

Getting Started

The Prime Machine & The Prime Machine HD

Thank you for choosing *The Prime Machine*. We hope that you will find the software useful. As well as providing examples we hope that through it you will discover interesting things about language and the way language works.

What is a corpus?

Essentially, a corpus is a collection of texts which have been chosen to provide examples of how words and phrases are used. The texts may be from written or spoken language. They may be complete texts, or extracts from longer documents. The plural form of *Corpus* is *Corpora*.

What is a concordancer?

A concordancer is a computer program which gathers together instances of particular words or phrases so a student or researcher can see all the ways in which they are used across the whole corpus. It works in a similar way to a search engine or the Find function of a word-processor, but the aim is to provide real examples of how language is used.

What can I use The Prime Machine for?

The Prime Machine was initially developed with language learners and teachers in mind. You can use it to find examples of **real language use** (also called **naturally occurring language**). It has also been further developed to offer a good range of tools for English majors and students studying linguistics or TESOL to use for corpus research projects. The patterns and summary information of features of Lexical Priming can be used to compare specific instances of English language use (perhaps an expression in an English learner text or a test item, or creative uses of language in speeches, articles or literary extracts) against patterns representing a norm in one of the ready-made online corpora. For example, a combination of words found in a single text outside the corpus can be compared with patterns of collocation and co-text of these words in a corpus of similar text types. The “art” of a sentence from a novel not in the corpus can be explored to see how some patterns in the sentence follow conventions found in similar text types, while other features of the patterns deviate from norms. Projects may also involve the construction of new corpora. Typical DIY corpus projects may focus on the exploration of differences between genres/registers, the exploration of different authors’ styles as well as differences in English translations (corpus stylistics) or the exploration of vocabulary and language patterns for English language teaching materials for a specific subject discipline.

It should not replace other forms of language support, but can complement a variety of resources. The table below shows some differences between information provided by various other sources and the kind of information you can get from *The Prime Machine*.

| Resource | <i>The Prime Machine</i> |
|----------------------------|---|
| Dictionaries | <ul style="list-style-type: none"> • Many more examples; • Examples from specialist fields; • Information about typical contexts (primings, labels and associates). |
| Search Engines | <ul style="list-style-type: none"> • Results from texts selected for language analysis not just websites; • Information about how words often go together (collocation); • Results are not ranked for commercial purposes. |
| Vocabulary & grammar books | <ul style="list-style-type: none"> • Insights into actual use; • Variation and typicality; • Differences across genres and text types. |

Some possible uses might be:

- Looking up words to find useful phrases or collocations;
- Comparing the use of two or more synonyms;
- Seeing how words are used differently in different genres or text types;
- Finding examples from other texts to compare the choice of language with your own.

System Requirements

The Prime Machine is a client-server application, meaning that you will need to have internet access in order to use it. It is a Windows application, and will run on Windows Vista and above. It has been tested with Windows 7 and Windows 10.

The Prime Machine HD is a client-server application, meaning that you will need to have internet access in order to use it. There are packages available for MacOS, Windows 10, iOS and Android.

Installation

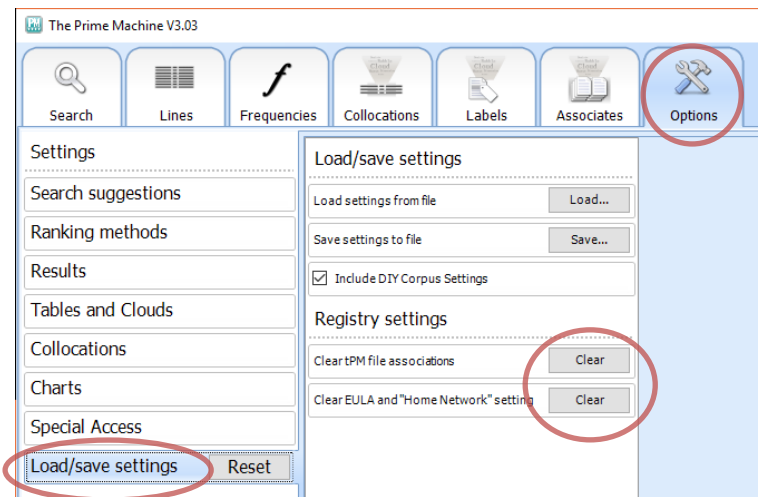
Windows 32 bit ZIP download - The application file is a single .EXE executable file. If you are copying the application from a ZIP file, you should unzip it before running it.

Simply place **the application file** on the desktop (or elsewhere on your system) and double-click it to run.

Uninstalling

The Prime Machine does not store very much data in the system registry, but if you want to completely clear this, you can go to the Options Tab, choose the Load/save settings page and click **Clear** next to “Clear tPM file associations” and “Clear Home Network Setting”.

You can simply delete the application’s .EXE file.



Basic features

This section will introduce how to use *The Prime Machine* to get examples and how to understand the purpose of some of the main features.

Opening the Application

When you first use The Prime Machine, you almost certainly will want to connect to the server to access pre-prepared corpora and resources.

Students and Staff at XJTLU connecting from campus or using XJTLU’s VPN service should remember to tick the “Try tPM Home Network” box in order access additional resources.

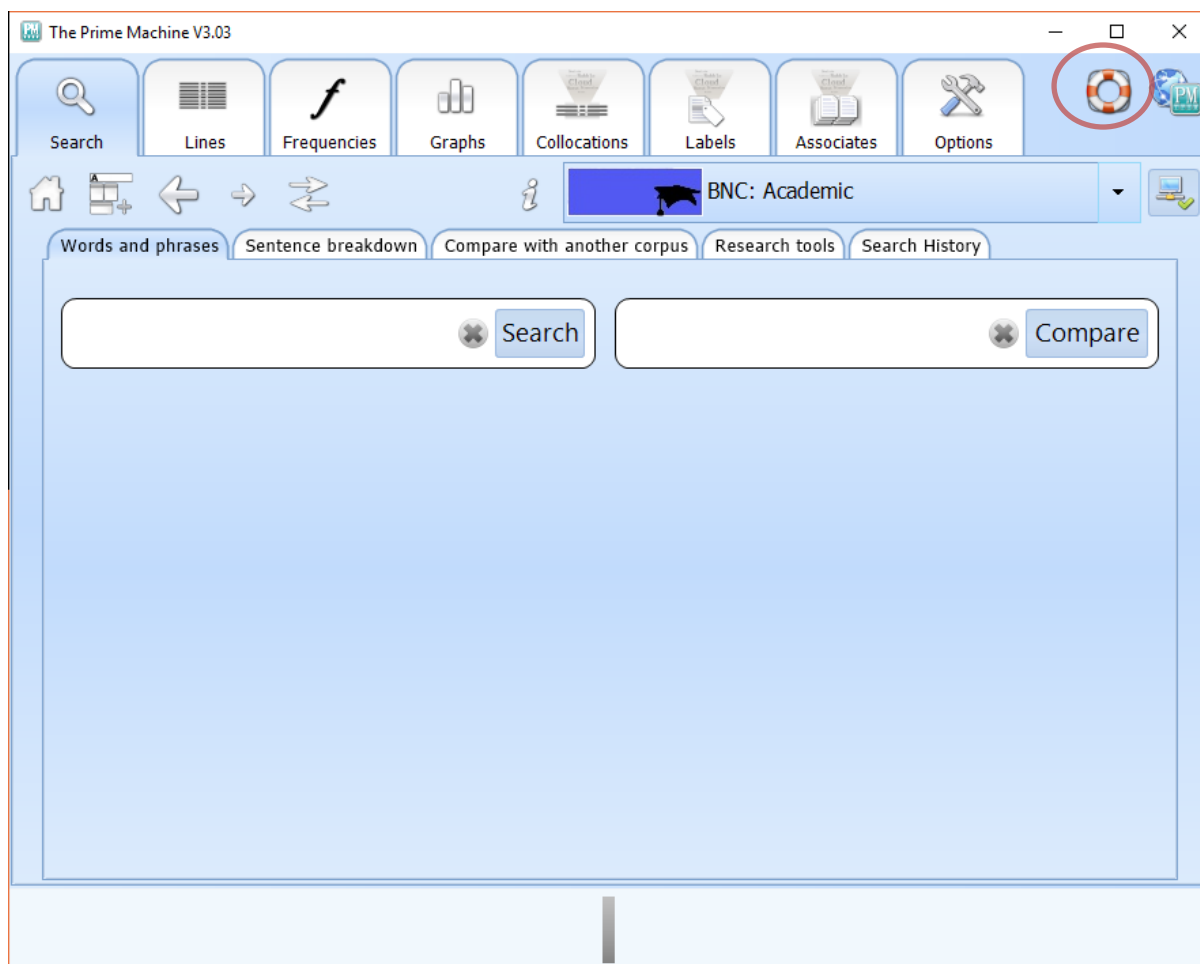


Click **Connect** and the application will connect to the server and retrieve information about the corpora which are available. (If you have already saved results or you want to create DIY corpora offline, you can click **Work Offline**.)

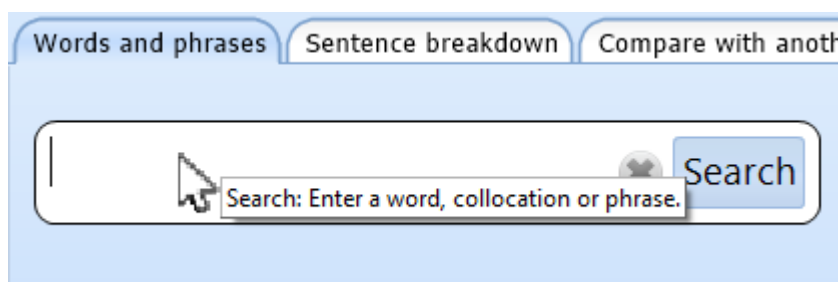
Getting around

After you have connected to the server, the main screen will appear. Across the top of the screen are a number of **tabs**. **These allow you to look at different kinds of information about the words and phrases.**



The life-ring button provides information about the current tab.









If you are not sure what a button or box is for, you can often get a **hint message** by pointing at it with your mouse.



Here is a list of the main tabs and there purposes:

| | |
|---|--|
|  <p>Search</p> | <p>Search Tab</p> <p>Enter a word, collocation or phrase to request results from the server.</p> |
|  <p>Lines</p> | <p>Lines or Cards Tab*</p> <p>The results are the same as for the Cards Tab, but appear with one result for each line of a table. In other software programs, these are called KWIC (Key Word in Context). The currently selected concordance line is also show as a card (see Concordance lines below).</p> |

| | |
|---|--|
|  Frequencies | Frequencies Tab* The Frequencies Tab helps you see whether the words and collocations you are looking up are very common or rather rare. When looking at results from two corpora of different sizes the frequencies per million words will help you compare. |
|  Graphs | Graphs Tab* The Graphs Tab allows you to see information about typical contexts for the words you have looked up. These include aspects of position (headings, text position, etc.), grammatical structures (modals, voice, articles, etc.) and repetition. As well as seeing graphs showing proportions for the current set of results, you can also see the overall proportions for all the results stored in whole corpus. |
|  Collocations | Collocations Tab* The Collocations Tab shows words which are often used together, based on several statistical measures. |
|  Labels | Labels Tab* The Labels Tab provides information about the labels which seem to be important for the words you have looked up. Results include tags for the whole text, specific sections containing the words and information about the writer or speaker. |
|  Associates | Associates Tab* The Associates Tab shows the most common text categories for the words you have looked up. It also shows other words which are frequently key words in the texts where the words which you have looked up are also key. |
|  Options | Options A few options appear at the top of different tabs, but to save or load settings and to change advanced settings, you will need to use the Options Tab. |

Note: Tabs marked with a star (*) only show information after a query has been made and results have been downloaded from the server.

The Home Button

The **Search Tab** should be fairly clear, but there are some sub-pages and a range of different corpora to choose from. If you get lost, click the **Home Button** to clear the search boxes and return to the default corpus.



If you get lost, click the Home Button to clear the search boxes and return to the default corpus.

Simple searches

You can enter between 1 and 5 words in each search box. As you start to type a word, the program will show a list of **auto-complete** suggestions.

If you can see the word you want in the list, just click on it.

The screenshot shows a search interface with three tabs: "Words and phrases", "Sentence breakdown", and "Compare with another corpus". The search input field contains the text "cons" and a "Search" button. Below the input field, a list of suggestions is displayed:

- cons (25)
- considered (4437)
- consider (3636)
- considerable (3031)
- consent (2050)
- consideration (2017)

As soon as you have entered a string of letters which matches a complete word in the corpus, further suggestions will appear, showing:

- Longer words which begin with the same string of letters;
- Common **collocations** containing the word nearest the cursor;
- **Other word forms**
- Words with a **similar meaning** or **similar translation**

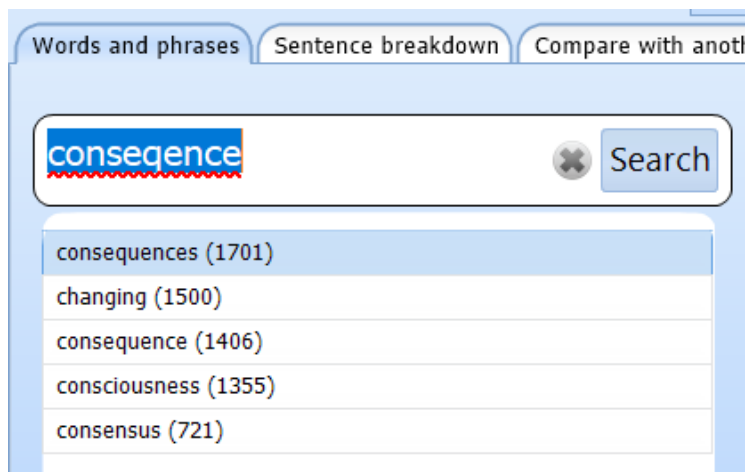
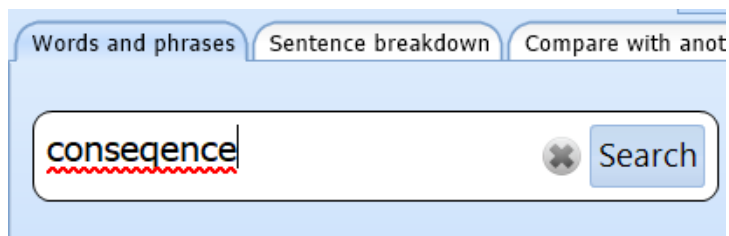
The screenshot shows a search interface with five tabs: "Words and phrases", "Sentence breakdown", "Compare with another corpus", "Research tools", and "Search History". The search input field contains the text "consequence" and a "Search" button. Below the input field, a list of suggestions is displayed:

- consequence (1406)
- consequences (1701)
- as a consequence (337)
- consequence of (866)
- inevitable consequence (24)
- direct consequence (33)
- a consequence of (388)

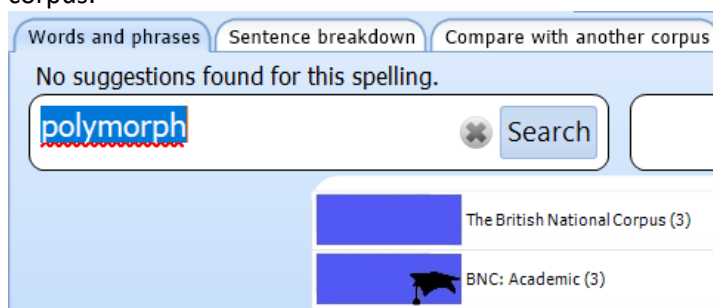
To the right of the search input field, there is a "Compare" button and a search input field. Below the "Compare" button, a list of suggestions is displayed:

- consequences (1701)
- sufferings (35)
- importance (3623)
- upshot (42)
- aftermath (139)
- event (2364)

If you enter a string of letters which cannot be found in the currently selected corpus, a red error line will appear. This does not mean that the word is misspelt, but rather that no results can be found. On the other hand, if you do make a **spelling mistake**, you can right-click on the word to get spelling suggestions.



When you right-click on a word marked with an error line, the program will also check all the other corpora to see if they contain the word. For example, if you look for the word *polymorph* in the **Spoken** section of the **British National Corpus**, it cannot be found and will be marked with an error. However, this **technical word** does occur in a number of academic texts, and its frequency in each will be displayed. If you double-click on one of the corpora listed, the program will start using that corpus.



What is a word?

It may seem like a simple question, but there are a few important points to understand regarding the way in which **search strings** are processed in *The Prime Machine*.

A search is based on a string of letters. This means:

- different word forms will not be grouped together in the results;
You will get different results if you look for *cause* compared with *causes*, and these will each be different from *causing*, *caused*, etc.
- different words which have the same spelling will be mixed together;
If you look for *bank* you may find examples of *river bank* mixed together with *financial bank*.

Capital letters and lower case letters are treated in the same way. This means results for *china* and *China* will be mixed together.

The only time capital letters will appear in suggestion boxes is if letters in the middle of the word are capitalised (e.g. PhD).

If you want to look at two word forms at the same time, you can use the **Compare** button (see below).

Searching for collocations and phrases

When you enter more than one word in the search box, it is important to understand the difference between **collocations** and other combinations of words. If it is stored as a collocation, the database also stores **priming summaries**.

The Prime Machine uses a statistical formula to determine whether or not to store priming information for a collocation.

If you type in or choose a phrase which meets this requirement, you will be able to see priming graphs, tags and associates for the collocation as well as just concordance lines.

For collocations, the order of the words is important. You should enter the words in the order in which they are to be used, and use two dots between words if gaps are permitted. For example:

- cause difficulties
- root cause
- cause .. action
- likely .. cause

The Prime Machine also stores collocations of three, four and five words in a row.

If the combination of words you enter in the search box can be found in the database as a collocation, the query will be made and the results will be downloaded.

However, if the phrase is not stored as a collocation, the program will check to see whether the combination occurs:

- all in a row (e.g. *President of the United States*)

The _ symbol means no other words can occur between each word and they must be in order.

president _ of _ the _ united _ states

- in order, but with gaps (e.g. *President _____ United States*)

The * symbol means other words may occur between the words, but they must be in the same order and the first word must be no more than 4 words away from the last.

president * united * states

- in any order within a 5 word window (e.g. *United President States*)

The | symbol means the words may occur in any order, but the first word must be no more than 4 words away from the last.

united | president | states

These search queries are quite complicated, but if you enter a phrase which is not stored as a collocation, suggestions for these will be given. If you click on the **Check other corpora** button, the program will check all the other corpora to see whether the phrase occurs as a collocation there.

Words and phrases Sentence breakdown Compare with another

No priming summary available. Try with |, _ or *

president united states

president .. united states (9)
president * united * states (>0)
president | united | states (>0)



When you use English you must pay attention to word forms.

When you use The Prime Machine, you should also be careful to use inflected forms - you should type in the form of the words as they will appear in your sentences.

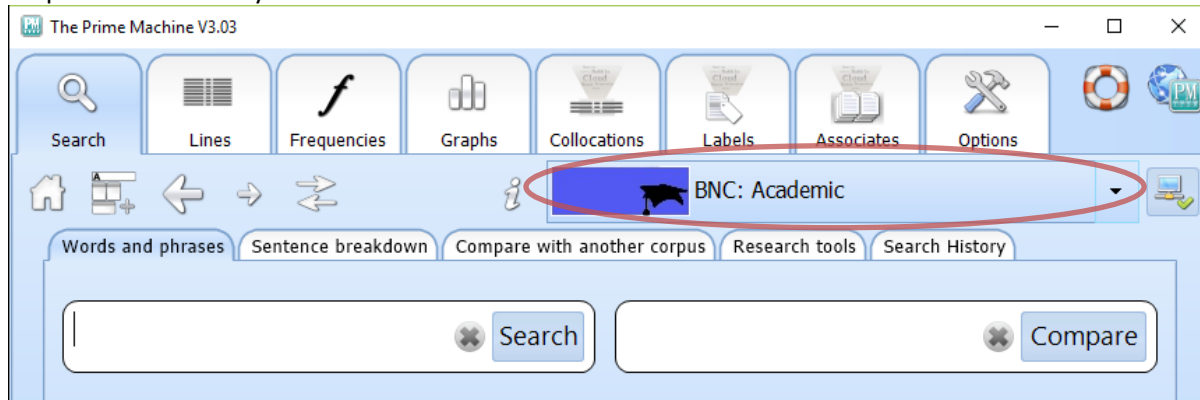


*If you are looking up a phrase and you see "**No priming summary available**", it could be because the phrase is common but the words in the phrase are often used in other phrases too. Check the suggestions box for how to search using |, _ or * so you can still see any concordance lines for the phrase even if it isn't stored as a collocation.*

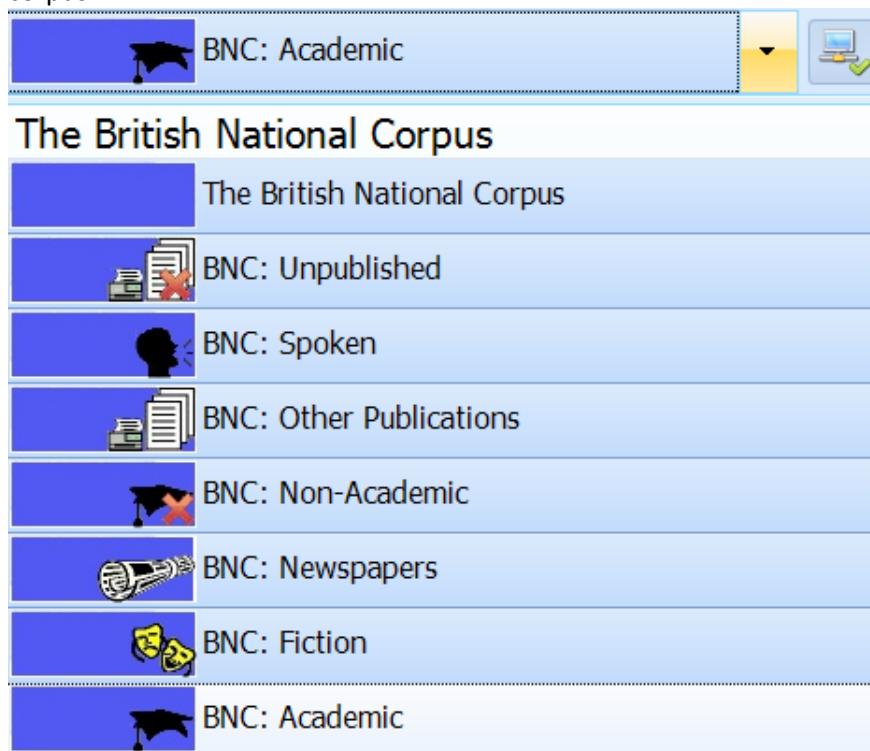
Choosing a corpus

For academic work, the BNC: Academic corpus will be a good choice. This is the default corpus and will already be selected each time you open the program.









If you click on the drop-down menu, you will be able to see a choice of other corpora which are available. If you are connected on The Prime Machine's Home Network (XJTLU), you will see more corpora here than if you connect from off-site without XJTLU's VPN.



Each section of the British National Corpus is available separately, but you can also search the whole corpus.



Other options may include specialist corpora, like those from the *Hindawi Academic Journals* shown below.

| Hindawi Academic Corpora | |
|---|---------------------------------------|
|  | Hindawi Mathematics |
|  | Hindawi Chemistry |
|  | Hindawi Physics |
|  | Hindawi Earth Science and Environment |
|  | Hindawi Engineering |
|  | Hindawi Computer Science |
|  | Hindawi Social Sciences |
|  | Hindawi Biological Sciences |

Each corpus has its own collocations and other statistics, so try to match the purpose of your search to a suitable corpus.

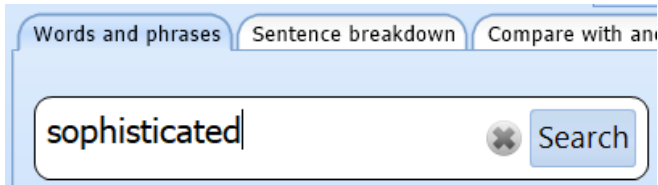


For more information about the latest corpora available check **tPM Help 003a** and **003b** from help.theprimemachine.com.

Concordance lines (LINES/CARDS tab)

The most important results are found on the Lines or Cards Tab. These are two ways to look at exactly the same set of results.

Type a query into the search box and click on **Search**. It will take a few moments for the concordance lines and other data to be downloaded from the **server**.



When it is ready, the program will jump to the Lines or Cards Tab. From there, you can switch between viewing the concordance lines as Lines or Cards. You can choose a different default view results on the Options Tab).

The Prime Machine V3.03

Search Lines Frequencies Graphs Collocations Labels Associates Options

BNC: Academic: **sophisticated** 200

| | Text to the left of node | Node | Text to the right of node |
|----|---|----------------------|---|
| 1 | stances amounts to a misrepresentation of the more | sophisticated | falsificationist's position. More than a hint of this is co |
| 2 | 19)/Dearlove (1979) seeks to develop a rather more | sophisticated | analysis in his discussion of reorganization, but his co |
| 3 | of industrial land markets in comparison with the more | sophisticated | understanding of labour and capital markets./ Prelimi |
| 4 | is century by Robert Walcott, who a | sophisticated | structure u |
| 5 | t process of accumulation. This is c | sophisticated | cally invali |
| 6 | from the decades of dedicated wor | sophisticated | lop no furt |
| 7 | ot be cost-effective to pursue with | sophisticated | ng upon th |
| 8 | :state's role has led some Marxists | sophisticated | e specialist |
| 9 | omes more experienced, her/his ne | sophisticated | |
| 10 | ./ Technical Documentation: As aircraft become more | sophisticated | integrating many technologies, so the proliferation of |
| 11 | Diagnosis/Isolation: Aircraft have become ever more | sophisticated | , integrating many technologies within a single airfram |
| 12 | gh the tasks to be performed may have become more | sophisticated | ./ A third, early, contribution to our understanding ab |
| 13 | at the environment, as market demands become more | sophisticated | so there is a need for oper |
| 14 | ion of 'rational choice' introduces us to another, more | sophisticated | conception of the rational |
| 15 | that of 'ordinary prudent men', need to have a more | sophisticated | grasp of the factors that s |
| 16 | this respect, his pluralism can be regarded as a more | sophisticated | version of monism. The fla |
| 17 | ing a Marxist, has to provide an alternative and more | sophisticated | account of why military exp |
| 18 | mentioned earlier, the HMI's took an altogether more | sophisticated | view of the relationship of education to society and v |
| 19 | ill see later, subsequent experiments have used more | sophisticated | measures of recall.) Knowing or not knowing the title |

The word you looked for is centred in the table, with each example in one row in the table

If you choose Lines + View, the card for the currently selected line appears on the right.

more sophisticated falsificationist
Humanities and Arts
What is this thing called science? Chalmers, A F
Open University Press
Milton Keynes 1992 1-120
... The self-avowed falsificationist Popper says as much in the passage quoted on p.44, where the italics are his. However, exclusive attention to falsifying instances amounts to a misrepresenta

The Prime Machine V3.03

Search Lines Frequen

Collocations

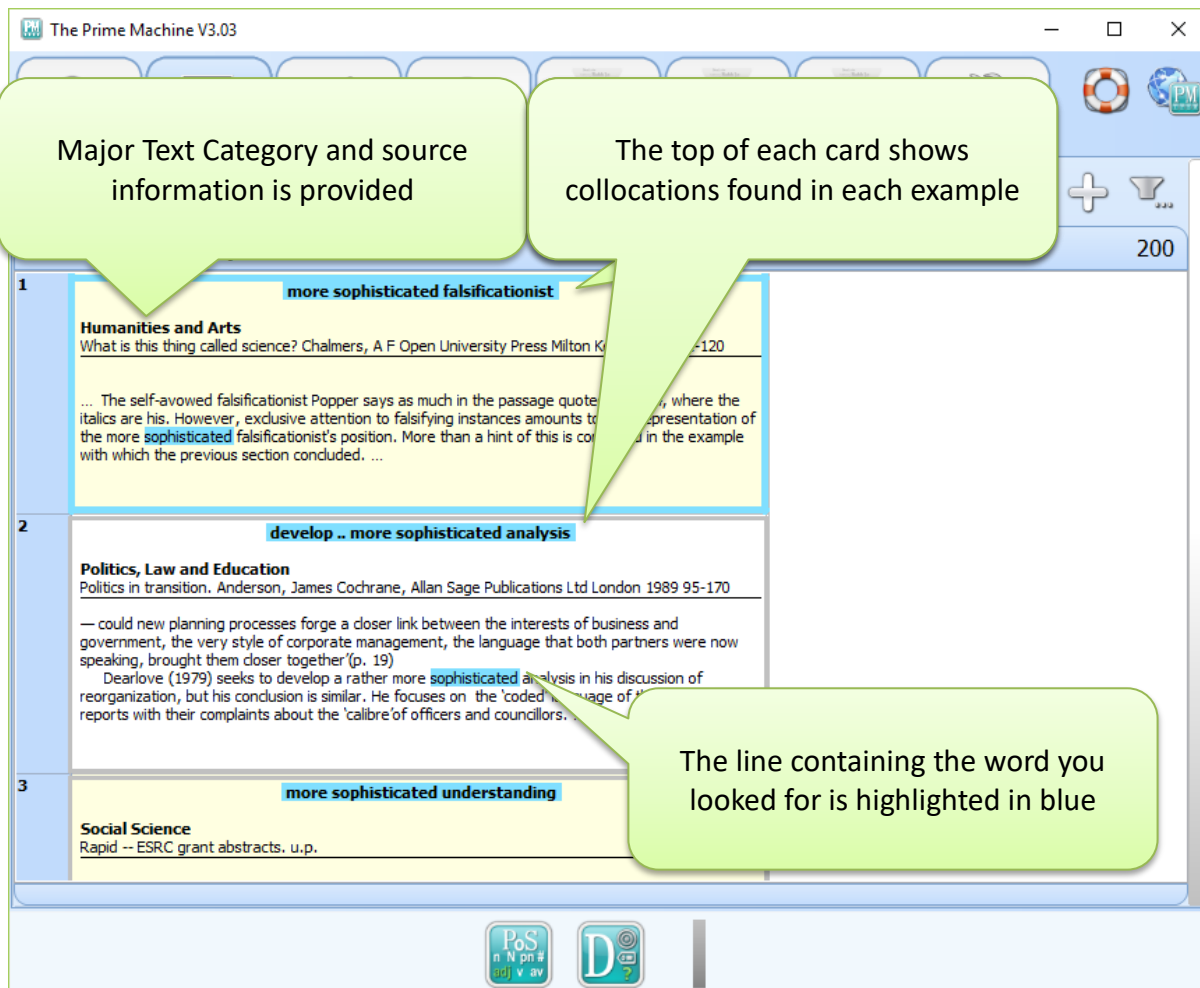
BNC: Lines + Lines Cards

1 st
2 19
3 of i
4 is c
5 t p
6 fro
7 ot b
8 : st
9 om
10 ./ /
11 : Diagnosis/Isolation: Aircraft have become

Wider context
 Sentence only
 Show Citations
 Hide Citations

The View Options button allows you to change the way concordance lines are displayed. Click to access the drop-down menu.

The same concordance lines can be viewed in a different way as a list of cards.



More about Lines

The Lines are useful if you want to look at many examples quickly. The / symbol is used to show **paragraph breaks**.

Category and source information will appear if you hover with the mouse over the line number.

| | Text to the left of node | Node | Tex |
|---|--|---------------|-------|
| 1 | g instances amounts to a misrepresentation of the more | sophisticated | falsi |
| 2 | n statements that are inconsistent with them. The more | sophisticated | falsi |
| 3 | as from the decades of dedicated work, using ever more | sophisticated | tech |
| 4 | ament police officers, from civilian support staff to more | sophisticated | tech |
| | | sophisticated | tech |
| | | sophisticated | ana |
| | | sophisticated | indu |
| 8 | h to adhere to that position. None of the modern, more | sophisticated | indu |

i Humanities and Arts
 What is this thing called science? Chalmers, A F Open University Press Milton Keynes 1992 1-120

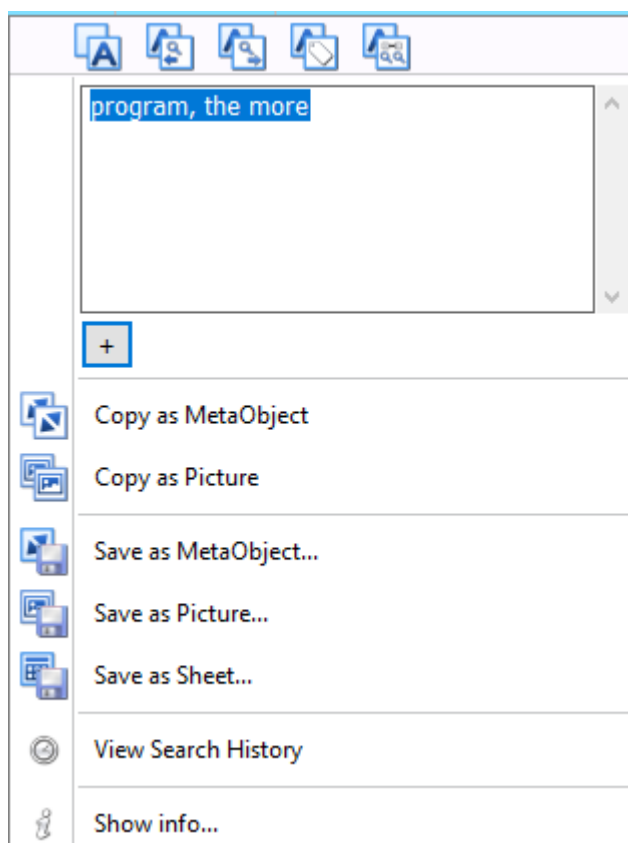
As with **all the other results pages**, right-clicking or double-clicking brings up a choice of ways in which you can copy or save the results.

MetaObjects allow you to re-size the display in another computer program without losing picture quality.

Pictures are saved as JPEG images of the text, so cannot be edited and may not appear very smoothly if enlarged.

For the Lines tab (and other tables of results), you have the option to copy or save as a spread sheet.

The default file format for saving as a sheet is as a Microsoft Excel 97-2003 Workbook (.XLS file). However, you can also choose to save as a CSV (comma delimited file).



Concordance lines on the Lines Tab will show as much text as possible to the left and right of the word, up to one sentence before and one sentence after. However, if you want to just see complete sentences, you can use the **Sentence Only** option from the View Settings drop-down menu.

Wider context:

| | Text to the left of node | Node | Text to the right of node |
|---|---|---------------|--|
| 1 | to falsifying instances amounts to a misrepresentation of the more | sophisticated | falsificationist's position. More than a hint of this is contained in th |
| 2 | together (p. 19) / Dearlove (1979) seeks to develop a rather more | sophisticated | analysis in his discussion of reorganization, but his conclusion is sir |
| 3 | the workings of industrial land markets in comparison with the more | sophisticated | understanding of labour and capital markets. / Preliminary research |
| 4 | : middle of this century by Robert Walcott, who argued for a more | sophisticated | understanding of the nature of the party structure under William a |
| 5 | : the capitalist process of accumulation. This is clearly a much more | sophisticated | analysis of the phenomena than the empirically invalidated instrum |
| 6 | that emerges from the decades of dedicated work, using ever more | sophisticated | technology, is that the creature skills develop no further. At a con |
| 7 | e, it would not be cost-effective to pursue without access to more | sophisticated | technology. / Sorting Options |

Limited to the same sentence:

| | Text to the left of node | Node | Text to the right of node |
|---|---|---------------|---|
| 1 | to falsifying instances amounts to a misrepresentation of the more | sophisticated | falsificationist's position. |
| 2 | / Dearlove (1979) seeks to develop a rather more | sophisticated | analysis in his discussion of reorganization, but his conclusion is sir |
| 3 | the workings of industrial land markets in comparison with the more | sophisticated | understanding of labour and capital markets. / |
| 4 | : middle of this century by Robert Walcott, who argued for a more | sophisticated | understanding of the nature of the party structure under William a |
| 5 | This is clearly a much more | sophisticated | analysis of the phenomena than the empirically invalidated instrum |
| 6 | that emerges from the decades of dedicated work, using ever more | sophisticated | technology, is that the creature skills develop no further. |
| 7 | e, it would not be cost-effective to pursue without access to more | sophisticated | technology. / |

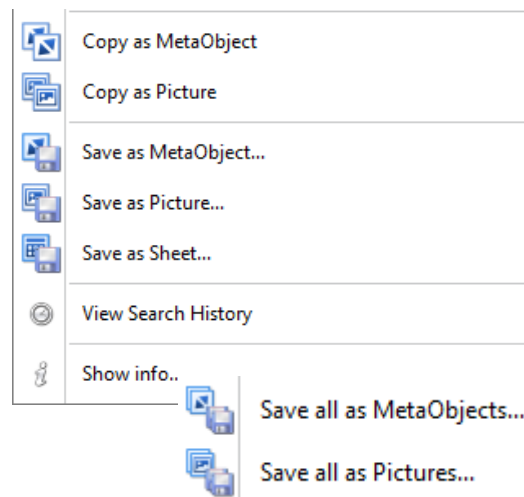
More about Cards


The Cards are useful if you want to see a wider context, and they also show you the word in its **position in a paragraph** or **heading**.

If you want to copy or save a card, you can right click or double-click on it.

Cards viewed as a long list can be copied or saved as pictures or as a spreadsheet.

If you right click on the single card from the Lines + screen, you can choose to save this card or all the cards as pictures.



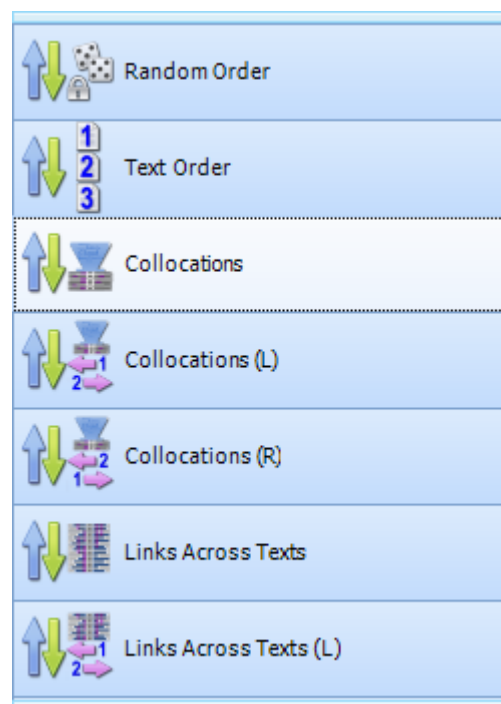
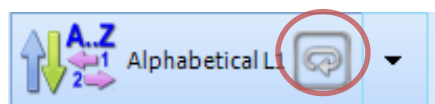
 The **Text Category** and **source information** may be useful, but it does make the cards quite large. If you want to hide this information, you can choose to hide citations from the View Settings drop-down menu.

Sorting Lines and Cards













When you perform a search, *The Prime Machine* server will send a fixed random selection of concordance lines. If the query has fewer results than your set limit (or if you use a DIY Corpus), you will receive all the concordance lines.

One of the most useful functions is being able to sort these lines and cards in different ways to help you see different patterns in the results.

There are 12 different ways to sort the results, and if you choose **Alphabetical L1** or **Alphabetical R1**, you can click the button to change the column for the sort order.

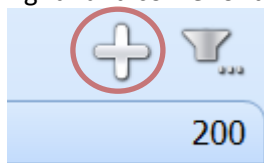


Here is a list of the different ways in which concordance lines can be sorted.

| | | |
|---|-------------------------------|---|
|  | Random Order | A fixed random order - This order will only change if the corpus is reprocessed (i.e. after a major version upgrade or if you import texts again into a DIY Corpus). |
|  | Text Order | The order in which the texts and sentences were loaded into the corpus. |
|  | Collocations | An order based on collocation strength of words 4 words to the left or right of the node - Concordance lines containing strong collocates will be pushed to the top (lines will then be promoted further based on the Links Across Texts ranking). |
|  | Collocations (L) | Same as Collocations , but only for words to the left. |
|  | Collocations (R) | Same as Collocations , but only for words to the right. |
|  | Links Across Texts | An order based on word forms frequently occurring in specific positions 4 words to the left or right of the node – Concordance lines matching others in the entire set of lines for the item will be pushed to the top (lines will then be promoted further based on the Collocations ranking). |
|  | Links Across Texts (L) | Same as Links Across Texts , but only for words to the left. |
|  | Links Across Texts (R) | Same as Links Across Texts , but only for words to the right. |
|  | Links Within Texts | An order based on links between the sentence and other sentences in the same text – Concordance lines where words in the sentence are repeated in many other sentences in each text will be pushed to the top. This is not available for DIY corpora. |
|  | Dictionary Style | An order based on shorter sentences, with fewer low frequency words – Concordance lines with many symbols, rare words, many names, etc. will be pushed to the bottom. |
|  | Alphabetical L1-L4 | Alphabetical order for words 1, 2, 3 or 4 slots to the left of the node. |
|  | Alphabetical R1-R4 | Alphabetical order for words 1, 2, 3 or 4 slots to the right of the node. |

How to get more results

If you want to see **more concordance lines** (more cards and lines), click on the **plus** button in the top right hand corner of the results window.



You can also increase the **number of results** to return each time by changing the setting on the **Options Tab**.

Filtering results

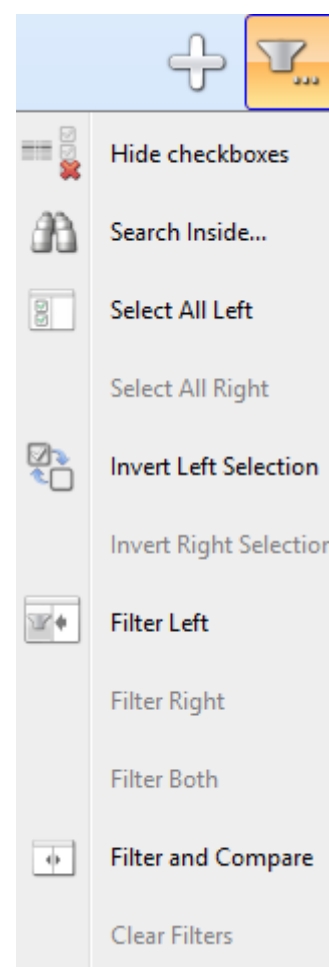
Access the Filter drop-down menu by clicking on the icon in the top right-hand corner on the Lines/Cards Tab.

There are a number of different ways to select and filter down some of the concordance lines.

- You can Hide/Show checkboxes to highlight lines yourself.
- You can use the **Search Inside...** button to highlight (or clear) concordance lines containing specific text in the co-text.
- You can use the buttons on the **Graphs Tab** to deselect some priming features and then highlight concordance lines matching your requirements (see page 21 below).

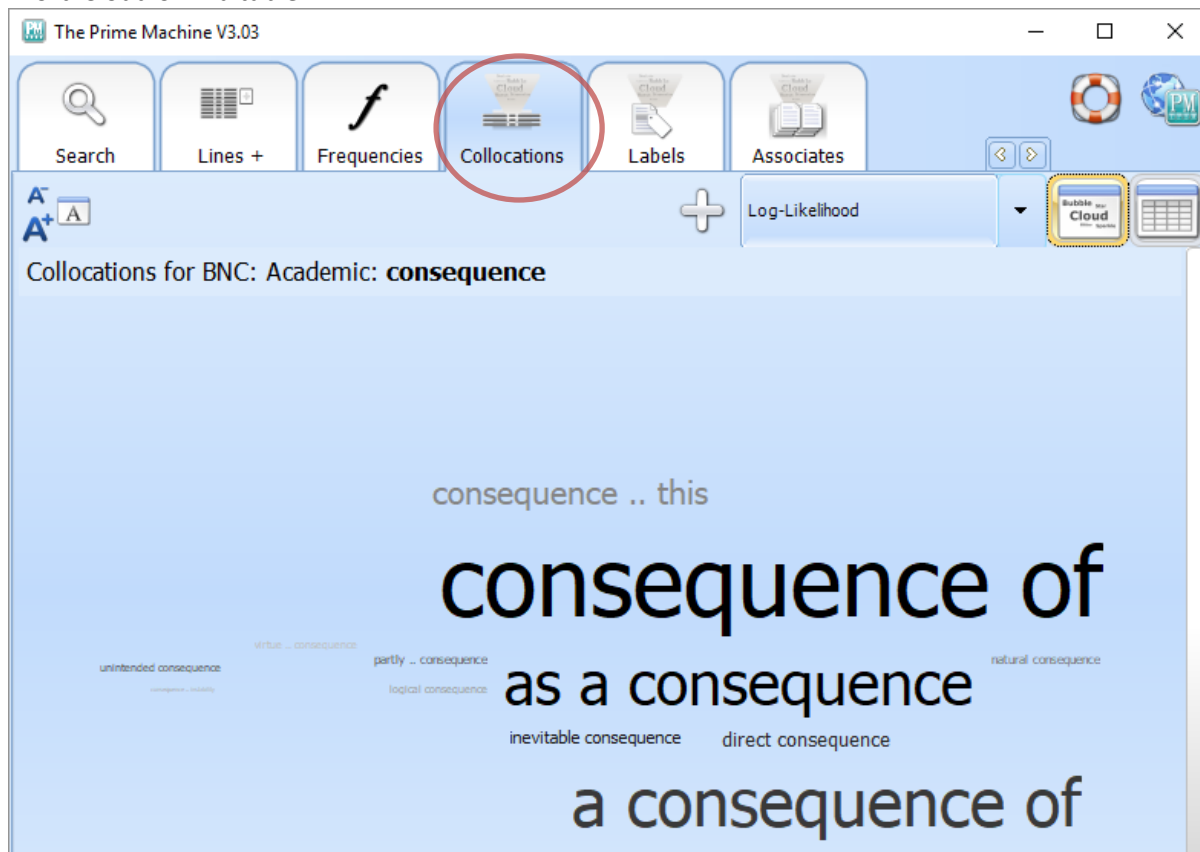
After highlighting one or more lines you can filter the results to only show the selected items by using **Filter Left** (or Filter Right, etc.).

If you want to compare the highlighted lines with those which are not highlighted, clicking **Filter and Compare** will split the results with those on the left only showing those which had been highlighted, while the remainder appear on the right.



Collocations

The **Collocations** tab shows words which are frequently used together. You can view the results as a Word Cloud or in a table.



Use the **Show** table button to see the results in a table.

| | Collocation | Frequency | Log-likelihood ▾ | Bayes Factor |
|----|----------------------------|-----------|------------------|----------------------|
| 1 | as a consequence | 337 | 464.42 | Very strong evidence |
| 2 | consequence of | 866 | 425.16 | Very strong evidence |
| 3 | inevitable consequence | 24 | 160.23 | Very strong evidence |
| 4 | direct consequence | 33 | 130.06 | Very strong evidence |
| 5 | a consequence of | 388 | 62.94 | Very strong evidence |
| 6 | unintended consequence | 7 | 59.62 | Very strong evidence |
| 7 | partly .. consequence | 13 | 44.31 | Very strong evidence |
| 8 | natural consequence | 17 | 43.89 | Very strong evidence |
| 9 | consequence .. this | 106 | 38.95 | Very strong evidence |
| 10 | logical consequence | 9 | 34.57 | Very strong evidence |
| 11 | consequence .. instability | 6 | 34.16 | Very strong evidence |
| 12 | virtue .. consequence | 8 | 31.99 | Very strong evidence |
| 13 | necessary consequence | 16 | 27.89 | Very strong evidence |
| 14 | consequence .. unsolicited | 3 | 26.36 | Strong evidence |
| 15 | consequence .. breach | 9 | 23.10 | Strong evidence |
| 16 | probable consequence | 5 | 21.18 | Some evidence |

There are several ways to calculate collocations.

The Log-likelihood measure takes into account the position of the words and also provides longer collocations of up to 5 words.

Delta P takes into account the order (A then B) and shows collocations of 2 words in length.

The other three measures (MI3, T-Score and Dice) are standard collocation measures which are also available in other concordancer programs.

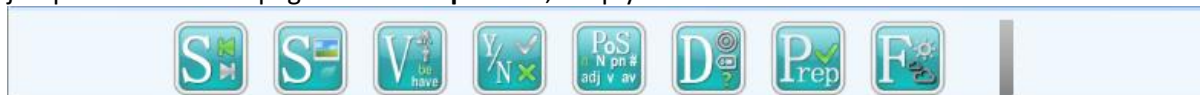
The table for these shows the proportion of occurrences for each word in 4 slots to the left and right of the main word.

Clusters (also known as n-grams) are not based on a statistical measure; they are repeated strings of words which occur above a minimum frequency, sorted by length.

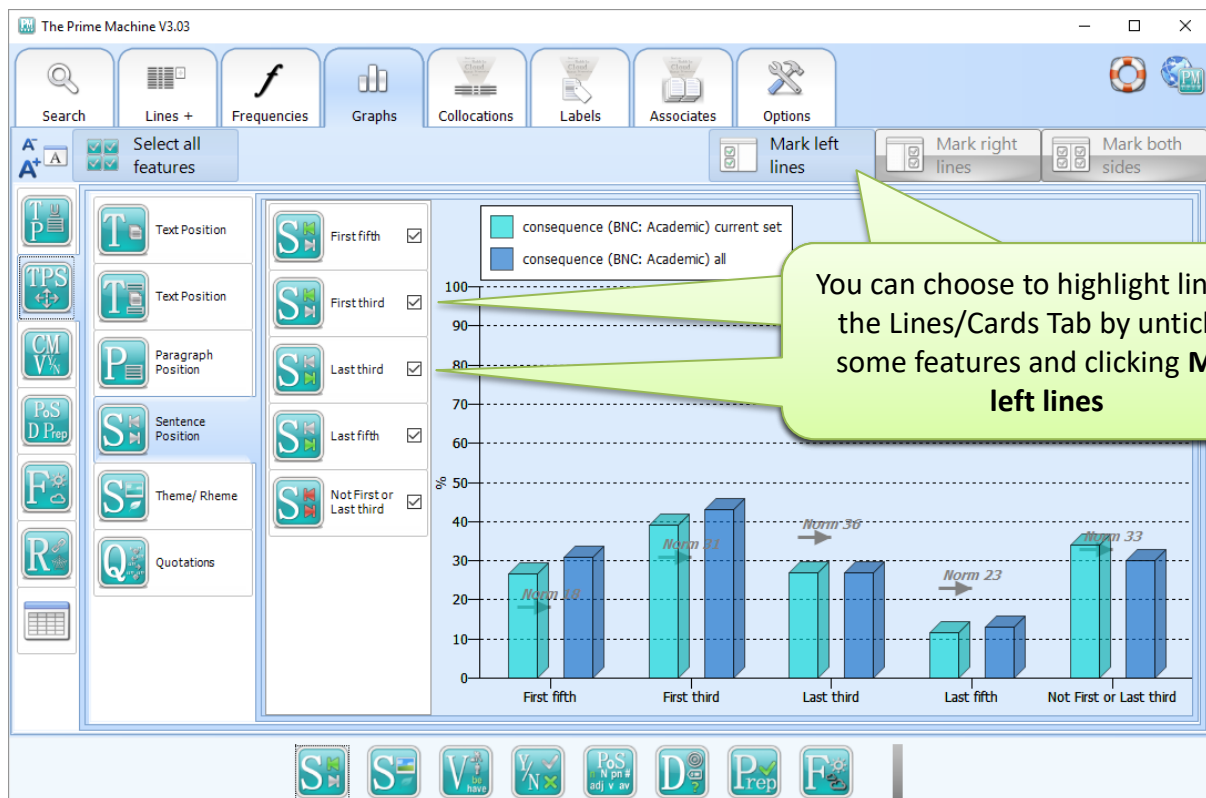


Graphs

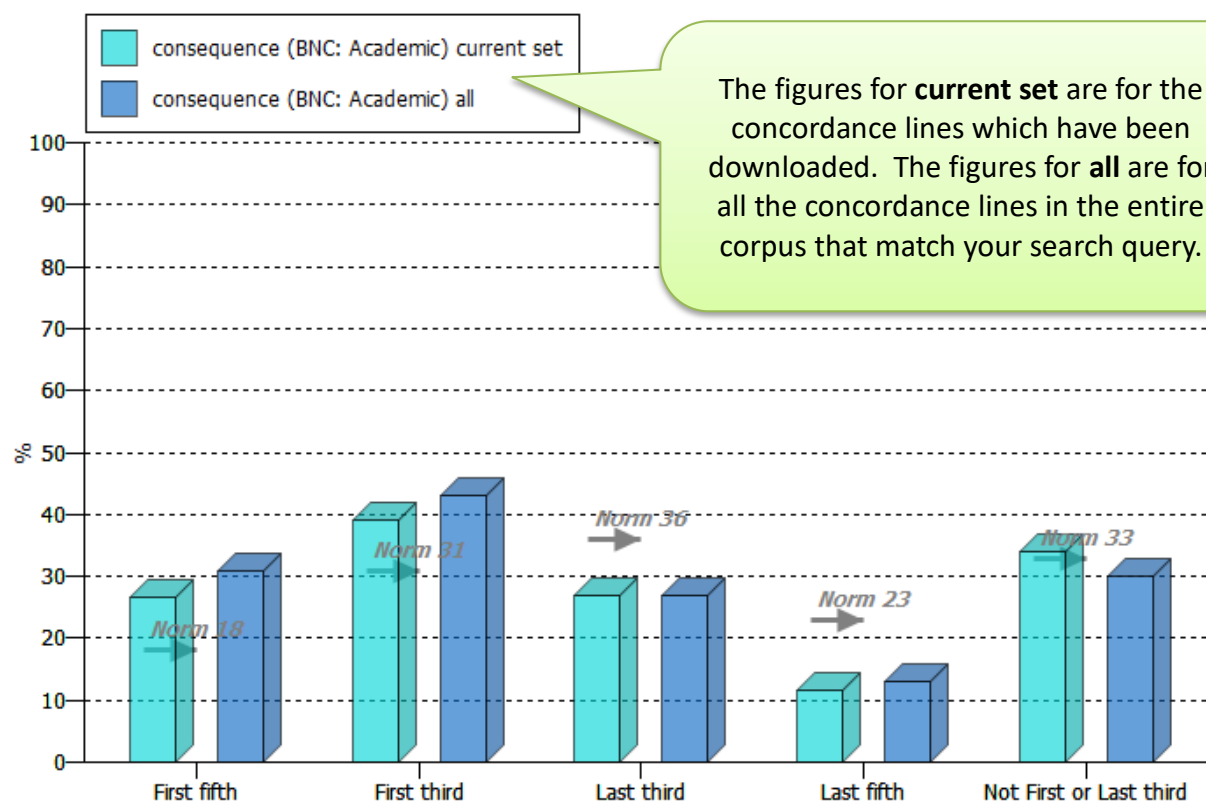
At the bottom of the screen a number of icons will appear showing some of the typical contexts. This set of icons is designed to encourage you to explore the information on the **Graphs Tab**. To jump to the relevant page on the **Graphs Tab**, simply click on one of the icons.



The menus and sub-menus on the **Graphs Tab** can be quite complicated at first, but they provide access to summary information about a wide range of features. Each sub-section of the **Graphs Tab** has its own **Life-ring** information.



For example, the graph below shows how the word *consequence* in the BNC: Academic corpus occurs very frequently in the first fifth and first third of a sentence. The **Norms** show the proportion of all words in the corpus which occur in these contexts, so it is easy to see how the tendencies differ.



Comparing results side-by-side

With *The Prime Machine* it is easy to compare two words or phrases at the same time. Simply, enter one search term in each of the search boxes and click the **Compare** button.

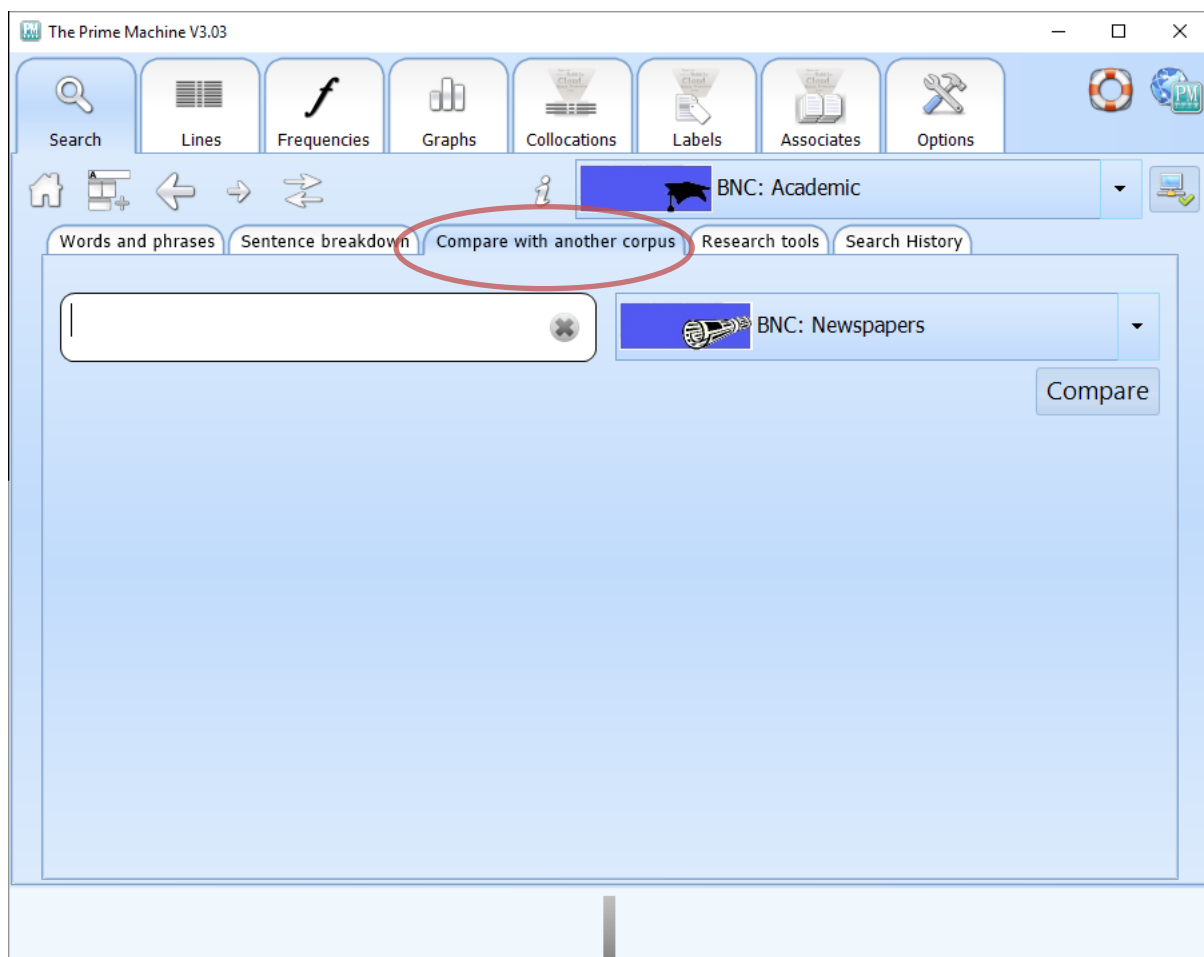
Suggested uses:

- See how different word forms of the same word are used differently
For example: *analysis vs. analyse*
- See how a word is used differently when it is part of a collocation
For example: *effective vs. cost effective*
- See how words with a similar meaning or translation are used differently
For example: *consequence vs. result*

Comparing results from two different corpora

As well as looking at results from just one corpus, with *The Prime Machine* you can also see results from two different corpora side-by-side. This will allow you to see how words and phrases may be used differently in different genres or text types.

To use this feature, simply click on the **Compare with another corpus** tab inside the **Search Tab window**. Enter the word or phrase in the box and then choose a corpus to use for comparison.



More advanced features

Labels

Sentences in a corpus often have labels attached to them giving details about the text, the section or the producer. A statistical measure can determine whether a word or phrase appears in particular kinds of text more often than expected by chance.

- Text labels include the main category of a text (e.g. *Fiction* or *Academic*). They may also include information about the publisher, the source or the genre;
- Section labels are the sub-headings of a text (e.g. *Abstract*, *Introduction* or *Conclusion*);
- Producer labels provide information about the writer or speaker (e.g. their name, age, gender, etc.)
- Neighbourhood labels are based on repeated semantic tags for words and phrases in the current card.

Sometimes, the labels represent only a few instances of the word; sometimes they represent most of the instances of the word. The line and arrows above each box show the proportion of occurrences of the word which match the labels that are visible in the cloud. If the percentages are very high, it means that labels in the cloud account for most of the concordance lines available. If the percentages are very low, it means the word occurs in many other contexts as well.

Detailed notes:

Since some occurrences of the word may be connected with more than one label, the figures are shown as a range of values rather than a single percentage. The lower end of the range shows the proportion of occurrences represented by the label which matches the most instances of the word. The higher percentage shows the maximum proportion of occurrences which could be represented in the cloud since it gives the combined frequencies for all the labels.

Associates

Collocation measures look at a few words either side of the main word, but you can also get a sense of how a word is used by seeing which other words or phrases often appear in the same texts. The words and phrases in the clouds are **key words** in the same texts as the main search word.

This tab also shows you the top text category for your search and the percentage of results which belong to that category.

The **Norm** arrows show whether this is above or below the expected proportion if all words were divided equally between all the categories.

If your word never occurs in a category, it will be listed in the **Empty Categories box**. This might suggest that the word is not usually used in that kind of text.

If you are interested in seeing a list of key words for a specific category or tag, you can use the **Tags** search on the main **Search Tab window**.

Options

There are a number of changes you can make to customise your searches.

The most useful options are related to **Ranking methods** and **Results**.

Each sub-section of the **Options page** has its own **Life-ring** information, so you can view additional information by clicking on this.

Settings

Search suggestions

Ranking methods

Results

Tables and Clouds

Collocations

Charts

Special Access

Load/save settings

Reset

Support

The Prime Machine is still undergoing development.

For further information see <http://help.theprimemachine.com>

Last Updated: 14 February 2022