scroll down quite far. It's probably more useful to have these word lists on your desktop than in your downloads folder so move them before you start.

2. If you are using a corpus you made in BootCaT, it is not tagged yet. If you are using a corpus that you know is tagged, skip to step 3.

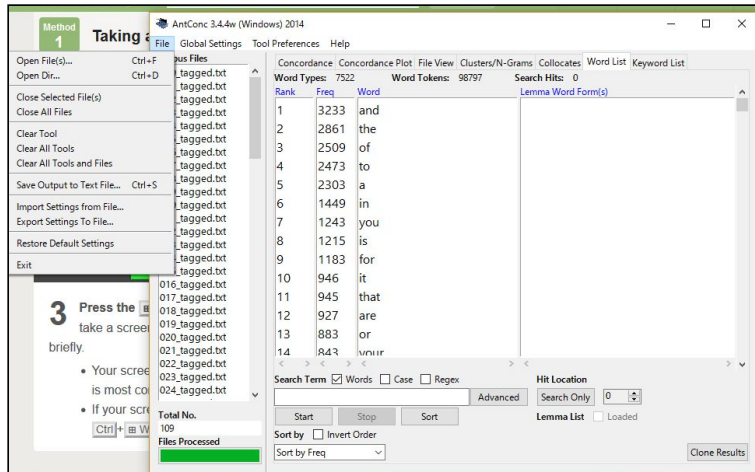
Open TagAnt. Your corpus you made is in a directory that is usually at My Documents/BootCaT Corpora/Name of your corpus project/ so go to File > Open Directory and choose the directory your corpus is in. Load and start. It shouldn't take much longer than a couple of minutes.



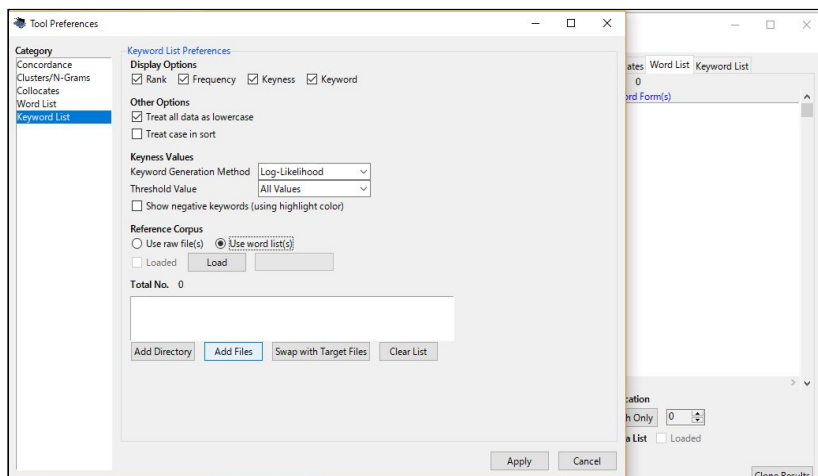
3. AntConc runs as an executable file which means you can run it from a USB flash drive or your computer. Open the file and tell your computer that you really do want to run the file.
4. Go to File > Open Directory and choose the directory your corpus is in.
5. Next, open Global Settings. We'll start with tags hidden. Click Apply after you change this setting.

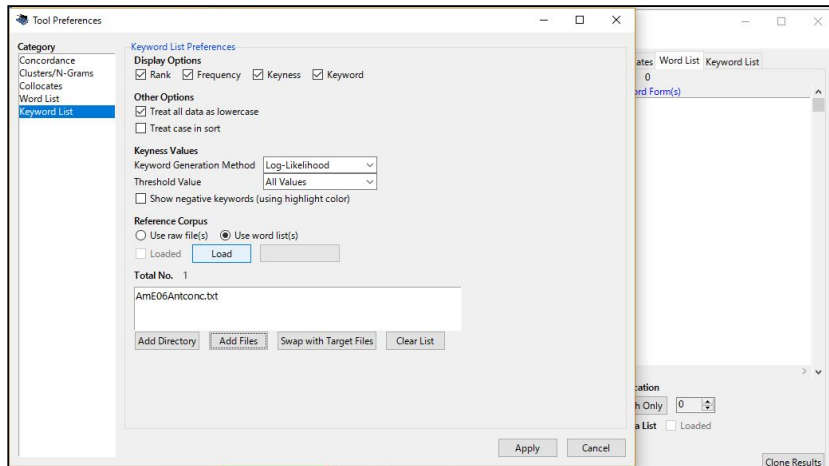


- Go to **Word List**. Click start. You have your word list and you can export it as a text file by clicking File > Save Output As.

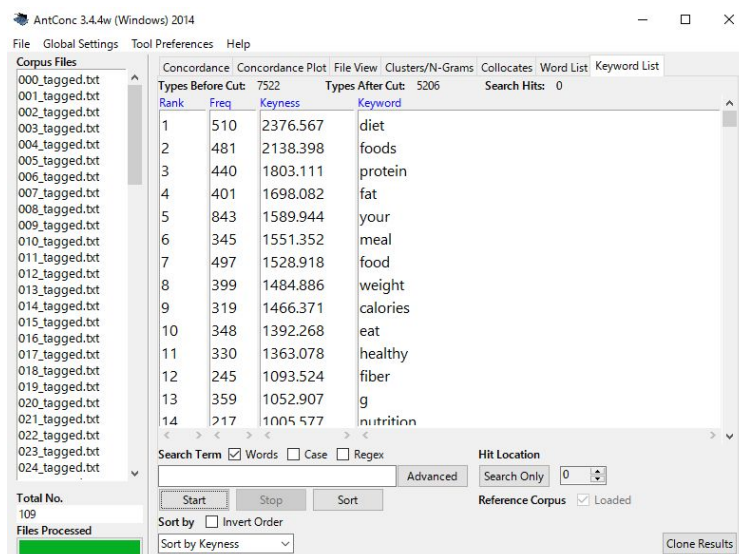


- To generate your **keywords** list you need to load a word list or a reference corpus (see part 1 of this section), then click apply. You probably don't have a whole reference corpus but you have word lists. Go to **Tool Preferences**, choose your word list, load it and then click **Apply**.



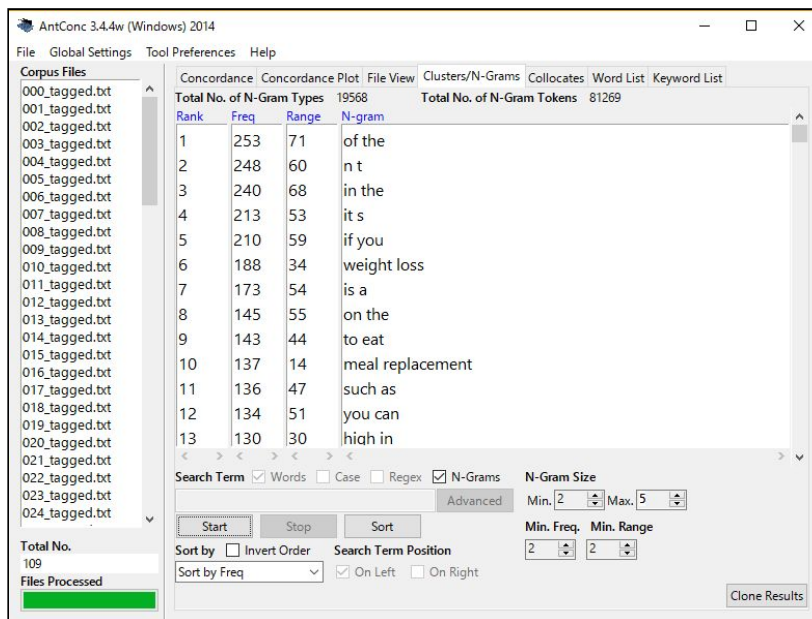


8. Generate your keyword list. Save your output again.

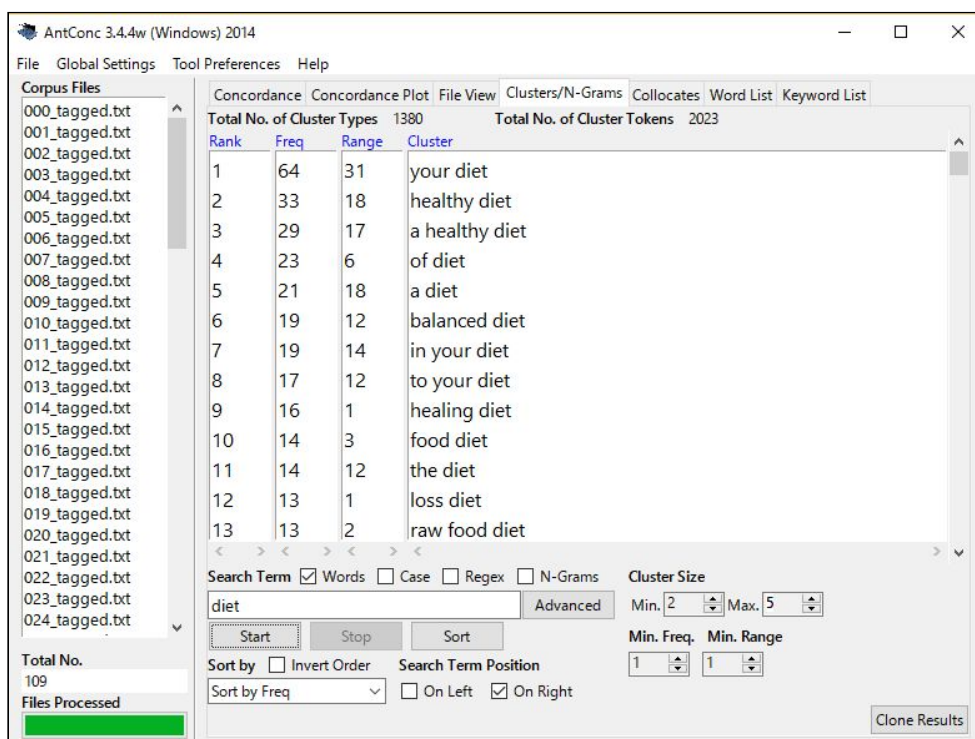


If you are using a corpus which is just one text file (and some corpora are set up like this) set the minimum range (number of text files in the corpus) to 1 in the following steps, otherwise you will have no output.

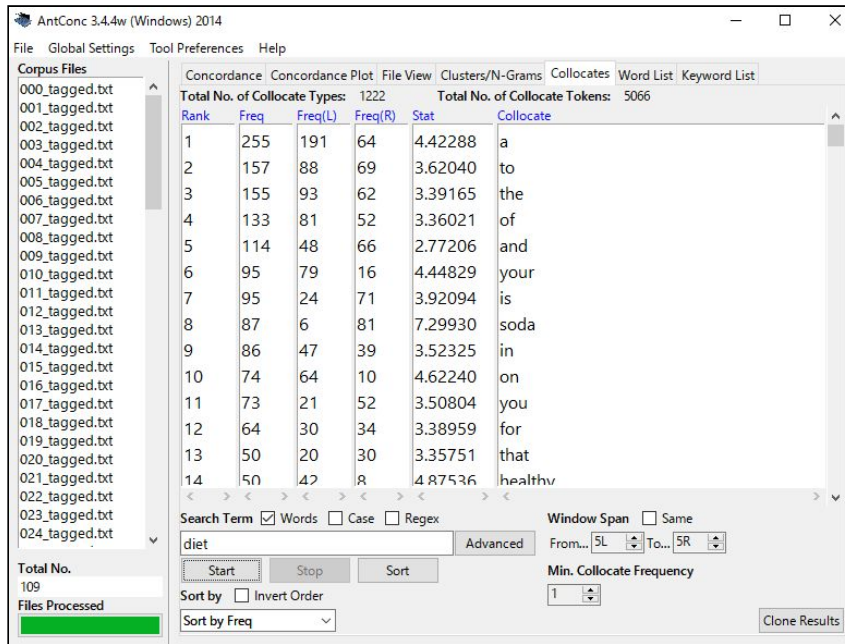
9. You can also generate N-Grams and clusters. N-Grams are easiest. Choose the minimum and maximum size of N-Gram and minimum range you want each one to occur in. Again, do not be afraid to fiddle with things here. When you are ready you can save your output again.



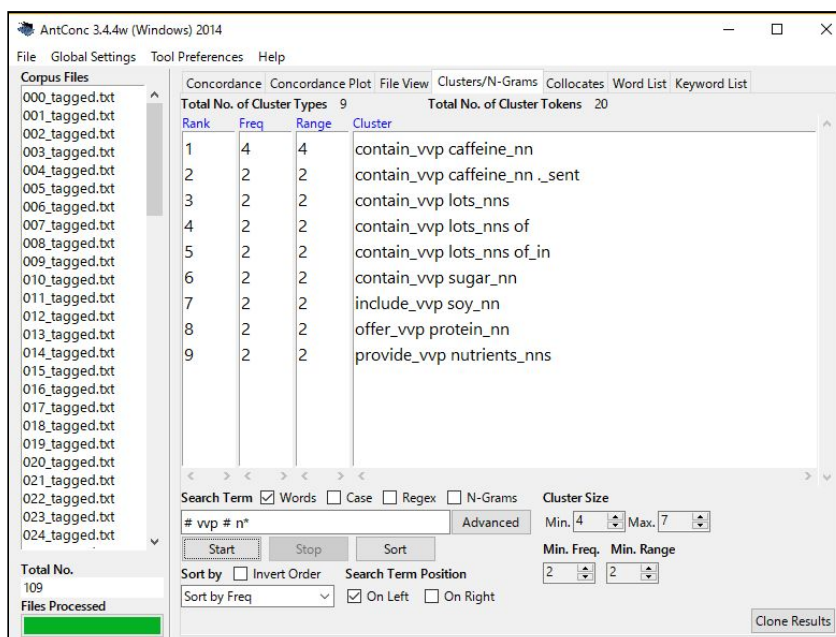
10. Clusters are groups of words with your chosen word at either the left of the cluster or the right. Check the cluster box, set your size and range (quite low is usually better). Type a word then press Start. You can save these to text files in exactly the same way. The same applies regarding minimum range here.



11. **Collocates** in AntConc generates words that appear within different proximities to your search term. The default search is 5L,5R, or five words to the left and to the right of your search term. You get to see if the collocating words occur to the right, the left and their frequencies. You can also save this as a text file.

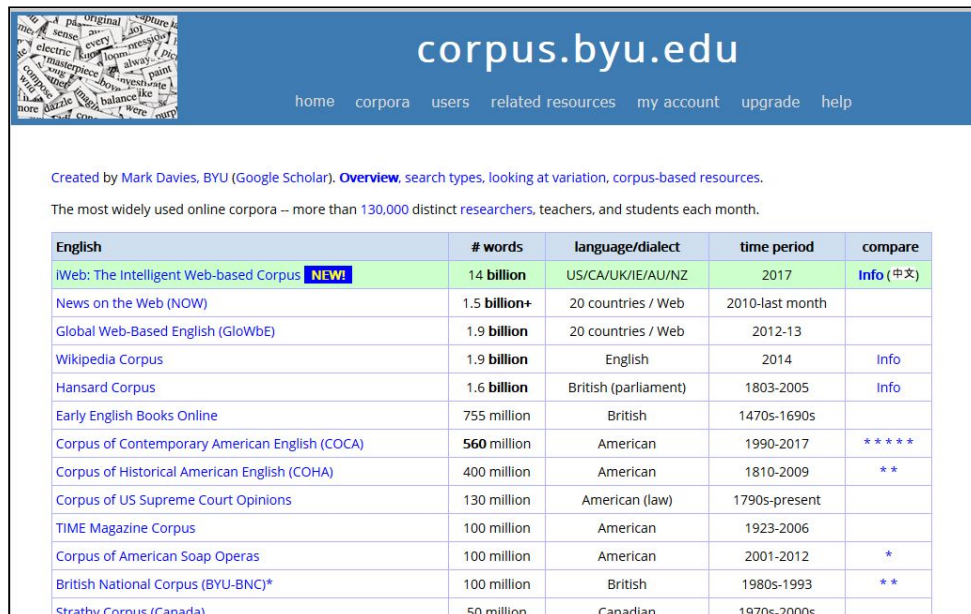


12. To search for grammar patterns, you can go to Global Settings and choose show tags. Cluster, collocates and concordance tools are the most useful for syntax (word order and pattern) investigations. You might want to see if a given noun is more usually preceded by adjectives or nouns. You can click to see examples of these. Tree Tagger tags (the tags used in TagAnt) for simple adjectives are ‘_jj’ and for singular nouns ‘_nn’. To search for [infinite verb + noun] patterns, search “# vvp # n*”. # and * are wildcard characters. # means ‘any one word’ and * means ‘one or more characters’. This means we search for [‘any one word’ ‘vvp’ ‘any one word’ ‘any noun tag’] because all noun tags start with n and are two or three characters long. It may be useful to search this in the cluster tool on the left as well as the concordance tool.



BYU Online Corpora

There are several different corpora available at <https://corpus.byu.edu/>. These are handy for more general applications, as well as for getting used to how corpora work. If you need to analyse 'general' written English, the iWeb or GloWBe corpora may be useful. If you have certain localities in mind, like the US, COCA could be useful and for Canada, there is the Strathy corpus. The Corpus of American Soap Operas could also be handy for generalised spoken language.

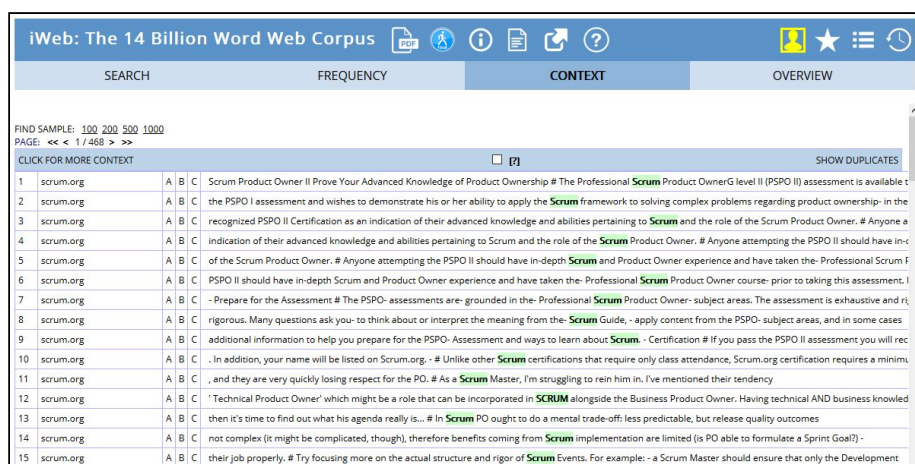


Created by Mark Davies, BYU (Google Scholar). [Overview](#), [search types](#), [looking at variation](#), [corpus-based resources](#).

The most widely used online corpora -- more than 130,000 distinct researchers, teachers, and students each month.

English	# words	language/dialect	time period	compare
iWeb: The Intelligent Web-based Corpus NEW!	14 billion	US/CA/UK/IE/AU/NZ	2017	Info (中文)
News on the Web (NOW)	1.5 billion+	20 countries / Web	2010-last month	
Global Web-Based English (GloWBe)	1.9 billion	20 countries / Web	2012-13	
Wikipedia Corpus	1.9 billion	English	2014	Info
Hansard Corpus	1.6 billion	British (parliament)	1803-2005	Info
Early English Books Online	755 million	British	1470s-1690s	
Corpus of Contemporary American English (COCA)	560 million	American	1990-2017	****
Corpus of Historical American English (COHA)	400 million	American	1810-2009	**
Corpus of US Supreme Court Opinions	130 million	American (law)	1790s-present	
TIME Magazine Corpus	100 million	American	1923-2006	
Corpus of American Soap Operas	100 million	American	2001-2012	*
British National Corpus (BYU-BNC)*	100 million	British	1980s-1993	**
Strathy Corpus (Canada)	50 million	Canadian	1970s-2000s	

You can search for words and collocates, much like in SketchEngine or AntConc above. Unfortunately, there is no N-Gram search in the BYU Online corpus interface. You may want to look at the KWIC concordances.



iWeb: The 14 Billion Word Web Corpus

SEARCH FREQUENCY CONTEXT OVERVIEW

FIND SAMPLE: 100 200 500 1000
PAGE: << < 1/468 > >>

CLICK FOR MORE CONTEXT SHOW DUPLICATES

1	scrum.org	A B C	Scrum Product Owner II Prove Your Advanced Knowledge of Product Ownership # The Professional Scrum Product OwnerG level II (PSPO II) assessment is available t
2	scrum.org	A B C	the PSPO I assessment and wishes to demonstrate his or her ability to apply the Scrum framework to solving complex problems regarding product ownership- in the
3	scrum.org	A B C	recognized PSPO II Certification as an indication of their advanced knowledge and abilities pertaining to Scrum and the role of the Scrum Product Owner. # Anyone a
4	scrum.org	A B C	Indication of their advanced knowledge and abilities pertaining to Scrum and the role of the Scrum Product Owner. # Anyone attempting the PSPO II should have in-
5	scrum.org	A B C	of the Scrum Product Owner. # Anyone attempting the PSPO II should have In-depth Scrum and Product Owner experience and have taken the- Professional Scrum F
6	scrum.org	A B C	PSPO II should have in-depth Scrum and Product Owner experience and have taken the- Professional Scrum Product Owner course- prior to taking this assessment. I
7	scrum.org	A B C	- Prepare for the Assessment # The PSPO- assessments are- grounded in the- Professional Scrum Product Owner- subject areas. The assessment is exhaustive and r
8	scrum.org	A B C	rigorous. Many questions ask you- to think about or interpret the meaning from the- Scrum Guide, - apply content from the PSPO- subject areas, and in some cases
9	scrum.org	A B C	additional information to help you prepare for the PSPO- Assessment and ways to learn about Scrum. - Certification # If you pass the PSPO II assessment you will rec
10	scrum.org	A B C	. In addition, your name will be listed on Scrum.org. - # Unlike other Scrum certifications that require only class attendance, Scrum.org certification requires a minim
11	scrum.org	A B C	, and they are very quickly losing respect for the PO. # As a Scrum Master, I'm struggling to rein him in. I've mentioned their tendency
12	scrum.org	A B C	Technical Product Owner' which might be a role that can be incorporated in SCRUM alongside the Business Product Owner. Having technical AND business knowled
13	scrum.org	A B C	then it's time to find out what his agenda really is... # In Scrum PO ought to do a mental trade-off: less predictable, but release quality outcomes
14	scrum.org	A B C	not complex (it might be complicated, though), therefore benefits coming from Scrum implementation are limited (is PO able to formulate a Sprint Goal?) -
15	scrum.org	A B C	their job properly. # Try focusing more on the actual structure and rigor of Scrum Events. For example: - a Scrum Master should ensure that only the Development

You can, however, gain an excellent overview on the main page of the iWeb corpus with some main clusters and collocates.

iWeb: The 14 Billion Word Web Corpus

SEARCH COLLOCATES CONTEXT OVERVIEW

COLLOCATES **SCRUM** NOUN

Advanced options Collocates Clusters Topics Dictionary Websites KWIC

+ NOUN	NEW WORD	?	+ ADJ	NEW WORD	?	+ VERB	NEW WORD	?	+ ADV	NEW WORD	?
5332	4.40	team	3191	9.53	agile	316	3.23	implement	67	4.09	half
4588	6.53	master	1050	6.40	certified	217	3.34	adopt	29	3.58	forwards
1640	3.30	project	995	4.55	daily	187	4.76	scale	26	2.51	backwards
1448	2.50	product	184	4.81	dominant	165	3.67	facilitate	16	2.54	front
1021	3.04	development	155	4.74	lean	165	6.49	sprint	12	6.10	offside
923	3.82	owner	140	6.80	attacking	153	3.89	dominate	5	2.71	superbly
923	5.89	framework	134	2.91	extreme	124	4.71	collapse	4	2.52	purposefully
889	7.18	methodology	122	2.64	Irish	118	2.68	award	4	2.59	optimally
877	4.10	meeting	110	4.21	resulting	110	2.92	opt	4	3.04	comprehensively
848	3.24	role	91	8.56	uncontested	103	3.76	certify	4	3.05	deceptively
770	6.16	alliance	84	7.27	iterative	102	3.72	lean	3	2.58	interchangeably
768	5.50	penalty	72	6.03	ensuing	96	3.01	coach	3	2.68	hotly
758	4.76	half	72	8.97	kanban	87	3.02	practice	3	2.72	inexplicably
666	3.67	guide	71	2.69	defensive	68	3.12	master	3	4.15	uncharacteristically
656	2.73	media	69	5.78	halfway	61	4.54	concede	3	6.47	diversely
547	7.01	sprint	59	5.64	retrospective	56	2.65	reset	2	2.67	ruthlessly
539	3.80	coach	58	4.81	empirical	50	2.69	collaborate	2	2.86	tactically
402	2.97	ball	53	4.30	Welsh	35	4.50	ensue	2	3.09	headfirst

iWeb: The 14 Billion Word Web Corpus

SEARCH **WORD** CONTEXT OVERVIEW

Collocates Clusters Topics Dictionary Websites KWIC HELP

scrum (NOUN) #10333

1. (rugby football) the method of beginning play in which the forwards of each team crouch side by side with locked arms

M O C G

PlayPhrase YouGlish Yarn

Translate: choose language

SYNONYMS (more)

crowd struggle, fray, scrimmage, free-for-all, tussle, scrum, jostle

TOPICS (more)

agile, team, try, penalty, ball, sprint, half, score, project, rugby, software, development, backlog, defence, match, kick, developer, kick, minute, master

COLLOCATES (more)

NOUN team, master, project, product, development, framework, owner, methodology

VERB implement, adopt, scale, facilitate, sprint, dominate, collapse, award

ADJ agile, certified, daily, dominant, lean, attacking, extreme, Irish

ADV half, forwards, backwards, front, offside, superbly, purposefully, optimally

NOUN + NOUN (more)

scrum NOUN	scrum master • scrum team • scrum teams • scrum alliance • scrum masters • scrum half •
NOUN scrum	media scrum •

Analysing corpora for TBLT

Before you start analysing the corpus, you need to bear in mind the task for which you are searching for related discourse. In this example, let's take the corpus we made using BootCaT and think of a task related to English for nutrition students, *Give general advice regarding diet*.

First it would be useful to select some words to search in the collocates and clusters tools in AntConc or in SketchEngine. These words can be based upon our own intuitions regarding the keywords, e.g. diet, eat, healthy, you* (equals you/your). Among other items were:

for a healthy	plans	fewer
of a healthy	keeps	smaller
(help) maintain a healthy	(may/can/to) help you	foodists
as part of a healthy	good for you	choosing
healthy diet	boost your	regularly
healthy eating	increase your	what you eat
healthy fats	you lose weight	foods to eat
healthy weight	your metabolism	foods you eat
heart	your diet	that people who eat
choices	your body	feel free to eat
living	you can	the foods you eat
maintaining	you to	eat fewer calories
	you should	eat plenty of

We then can look at the words and patterns that are found with our search terms. We then have more to look at and can search our corpus concordances for these and they will give use some ideas of written discourse which we shall have to evaluate as to whether it is appropriate for spoken output in our prototypical task.

What comes back to add to the above is:

Eat plenty of fruits and vegetables
 Eat plenty of fruit, vegetables and wholegrains
 Eat plenty of soy and bean products
 As part of a healthy diet
 Help you maintain
 Maintain a healthy weight

Maintain muscle

The (comparative adjective) food you eat, the (comparative adjective)

To boost your metabolism

Foods in your diet

People eat fewer calories (preposition)

You can also/easily/get

We now have a large amount of authentic language to help create a prototype task featuring these main patterns. The task could look something like the below, although bear in mind we would need to take other factors into consideration apart from the key vocabulary and collocations, i.e. the modality of the task (spoken or written), the audience, and the typical structure (e.g. turns, moves etc.). See the Session 6 presentation for more on this.

Prototype Task: Give general advice regarding diet.

As part of a healthy diet, you should eat less fewer salty foods and drink less caffeine. The more salt and caffeine you consume, the faster your heart beats and in our current environment, a lot of people have racing hearts. It's also important to eat plenty of fruit, vegetables and wholegrains to maintain energy throughout the day while also reducing the amount of heavier carbohydrates you eat, such as bread, rice, pasta and potatoes. The more vegetables you eat, the more that these count toward your carbohydrate intake. This will help you lose weight if necessary, while also maintaining enough calories to fuel you for the day. You shouldn't cut out fats and heavy carbohydrates completely: feel free to eat these in moderation. The important point is to reduce the foods that are unhealthy when eaten too much and to increase foods that will boost your metabolism and maintain your health.