**The Prime Machine HD Guide**

***I want to …*  learn more about collocation**

If you are an undergraduate or postgraduate student, using English as the language of instruction, you will need to develop your academic writing skills. Teachers and textbooks will often encourage you to pay attention to combinations of words which make the expression of ideas concise and clear. When we look in a corpus at the combinations of words which occur, we can see that words have a tendency to co-occur with a smaller set of other words; the computer can determine the strength of these relationships by calculating the likelihood of two or more words co-occurring in relation to their frequencies in the corpus overall. You can select a general corpus containing many different kinds of texts (such as the complete British National Corpus), or a corpus of academic texts from a variety of disciplines (such as the BNC: Academic), or a corpus of academic texts from a specialized field (such as the Hindawi academic corpora, or a corpus you build yourself). The results of these statistical processes mirror the relationships between words which must exist in the minds of the communities of speakers and writers as they share, discuss and create knowledge in their specialist areas. The results can help you notice combinations of words which are likely to seem more natural and help you become more aware of combinations which may be considered more rare or creative. If you put words together just based on basic meanings from wordlists or dictionaries, you will probably discover that your EAP teacher and subject professors find that your expressions seem unnatural or awkward or simply just don’t seem to fit together. When you are trying to communicate a new idea, or you want to use language more creatively, knowing the collocations which already exist will make your new combinations stand out. This guide will explain the steps you can take to deepen your knowledge of collocations through game activities and analysis using The Prime Machine HD corpus tool.

Steps to complete:

1. Learn how to spot collocations in concordance lines and other corpus data by playing some games about academic synonyms and prepositions.
2. Learn how to use the search support features to help you find combinations of words to compare.
3. Learn how you can build your own corpus, and view collocations of words of interest within it.
4. Explore the collocational strength between words in your own sentences, by using an online corpus as a baseline.

This approach can help you:

1. Learn how to notice the typical patterns of usage of different words and collocations in general English, academic English and in specialist fields.
2. Find combinations of words from your own writing which may need reformulation.
3. Discover the patterns of usage of collocations which may be distinct from the words from which the combinations are formed.

What you’ll need to get started:

* The Prime Machine HD for Windows, macOS, iPad, iPhone or Android (available free from <https://www.theprimemachine.net/> )
* Some combinations of words to inspire your exploration after you’ve tried the games - perhaps phrases you want to learn more about or some sentences from your own writing to check for collocational strength.
* Patience, enthusiasm and an open mind!

**Looking at your use of language through linguistic spectacles**

You don’t need to be a student of linguistics to be able to analyse the patterns of use of language (but of course if you’ve studied linguistics you can draw on your linguistic knowledge). When you complete an academic assignment, you are essentially showing to your professors, the institution and the world that you are becoming a member of an academic community – a community which communicates using academic conventions and through academic language. Your use of academic vocabulary and specialist terminology reveals your experience reading and hearing these words and phrases as they are applied in your specialist discipline. Some collocations can be found in different disciplines and across text types, but others can have special meaning for members of a specialist field of study.

**Are you ready to become a language detective?**

To make the most of the approach described here, you will need to try to think about your word choices as a detective might look at clues. As well as being an amazing expert in detection, Sherlock Holmes is also remembered in the stories as being an expert in disguise.

* As a detective, can you look at a hundred examples of a word or phrase and pick out some of the patterns which occur?
* As a detective, can you look at the context and co-text of examples from a corpus to see deeper meanings and typical uses of words and collocations?
* As an expert in disguise, can you dress-up your own writing, so as to blend into the norms and expectations of expert language users?

**Getting started**

The best place to get The Prime Machine HD (tPM) is from an official store. It is free!

|  |  |
| --- | --- |
|  | [Graphical user interface, text, application  Description automatically generated](https://play.google.com/store/apps/details?id=net.theprimemachine.tPM) |
| Qr code  Description automatically generated | Qr code  Description automatically generated |
|  |  |
| [Graphical user interface, text, application  Description automatically generated](https://www.microsoft.com/en-us/p/apps/9pf8710mxbh9?ocid=badge) | |
|  | |

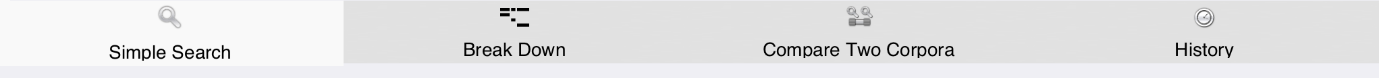
Windows and Android users can also download the App directly from the website: <https://www.theprimemachine.net>

The main ‘hamburger’ or 蒸笼menu in the top-right corner allows you to connect and change mode.

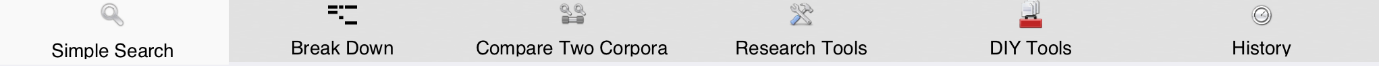
Graphical user interface, application

Description automatically generatedWhen you first use tPM, you almost certainly will want to connect to the server to access pre-prepared corpora and resources. There are three modes: Simple Mode, Full Mode and Game Mode. The Full Mode includes additional tabs and features for corpus research and DIY corpus work.

**Simple Mode**

****

**Full Mode**



**Game Mode**

Graphical user interface, text, application, chat or text message

Description automatically generatedGame Modeprovides a different user interface, giving you a list of groups of words to study before you *take a challenge* and try to guess which word matches a set of concordance lines. Two groups - **Academic synonyms** and **Academic prepositions** -are good places to start if you want to learn how collocations can help you choose between synonyms and find differences in the meaning of verbs when followed by different prepositions*.*

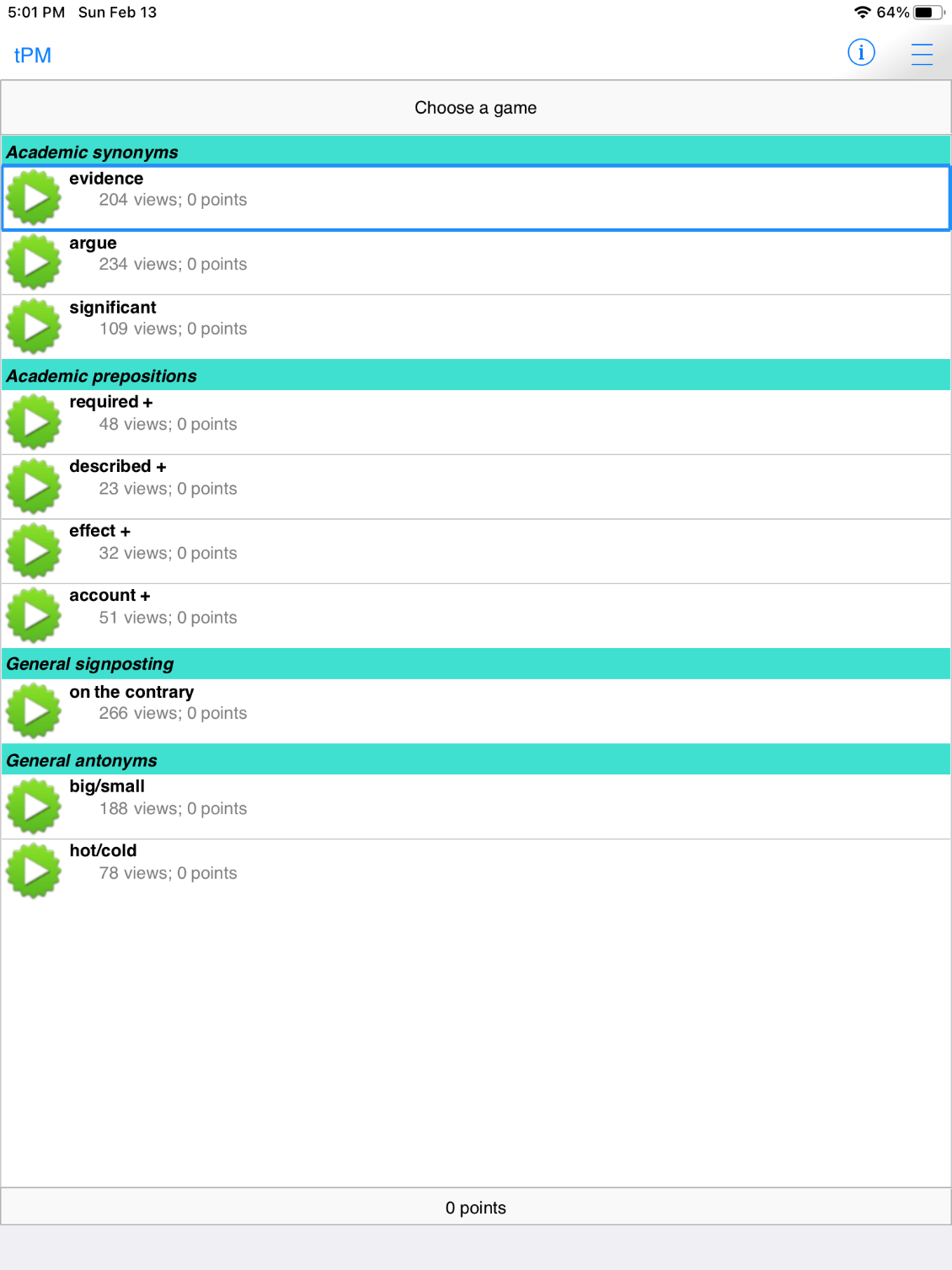
**Step 1: Learn how to read corpus data using tPM Game Mode**

**How to play**

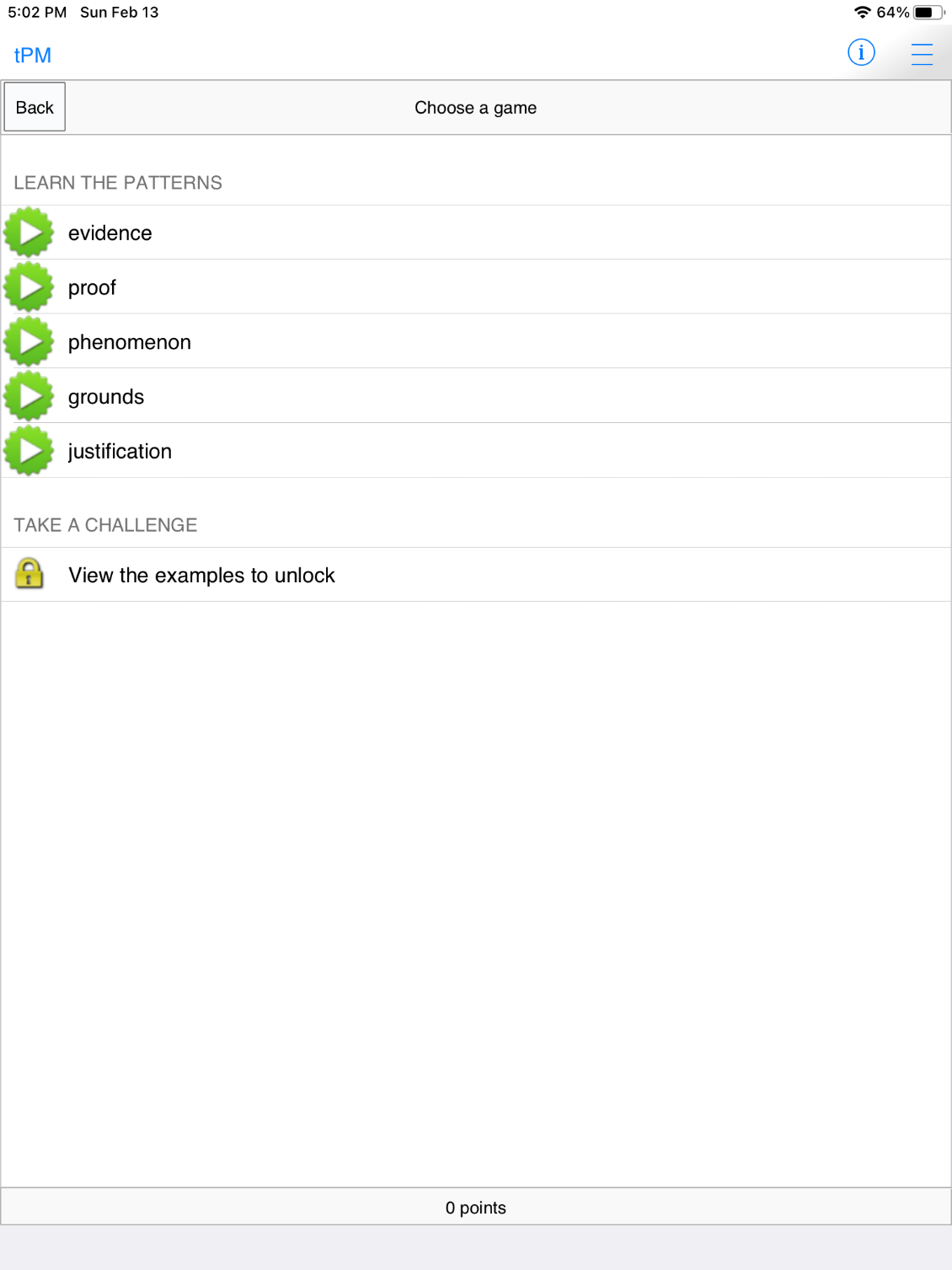
1. Choose Game Mode before connecting to the tPM Server. Then Connect.

2. The first time you try it, you will be shown the instructions for the game.

3. Choose a set of words from one of the games listed. In this worksheet, we are interested in academic synonyms and academic prepositions, so select one of the words from the first or second group.



4. When you have selected a game, you will see a submenu containing the words or combinations you have to learn. Open the items in the group one by one, looking at the concordance lines and thinking about the patterns you see. This is your chance to try to notice and remember the patterns of usage for each of the different words and combinations of verbs and prepositions.



5. When you select each word, 200 examples will be downloaded from the server. The displays for learning the words are the same as when you look the words up in Simple Mode.

You can sort the lines in different ways to help you notice patterns.

Table

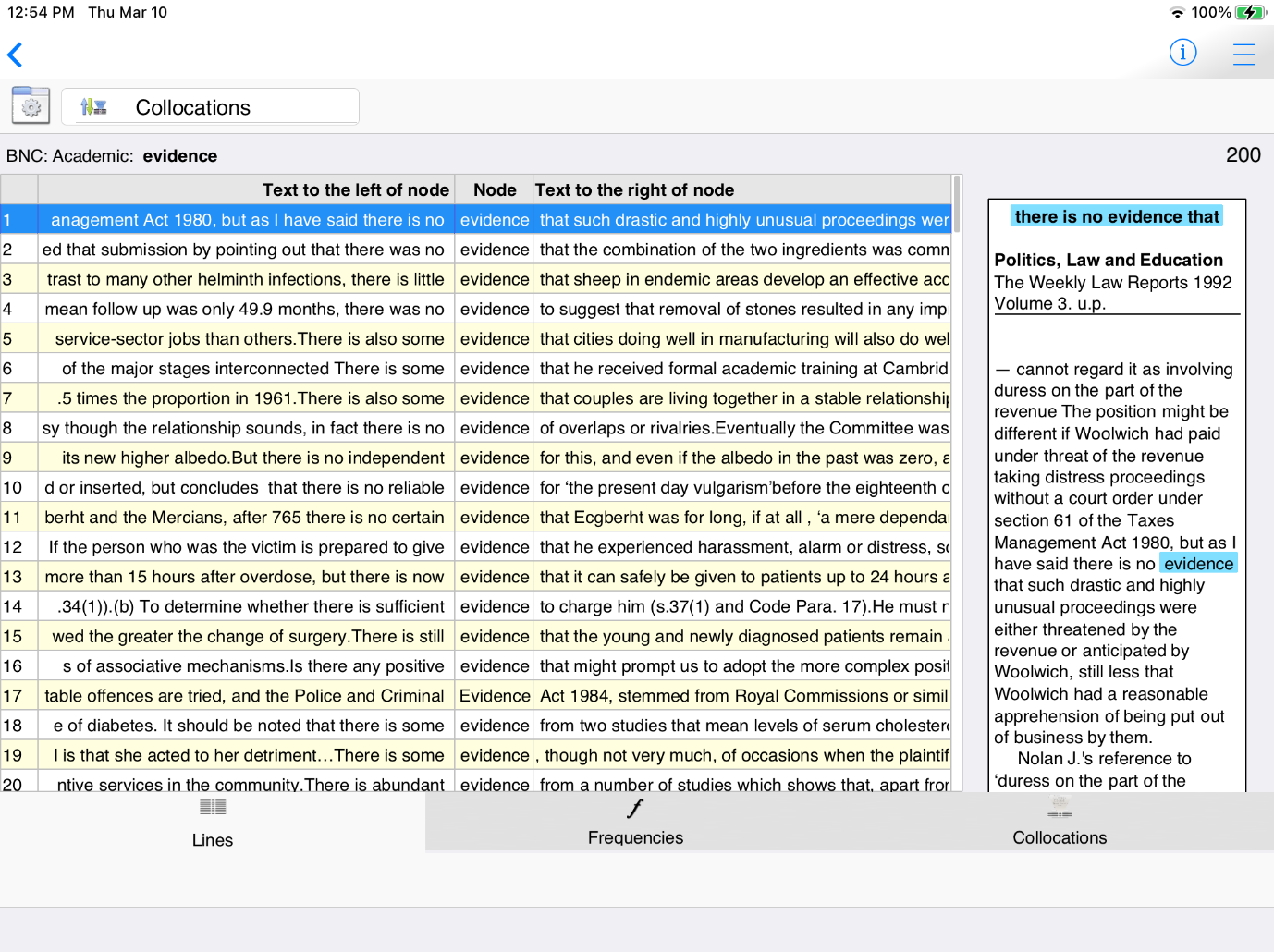
Description automatically generated

This menu button allows you to change the display settings for the current tab of results.

When learning words in Game Mode there are three kinds of data to view: Lines, Frequencies and Collocations.

This shows the number of examples already downloaded.

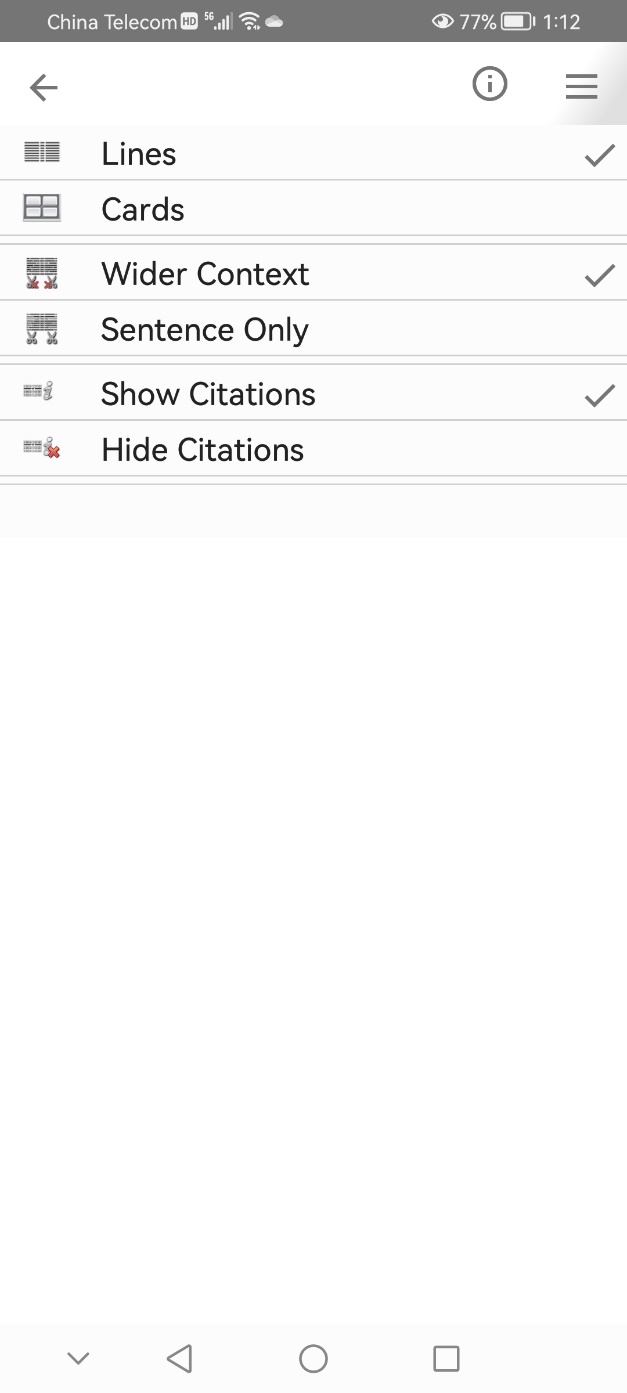
On a wider display (like this iPad), you can see a concordance card showing more information about the currently selected line.



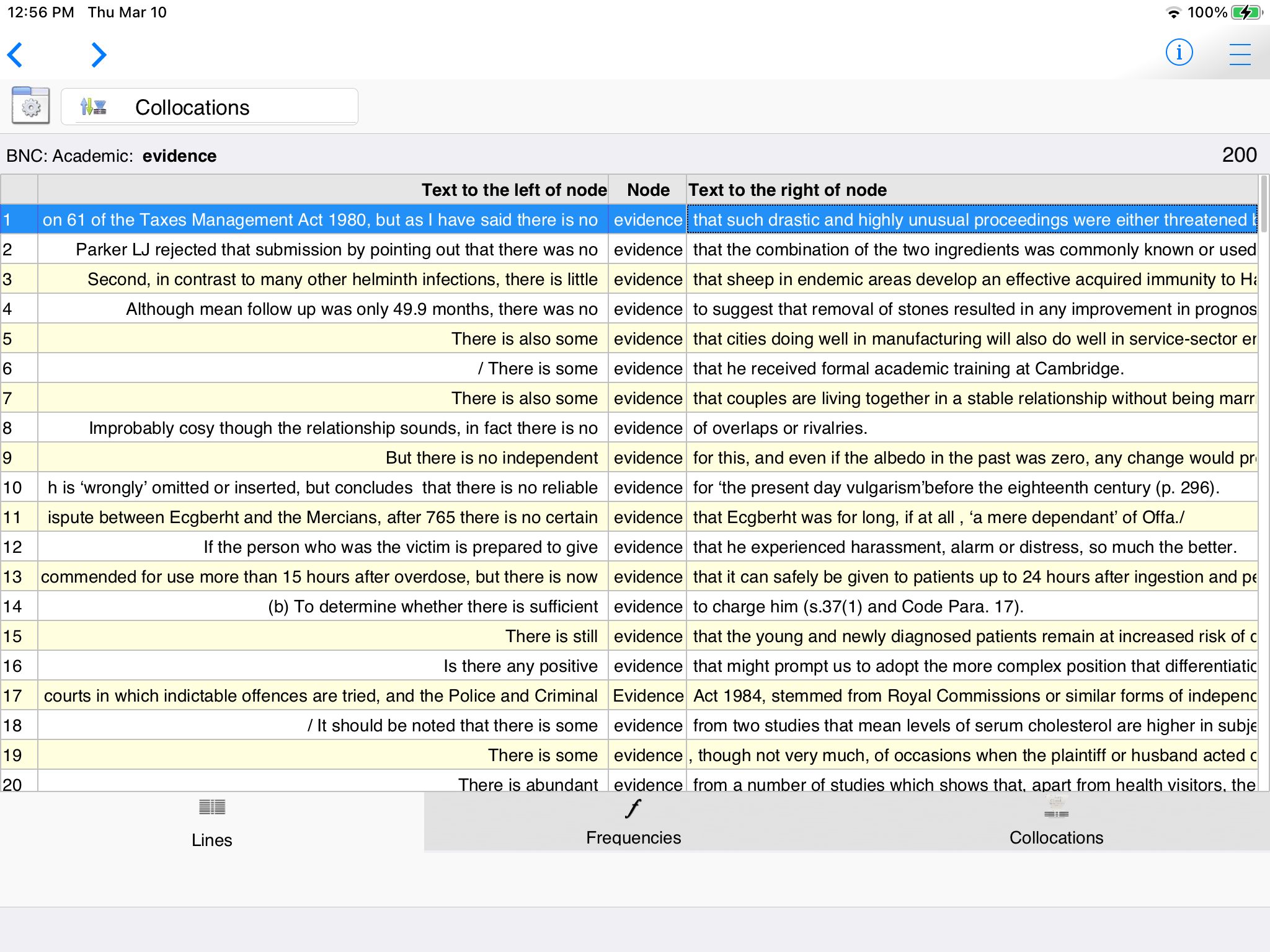
**Lines +**

On narrower displays (like this Android phone) you can view either lines or cards.

On all devices, switch between Lines + (if space allows), Lines and Cards using these settings.



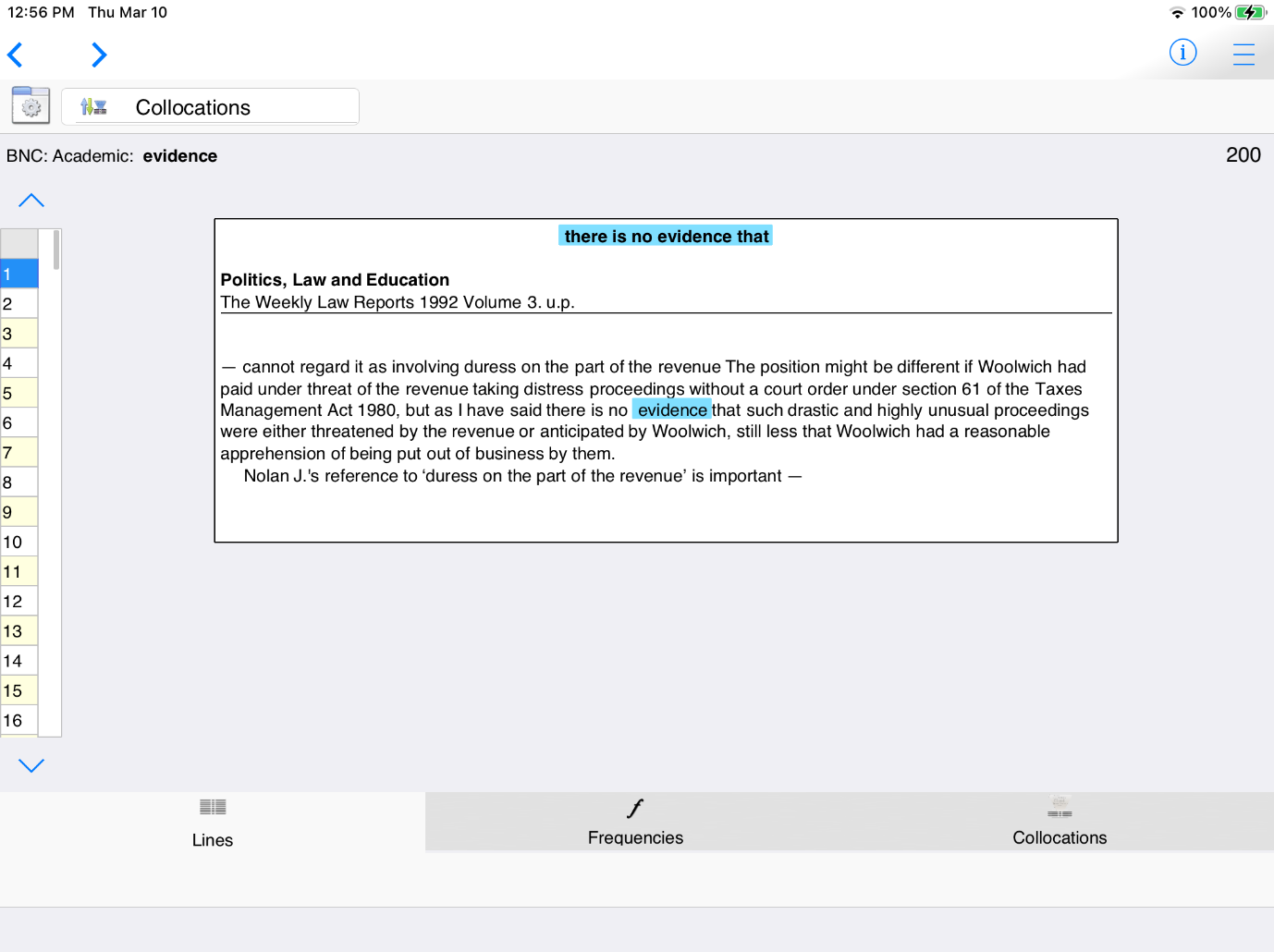
Other useful settings include just showing one sentence for each hit and hiding the citations.



When Sentence Only is selected, you can see if the word occurs near the beginning or end of sentences easily.

**Lines**

This shows the kind of text from which this example comes with reference information below.

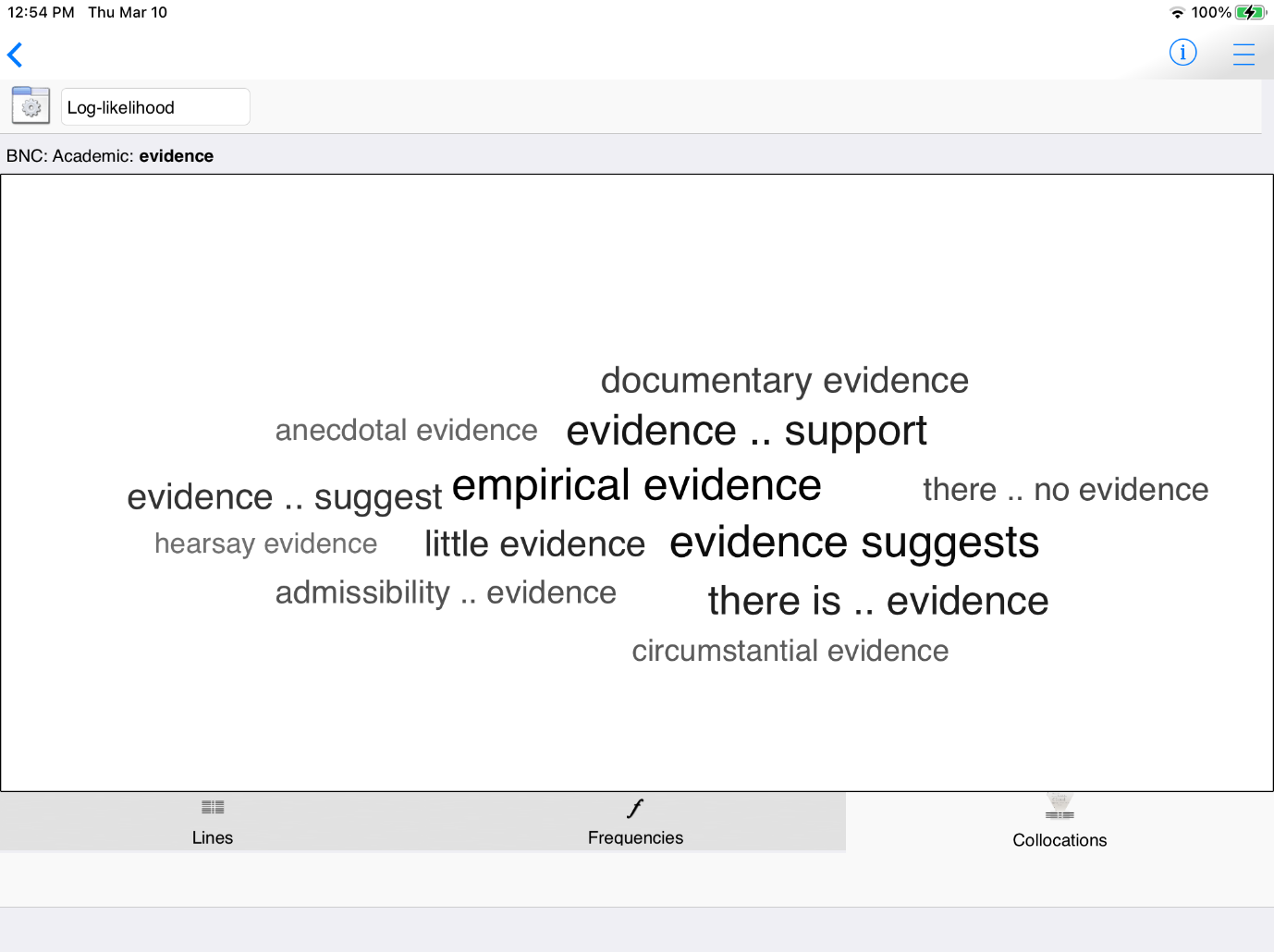


You can navigate through the examples using the sidebar.

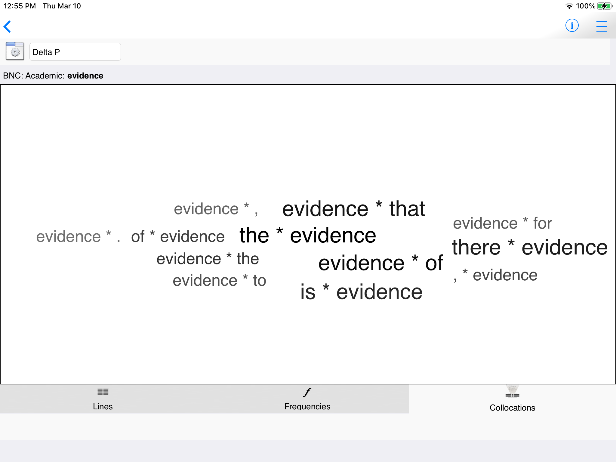
At the top of the card you can see strong collocations near the word of interest.

The card shows up to one sentence before and after the sentence containing the hit. Text is displayed with paragraphing.

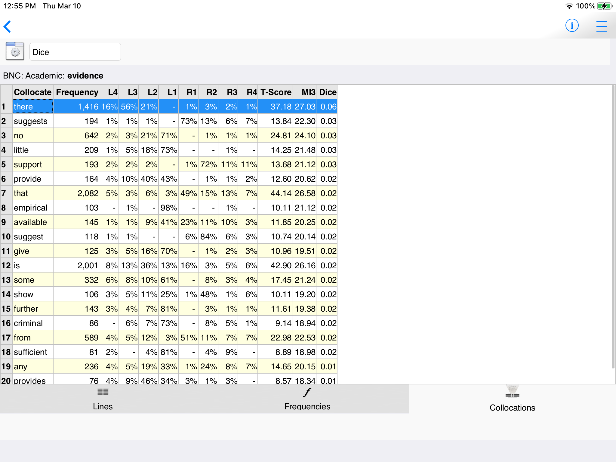
**Cards**



The Collocations Tab shows the word partnerships – the combinations of words which go together naturally in the corpus.

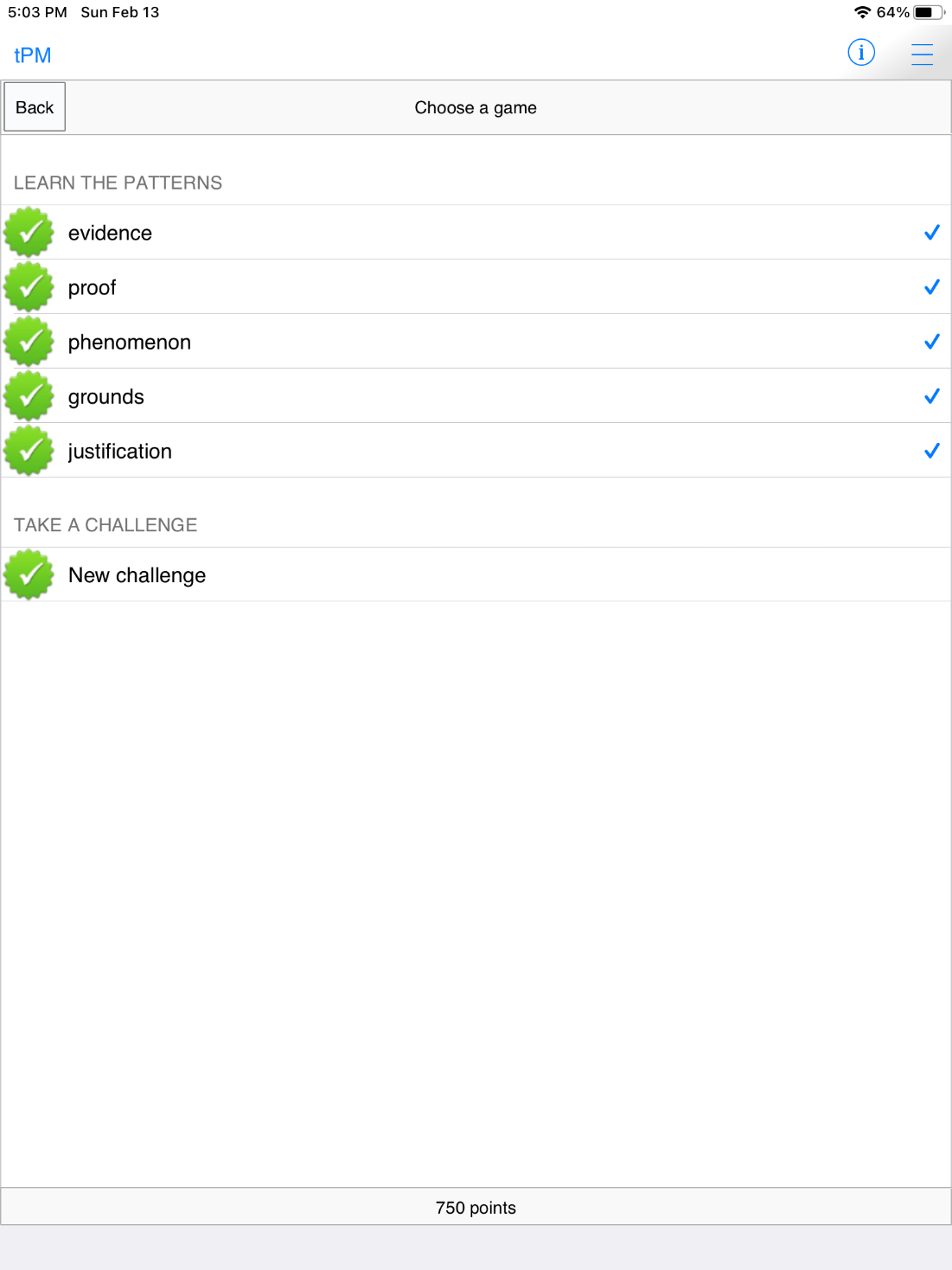
The menu next to the display settings button allows you to choose different statistics.



Different statistical measures can be used to create lists of collocations.

You can also view lists of collocations as tables.

6. When you have viewed concordance lines for all the words, **New challenge** will unlock.



7. Each time you take a new challenge, decide how confident you are that you will be able to guess each word correctly.

The more lines you have, the easier it will be to guess, but the fewer games you can play overall.

Graphical user interface

Description automatically generated with medium confidence

If you slide the confidence down to 1/5, you will get fewer points for a correct answer when this item is the correct answer, but you won’t lose so many points when you guess incorrectly.

If you slide the confidence up to 5/5, you will get more points for a correct answer, but face a heavier penalty if this is the missing word and you guess incorrectly.

Tips:

* Synonyms often have different collocations; if you learn the strong collocations, you should see some of these in the lines at the top (when the lines are sorted using Collocations – the default).
* Synonyms often have different grammatical patterns. Although they may all be the same word class, some nouns may be used with *the* or *this*, while others may be used with *a, an* or *some*. You may also find differences in use of modal verbs or prepositions, for example.
* For the academic prepositions games, it is better to try to explore the possible meanings and patterns of each combination of verb and preposition – while prepositions may have some general meaning, we often find that it is hard to generalize.
* If you find the challenge too hard, go back and read more examples of the words by selecting them from the **Learn** section. You can continue with the same challenge again later.

Graphical user interface, text, application

Description automatically generated

In landscape orientation, more co-text will be visible

Table

Description automatically generated with low confidenceTable

Description automatically generated

In portrait orientation, more lines will be visible and most collocations will still be in view.

Look for the strong collocations you’ve learned to the left and right of the gap.

8. Look at the set of concordance lines and guess the missing word.



When you are ready to guess, click or tap the Answer button.

**Step 2: Using search screen support and formulating your own queries**

After you have completed some of the games for learning about academic synonyms and academic prepositions, you should find that you have become more familiar with how to read and use concordance data. The next step is to make the most of the different ways to search for your own words and phrases.

There are four different search screens for you to use: Simple Search, Break Down, Compare Two Corpora and Advanced Search (under Research Tools on Full Mode).

**Simple Search**

This is the main search screen and it allows you to enter one or two search queries at a time. If you enter a word or phrase and tap the “hamburger” or 蒸笼menu inside the search box, suggestions will appear. You can see the number of examples in the currently selected corpus. You can also see collocations, other word forms and words with a similar meaning.

Frequency is important (and tells you the number of hits), but collocation strength is based on the statistical tests, trying to mirror the relationships between words of varying raw frequencies. The word *of* is very very frequent, but it will only show in a collocation if the statistical test shows its frequency in a combination with the search word is unexpectedly high.

The suggestions and the number of hits and the collocations will be different in different corpora. Click here to change to a different corpus.

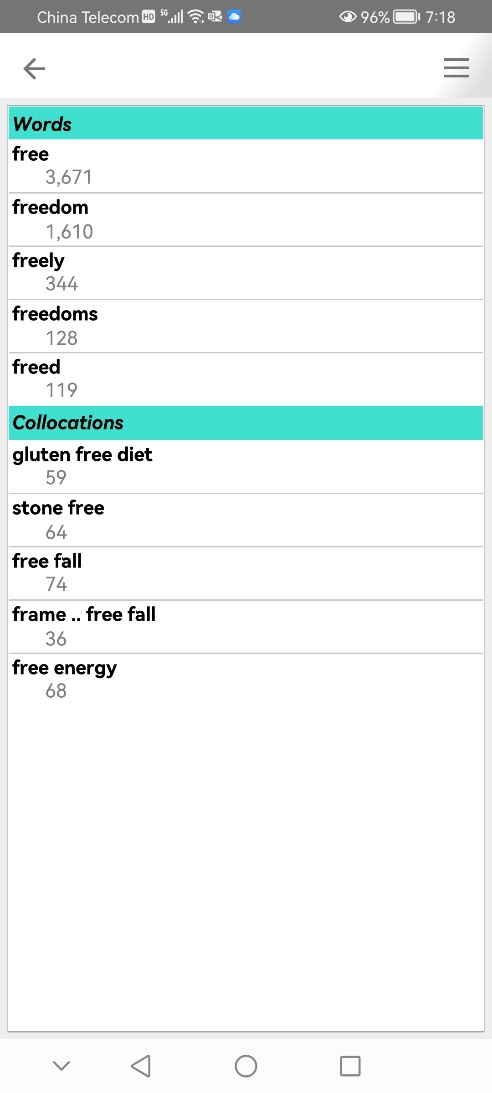
This button will give you more information about the currently selected corpus.

Graphical user interface

Description automatically generated

The left-hand side shows single word suggestions and strong collocations. The numbers are the number of hits in this corpus.

The right-hand side will show other word forms and words and collocations with similar meanings.



On narrower screens like this Android phone, the suggestions pop up on another screen. Select the item you want and then tap the back button.

Tips:

* When you enter one word, the suggestions will be for longer strings with the same letters (auto-complete) and strong collocations on the left; and other word forms and words with similar meanings on the right.
* The suggestions for similar meanings use several different resources including English-English resources and words which have the same English translation in a Chinese-English dictionary. You can only search for English words, but some suggestions are based on a Chinese-English dictionary.
* If you click or tap the hamburger/蒸笼 button again, you will be given more suggestions including more collocations.
* When you enter more than one word, tPM will try to find a match in the list of strong collocations in this corpus. If no match is found, it does not mean the combination does not exist.
* From the Simple Search screen, you can force the app to show concordance lines and frequency data for combinations of words not stored as a LL collocation in the database, by using three special symbols:
  + Putting \_ between words means these words must occur in order with no other words in-between.
  + Putting \* between words means these words must occur in order, but there could be zero or more words in-between (up to a total phrase length of 5 words).
  + Putting | between words means these words can occur in any order but all must occur within a 5 word window.
* More advanced searches can be performed using the Advanced Search tab.

When you are using a corpus tool, you will often need to perform several searches and to refine and retry different combinations. For example, if you look at the concordance line results for a noun you are interested in and you discover it has a strong collocation with a particular verb, you can return to the Simple Search tab and perform a new search for the two words in combination. Then up to 200 hits for the collocation will be displayed and you can explore the patterns of use for the collocation in more detail.

Text

Description automatically generatedIn BNC: Newspapers, the collocation cloud for *scheme* contains *scheme .. launched*.

This means that the verb *launched* has a strong statistical relationship with the noun *scheme*, taking into account the specific verb form and the word order.

Because there are over 2,000 hits for *scheme,* only a few of the concordance lines in the sample will contain the combination.

Graphical user interface, text, application

Description automatically generatedBut if you return to the Simple Search screen and look up the collocation, all 20 lines will be retrieved.

Graphical user interface, application

Description automatically generatedGraphical user interface, application

Description automatically generated

The interest in the combination *scheme ..*  *launched* may spark your interest in the verb *launched* – what other things are described as being *launched* in newspapers?

Returning to the Simple Search tab, you may want to explore the combinations of *launched* and also to explore the combinations of words with a similar meaning such as *established*, *published*, and *carried out.*

On the Break Down tab, we can enter a longer phrase or sentence and click the hamburger / 蒸笼 button inside the search box, and it will split the sentence into words, allowing us to look at combinations of up to five words starting at the currently selected item.

**Graphical user interface, application

Description automatically generated**

The second column shows combinations of words to the right of the selected word.

If the combination is stored as a LL collocation it shows the number of hits.

You could check concordance lines if you couldn’t guess “On the other hand” is better.

If you click the + button, suggestions may appear.

The first column shows the frequencies of each word in the sentence.

**Graphical user interface, application

Description automatically generated**

Perhaps *statistical techniques* would be better?

These two words don’t appear adjacent in this corpus, but they are nearby.

Choosing another word will show combinations with words to its right

On the Compare two Corpora tab, you can enter one word or a phrase and then retrieve concordance lines and other corpus data from two different corpora in one operation.

Results from the main corpus will be displayed on the left (A).

BNC: Newspapers is a useful corpus for finding collocations for social issues or topics for general EAP essays.

Graphical user interface

Description automatically generated

Results from the second corpus will be displayed on the right (B).

The suggestions will show the frequencies in both corpora, with collocation suggestions based on the strong patterns in the main corpus.

These buttons allow you to change the second corpus or view information about it.

On the Advanced Search tab (under Research Tools), you can perform a number of more advanced queries:

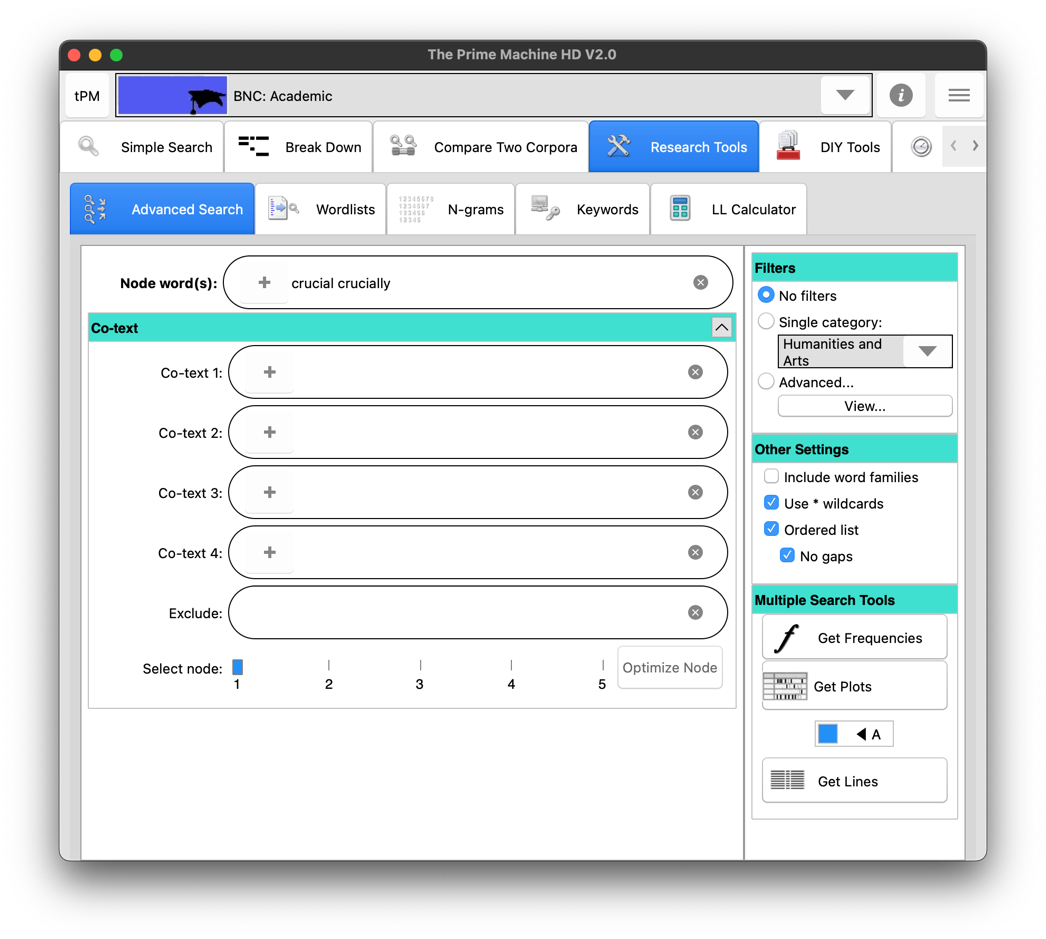
* Combining the results of more than one word form by entering multiple words separated by space, using wildcard pattern matching, or selecting the option to include word families;
* Forming patterns for phrases with multiple words in up to 5 slots, with options for the ordering of items and whether or not other words can occur in-between.
* Using filters to get results from one text category or advanced filters based on labels or features such as part-of-speech and position in text.

On the Advanced Search tab as well as getting concordance lines and displaying them on the left (A) or right (B) you can also get frequency tables and plots.

tPM has been designed to provide easy access to concordance lines and other summary statistics for strong LL collocations. While you can look at other combinations using the special \_, \* and | symbols, or use the Advanced Search tab to merge results for different word forms, you should be able to find many useful kinds of data by simply using the collocation searches on the Simple Search screen.

Results will come from the main corpus; click or tap here to change.

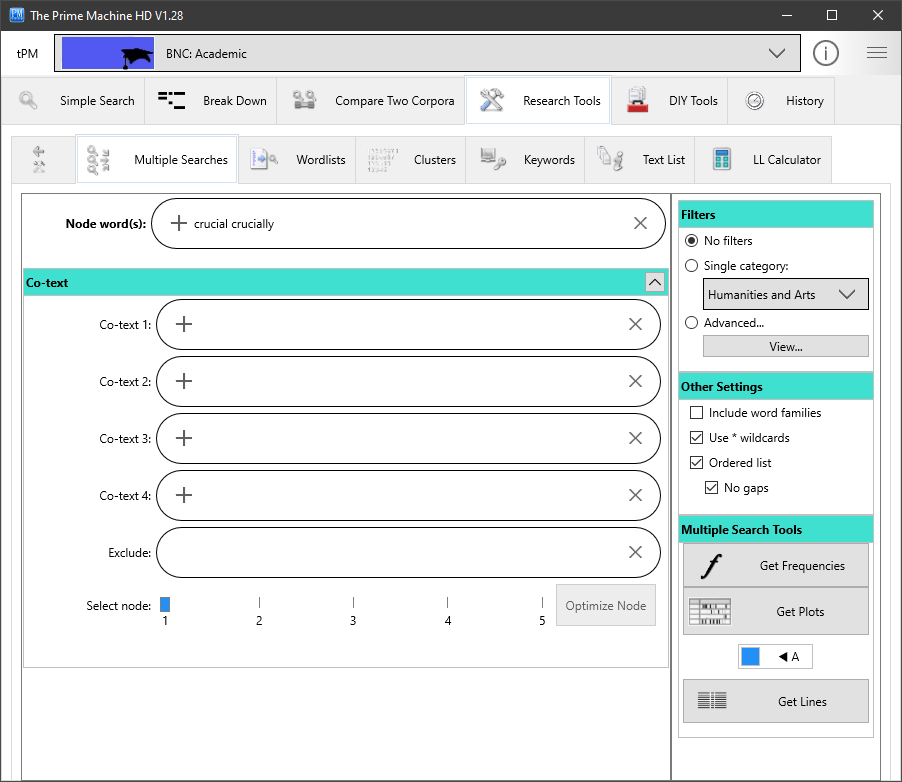
Some corpora are split into categories, so you can limit the search to specific groups of texts.



Enter lists of words separated by spaces to find any of these words and merge all the results.

If you enter one word and click or tap on the + button, other word forms for this word in the current corpus will be added to the box.

Clicking or tapping on the A/B button will send the results to the left (A) or right (B) so you can compare two sets of results side by side.

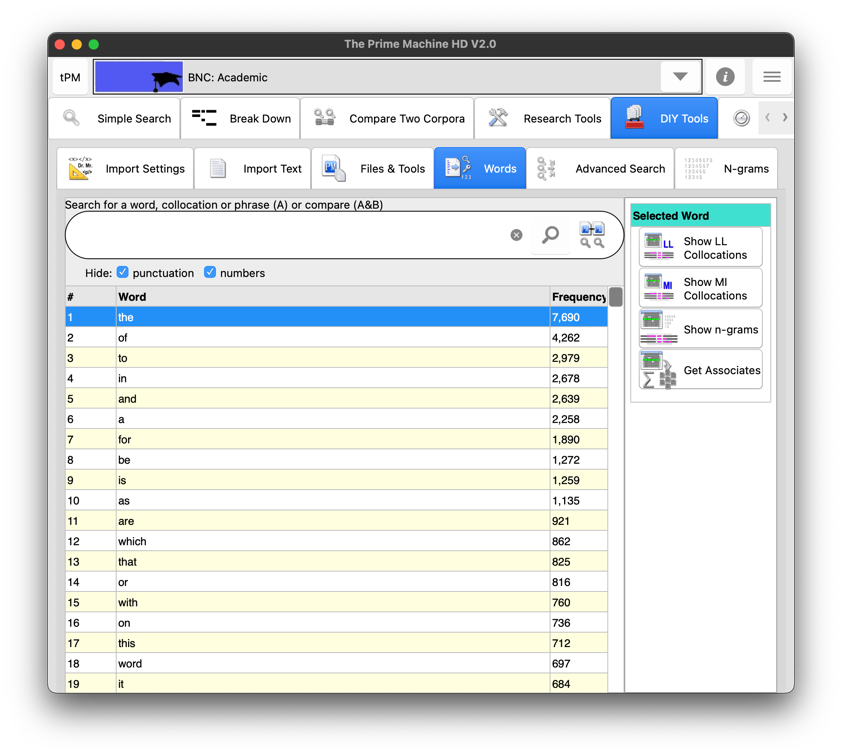
The other settings panel allows you to specify:

* Whether to include all the different word forms in the corpus for each word you have entered. For example, if you tick this option and enter the word *analysis* the results will include *analyses, analyse, analysing, analysed*, etc.
* Whether you want to use \* to represent zero or more letters. For example, *analy\** would include all words beginning with the letters analy.
* Whether the words entered in the co-text boxes need to occur in order or in any order and whether or not other words can occur in-between.

**Step 3: Building your own corpus, and viewing collocations of words of interest within it**

Building a corpus means gathering one or more texts and importing them into the corpus tool. For research, you would normally need to consider how you select your corpus texts and try to ensure your selection of texts is not made with unnecessary bias. But for a student or teacher wanting to explore collocations for a specific academic discipline, a good starting place would be to gather texts from the web pages, documents and articles used in a course. For step-by-step instructions on how to build your own corpus, see *The Prime Machine HD Guide I want to … build my own corpus.* You can build a corpus on any platform (Windows, macOS, iPad, iPhone and Android).

Once you have your corpus ready, you can select a word from the DIY word list or type it in, and use the buttons on the right to show different kinds of collocations, or use the magnifying glass button to retrieve concordance lines and view collocations on the Collocations tab.



or select a word from the list,

Type a word in this box…

These buttons will show tables of results for the selected word.

Click to get concordance lines; you can also compare hits in your DIY corpus with another corpus.

Graphical user interface, application

Description automatically generated

The Collocations tab for the results will be for your DIY corpus.

**Step 4: Exploring the collocational strength between words in your own sentences**

If you want to see which combinations of words in your own sentences have stronger or weaker collocational strengths, you can use the highlighting function in tPM. This will send your combinations of words and look up the collocation scores in the currently selected online corpus. The matches (or missing matches) found in the online corpus will then work as a baseline, showing which words have strong collocational relationships with their neighbours and which do not.

Remember, the collocational strengths of combinations are different in different corpora, so select a suitable corpus as a baseline before you start.

There are two ways to generate these results in tPM. For short strings of words or one sentence, you can use the Break Down tab.

Graphical user interface, application, Word

Description automatically generated

On the Break Down tab, type or paste the sentence into the search box and click or tap the MI & LL button. The combinations of words will be sent to the server and the results will appear as shown below.

Graphical user interface, text, application, email

Description automatically generated

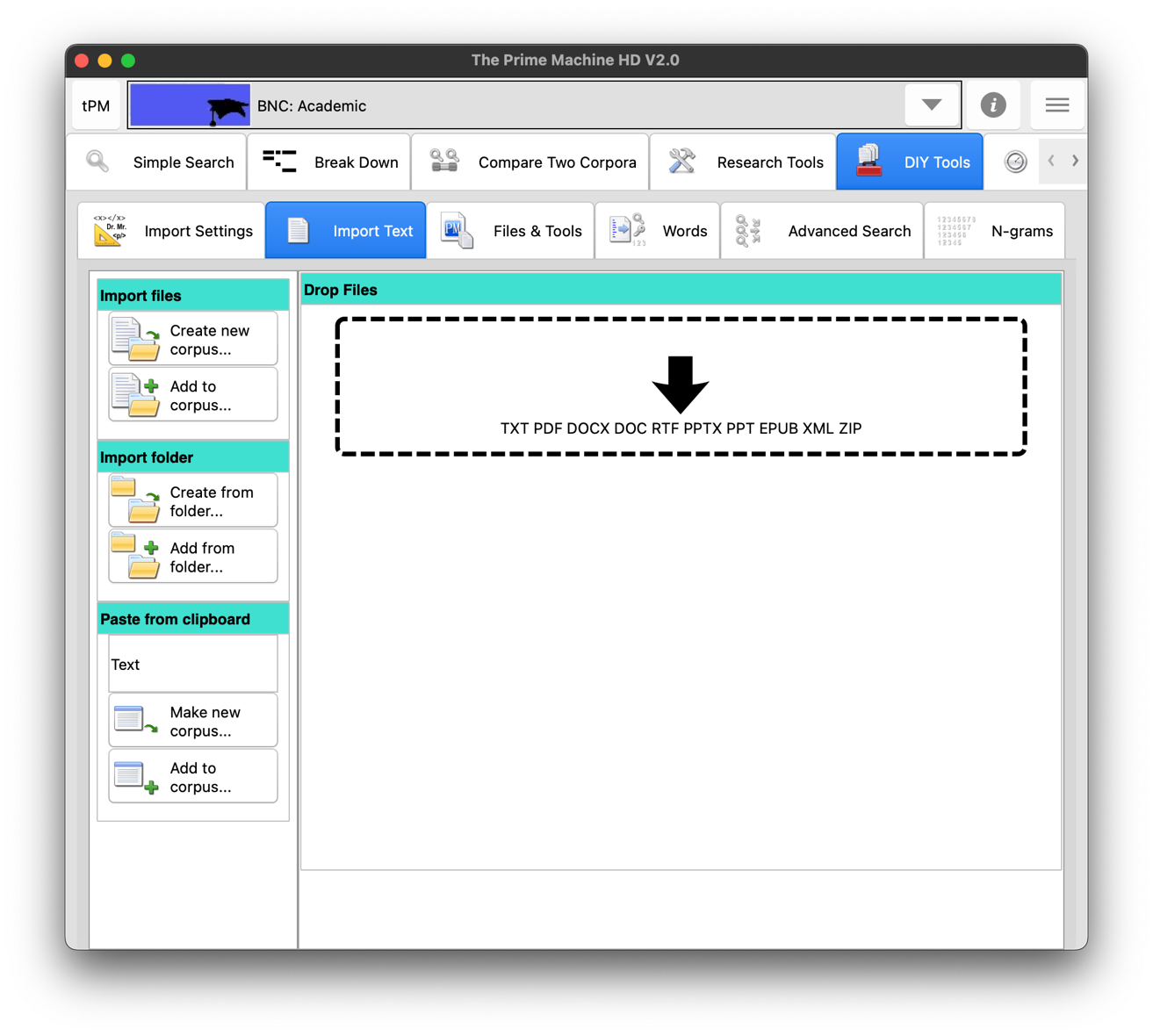
As well as viewing the results and noting down some combinations to explore later, the numbers in the columns for MI, DeltaP and LL have links, so you can see all the matching collocations and their scores.

In the table below, the LL score of 50 has been clicked, revealing a table of Log-likelihood collocations found in the BNC: Academic corpus which match combinations of words in our sentence.

Graphical user interface

Description automatically generated

The other way to get these scores is using one entire text from a DIY corpus. The “Paste from Clipboard” function is a quick way to make a small DIY corpus. You can set the name of the corpus before you paste text using the box above the button.



To generate collocation highlighting results for a DIY corpus text, go to the Files & Tools tab under DIY Tools and find the “Highlight MWUs” button.

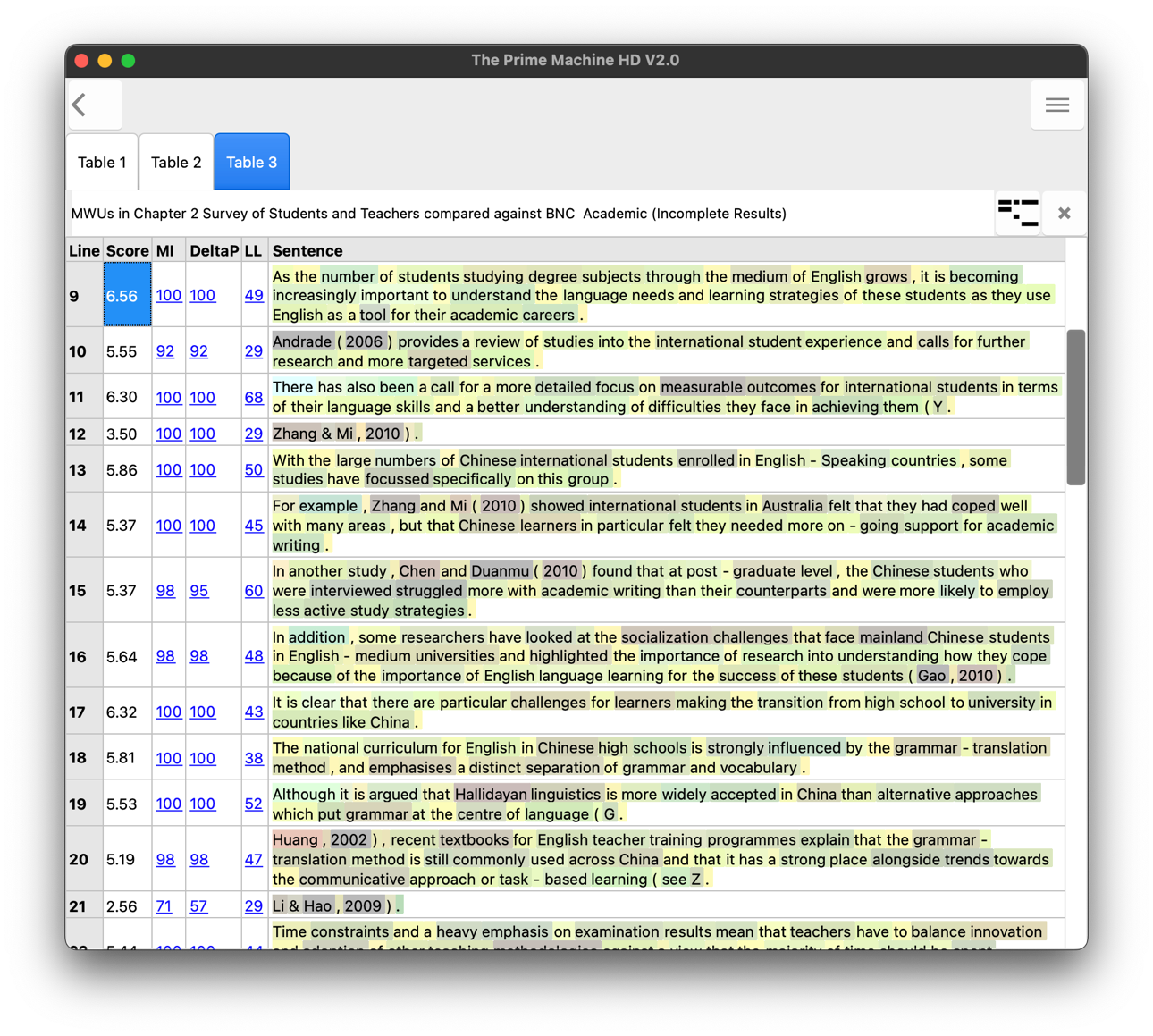
If you do want to send an entire text from a DIY corpus, you can navigate the texts on this tab by using the left and right arrows.

Graphical user interface, text, application

Description automatically generated

When you click the button, combinations of words will be sent to the server.

**Note: Using “Highlight MWUs” means chunks of your text will be transmitted over the internet. Efforts are made to make this secure, but you should be mindful of the risks of transmitting these kinds of data.**



This button will copy the selected line to the Break Down tab.

The results break the text into sentences and each sentence has a set of collocation scores. Unlike the Break Down tab, the full stop at the end of each sentence is also included in the scores (so we get slightly different results for the first sentence).

If we look at line 5 in the results above, we can see that the word *focussed* is in grey. This is actually because I have a personal preference for the less usual spelling of this word. It is correct, but *focused* with one *s* seems more common in the BNC: Academic. I can test this theory by changing the spelling and regenerating the results:

Graphical user interface, text, application, email

Description automatically generated

But in your own writing, you may be able to find some grey words that intrigue you – some combinations you thought were quite common and didn’t need any attention have been coloured grey by this process.

Try to spot some of these and break down the phrases into different combinations, trying to see whether there are synonyms or other wording which will make your expressions seem more natural. Don’t take away everything that is creative. But if an expression is grey and it is describing a common idea, you probably should try to change it.

If you are a teacher wanting to analyse your students’ work in this way, bear in mind that even in EAP classes the BNC: Academic may be to specialised for the kinds of general EAP topics which often come up in academic writing courses. It may be a good idea to use a more general corpus such as the complete British National Corpus.

tPM is a powerful tool for exploring collocations. Even by following just a few of these clues, you should be able to use your detective and reasoning skills to learn more about the contexts in which writers prefer one word combination over another, and the co-texts that are formed through collocation and grammatical patterns.

For some of the background to these methods and approaches, please see the tPM Help Selected Bibliography available from <https://www.theprimemachine.net/help.html>.

A picture containing outdoor, tree, sky, building

Description automatically generated

Qr code

Description automatically generated

Dr. Stephen Jeaco - 杰大海

[www.theprimemachine.net](http://www.theprimemachine.net)

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA. If you use this document, please include a link to the tPM Help Selected Bibliography available from <https://www.theprimemachine.net/help.html>.

First published: Thursday, 17 March 2022

Last updated: Saturday, 09 September 2023