

Getting Started

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.

Screenshots in this guide are from the iPad version.

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:

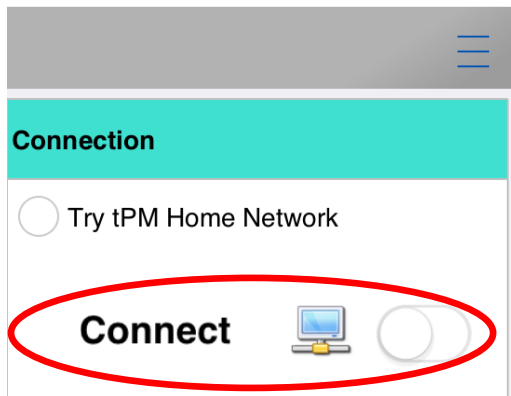
- Looking up words to find useful phrases or collocations;
- Comparing the use of two or more synonyms;
- Seeing how words are used differently in different genres or text types;
- Finding examples from other texts to compare the choice of language with your own.

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.

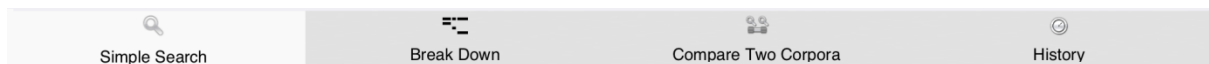


Click connect

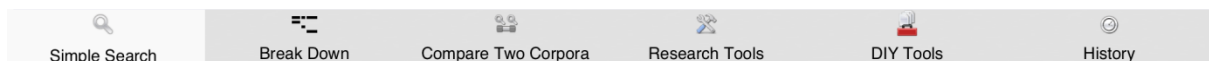
Click **Connect** and the application will connect to the server and retrieve information about the corpora which are available. (If you have already saved results or you want to create DIY corpora offline, you can work offline.)

Getting around

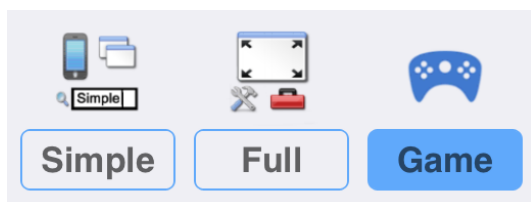
There are two main views for the search screen – Simple Mode and Full Mode. The Full Mode includes additional tabs and features for corpus research and DIY corpus work.



Simple Mode

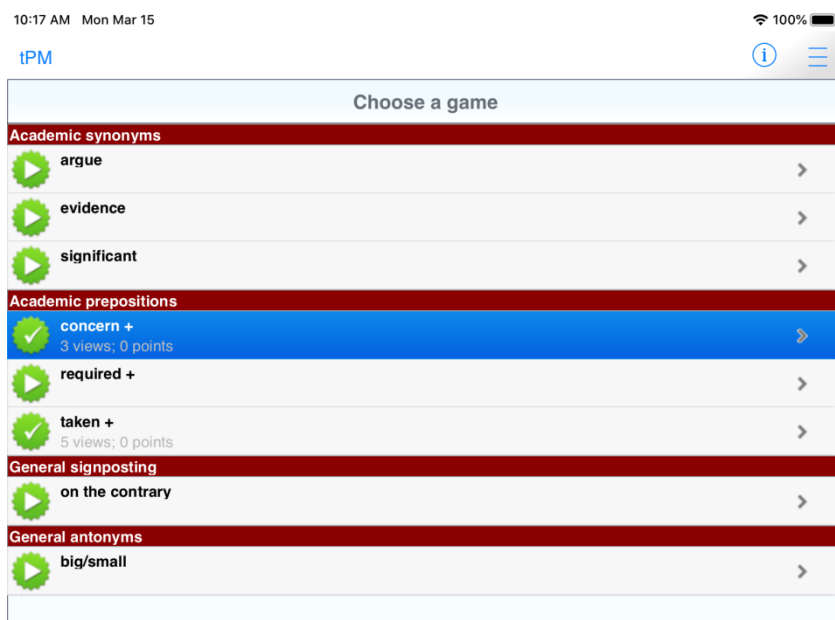


Full Mode

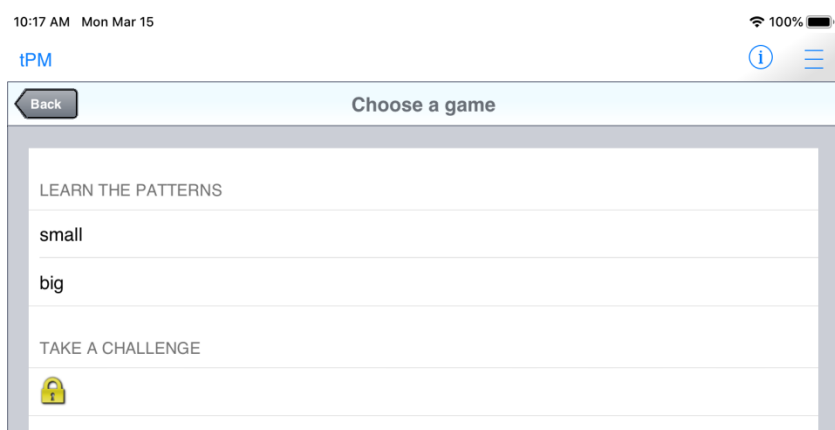


Game Mode

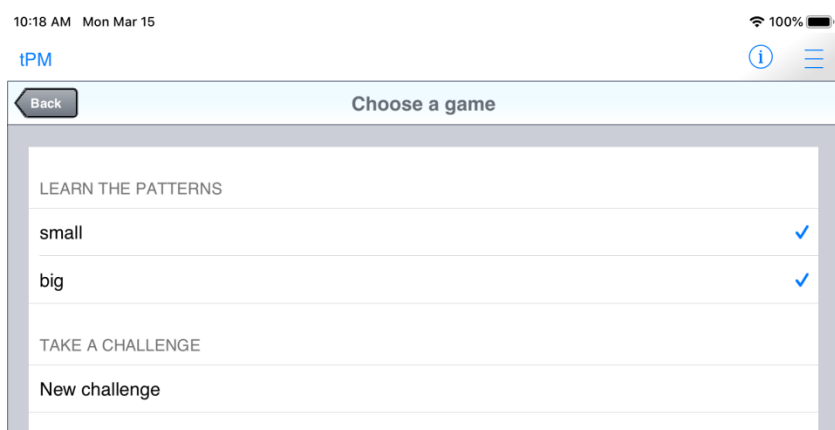
Game Mode provides a different user interface, giving you a list of groups of words to study before you *take a challenge* and try to guess which word matches a set of concordance lines.



First, choose a set of words for the game.



Next, choose each word from the list and look at concordance lines and other corpus data to learn its patterns of use.



You must look at each word for at least 15 seconds before it will appear ticked on the list.

When you have looked at examples for all the words in the set, the **New challenge** option will appear.

When you start a new challenge, you need to decide how many lines you think you need and how confident you are that you will guess each word correctly.

This is your opportunity to increase the number of challenges you can take (and increase the opportunities to win points) and to increase your rewards for choosing the correct answer.

9:53 PM Sun Feb 13

<

How many lines do you think you need?

1

25

50

75

100

How confident are you?

evidence

proof

phenomenon

grounds

justification

Confirm

The challenge looks like a normal set of concordance lines, but one or more words will be hidden.

You need to read the lines and guess the word from the patterns of use you can see in the lines.

If you choose four or five, you will get more points if this is the hidden word and if you guess it correctly. However, you will also face a heavier penalty if you guess this word incorrectly.

10:19 AM Mon Mar 15

Collocations

50

	Text to the left of node	Node	Text to the right of node
1	e underweight by more than a	---	amount./ In addition to the wei
2	on reservation pressure from a	---	minority of the library's clientel
3	al employer, especially if it is a	---	business, that you have arrang
4	ing of space, so use them in a	---	room that needs opening up/ S
5	to a piece of wire leading to a	---	silver sensor. You also get a ke
6	, it has been decided to start a	---	fund for a Christmas Party for t
7	n cinema screens represents a	---	slice of what is produced globa
8	IER IDEAS/ Another idea for a	---	gift is to decorate a bookmark.
9	Australia's Qantas and TAT, a	---	French airline. Buying into othe
10	y A pale pink floor can make a	---	area appear much larger than i
11	ised chewing behaviour and a	---	brain does not make it our dire
12	every-second Billy was with a	---	'b'. I goes Annemarie I doubt if
13	r themes, stressing the role of	---	businesses, small savers, char
14	ne months of this year, 12,359	---	businesses failed in London ar
15	shaped teeth, and a relatively	---	eye. Coloration is very variable
16	d to homoeopathy, where very	---	changes in the chemical comp
17	till's Prescription i/d. Feed him	---	quantities at regular intervals -
18	in both rural and urban areas.	---	firms have traditionally found it
19	e too large for a community of	---	fish — even the Flag cichlid, al
20	of PO Box numbers./ Read the	---	print; you never know what you
21	to 2 lb and large dabs. A lot of	---	codling in the Menai Straits, plu
22	hing than Kestrel, in pursuit of	---	birds, though may hover. Hills ;
23	ng of a large bay studded with	---	clay islands called 'eyes'. By th
24	to, sometimes the heirs of, the	---	- nationality movements directe
25	ie men had brought ropes and	---	axes with them so that, while th
26	he path is clearly marked with	---	piles of rocks at intervals, and i
27	atal, for example, it leaves two	---	marks that remain even when t
28	her with wire-brushing of some	---	areas of stonework where this
29	home when the children were	---	to make a bit of money.' Now s
30	Common name:/	---	-leaved Amazon Sword/ Easy p
31	table cruiser. Certainly it's one	---	car which I can drive for sever
32	vill have a picture of your ideal	---	farm with grass, crops, livestoc

Lines

Answer

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

Simple and Full Mode

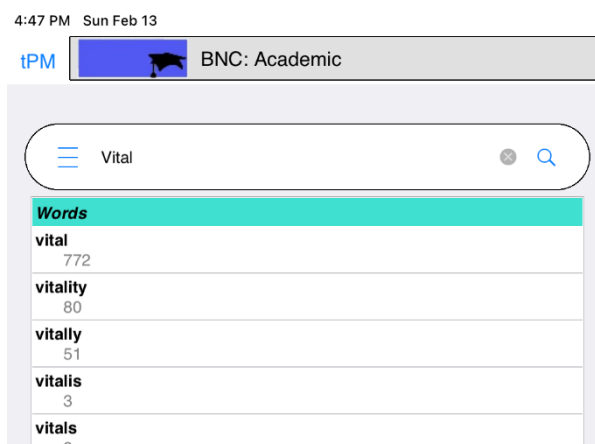
Here is a list of the main tabs and their purposes:

Simple Search	Enter a word, collocation or phrase to request results from the server.
Break Down	Enter a phrase or sentence and break it down into words and combinations of words. Frequencies of these combinations can then be checked in the corpus, along with suggestions for reformulations.
Compare Two Corpora	Look up a word or phrase in two corpora at once. Then view the results side by side (screen size permitting) or flip between them using the forward/back buttons on the results screen.
History	Save results from your searches to a file for viewing offline later or re-open previously viewed results. On iOS and Android these files can be shared.
Research Tools	This tab opens a new set of tabs for a variety of research-oriented corpus work using the readymade corpora held on the server.
DIY Tools	This tab opens a new set of tabs for a variety of research-oriented corpus work using your own DIY texts. These tools can allow you to compare your DIY corpus with a readymade corpus or other DIY corpora.

Simple searches

You can enter between 1 and 5 words in each search box. If you type a few letters, you can click the hamburger button inside the text box to get auto-complete suggestions. Each suggestion shows the frequency of the word in the corpus. This is the number of examples for that word found in the current corpus.

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



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**

4:47 PM Sun Feb 13 65%

tPM BNC: Academic ⓘ ☰

Compare two words, collocations or phrases (A&B)

☰ Vital ✕ 🔍

Words

vital 772

vitality 80

vitality 51

vitalis 3

vitals 3

Collocations

vital part 38

vital importance 24

vital role 26

it .. vital 110

play .. vital 13

☰ ✕ 🔍

Other forms

vitality 51

vitale 2

Similar meanings

critical 1,989

important 10,898

crucial 1,375

extremely important 59

must be 5,713

🔍 Simple Search
☰ Break Down
🔄 Compare Two Corpora
🕒 History

If you enter a string of letters which cannot be found in the currently selected corpus, it does not always mean that the word is misspelt, but rather that no results can be found. On the other hand, if you click on the hamburger button again you should get some spelling suggestions.

10:22 PM Mon Feb 14

tPM BNC: Academic

☰ consequence ✕ 🔍

Spelling suggestions

consequences 1,701

changing 1,500

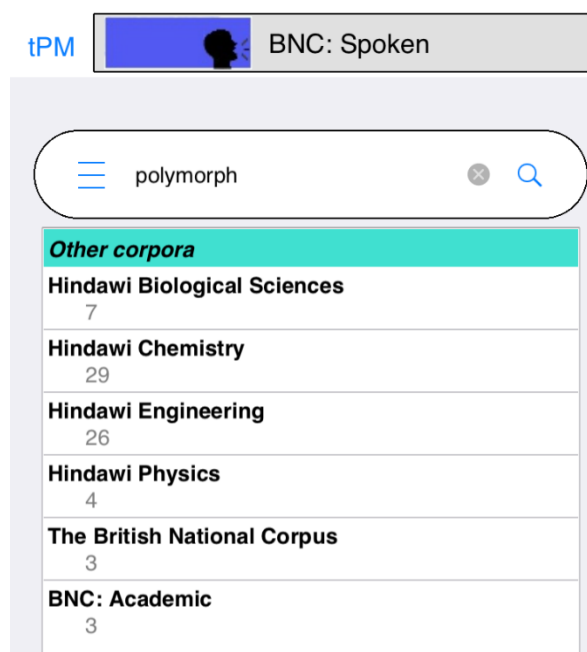
consequence 1,406

consciousness 1,355

consensus 721

If you search for a word which is not found in the current corpus, 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. However, this **technical word** does occur in a number of academic texts, and its frequency in each will be displayed. If you tap or double-click on one of the corpora listed, the program will start using that corpus.

10:22 PM Mon Feb 14



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.






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.

10:23 PM Mon Feb 14

tPM  The British National Corpus

 president united states  

Collocations

united states president
47
president .. united states
92
president * united * states
president united states



*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.*

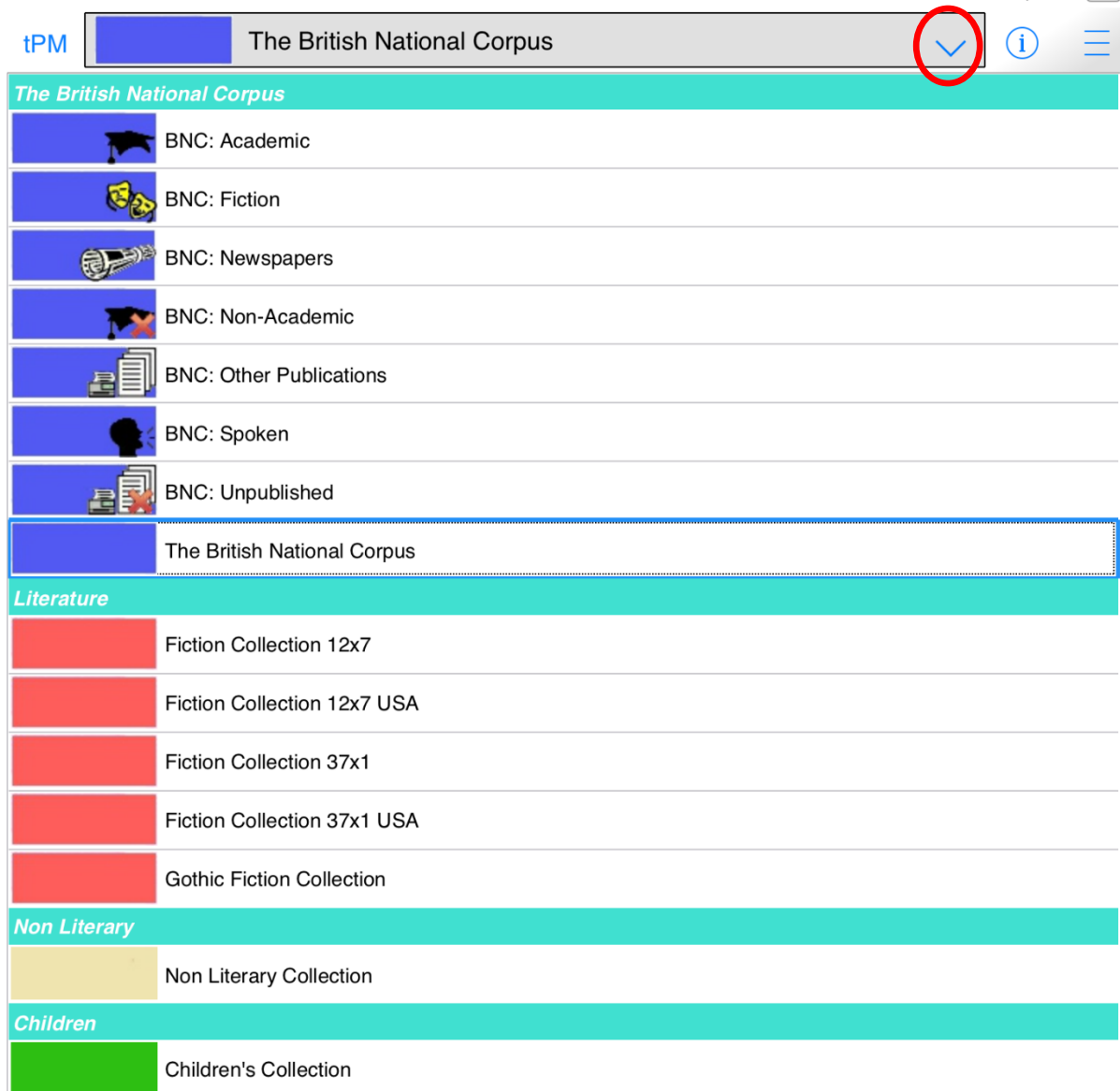
Choosing a corpus

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.

10:23 PM Mon Feb 14

62%



tPM The British National Corpus

The British National Corpus

- BNC: Academic
- BNC: Fiction
- BNC: Newspapers
- BNC: Non-Academic
- BNC: Other Publications
- BNC: Spoken
- BNC: Unpublished
- The British National Corpus

Literature

- Fiction Collection 12x7
- Fiction Collection 12x7 USA
- Fiction Collection 37x1
- Fiction Collection 37x1 USA
- Gothic Fiction Collection

Non Literary









- Non Literary Collection

Children

- Children's Collection

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 Social Sciences
		Hindawi Physics
		Hindawi Mathematics
		Hindawi Engineering
		Hindawi Earth Science and Environment
		Hindawi Computer Science
		Hindawi Chemistry
		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.

RESULTS

You can explore corpus data in several ways. Here are the different tabs of the results screen. Names shown with a * are only available when connecting using Full Mode.

Lines or Cards Tab*	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 shown as a card (see Concordance lines below).
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 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 Tab	The Collocations Tab shows words which are often used together, based on several statistical measures.
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 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.

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.

4:47 PM Sun Feb 13

Collocations

BNC: Academic: **vital** (Simple Mode) 200

	Text to the left of node	Node	Text to the right of node
1	pital services Acute non-psychiatric hospitals play a	vital	role in the provision of medical care for all members o
2	I towns (population less than 5,000 in 1801) played a	vital	role in provincial Britain and France as marketing, cra
3	tive.But, to admit this is to admit that theory plays a	vital	role prior to observation.The naive inductivist cannot e
4	, unpublished observations).Spleen size may play a	vital	role but other reports have shown that this relates poc
5	giving an intimate and unique atmosphere which is	vital	part of the essence of the music.Accompaniment is lik
6	/ There can be little doubt that th		
7	s a cornerstone of the German w		
8	iewed suggest that residential w		
9	t must inevitably appear at times		
10	and the rationing that must be a		
11	rns. Fabians also believed that		
12	nd, more importantly, advertising revenue, became a	vital	ingredient in the make-up of the newspaper.It contribu
13	y control in GIS. In summary, the user interface is a	vital	element of any GIS.Long ignored as an esoteric aspe
14	n activity in art and design.In this regard they are a	vital	element in the system and serve two very important f
15	ation of Latin America and, since values constitute a	vital	component of culture, the H
16	ke some more specific ground.(b) Procedure It is of	vital	importance that the procedu
17	e cases incur fruitless costs in the search for such a	vital	clue where none exists.But,
18	re it otherwise, the press would be unable to play its	vital	role of 'public watchdog'./M
19	om line in the pay negotiations.It was considered of	vital	importance to obtain the uni
20	t some of the texts examined show it playing a more	vital	role./Protective formulae of t
21	and phone numbers This is self-explanatory, but it is	vital	to get the spellings correct.There are sometimes curic
22	m any of the professions in the team However, it is	vital	that the therapist, whatever his or her profession, has

play a vital role

Social Science
Health and health care in later life. Victor, C Open University Press Milton Keynes 1991 14-156

Utilization of hospital services
Acute non-psychiatric hospitals play a **vital** role in the provision of medical care for all members of the community but especially older people. At its creation the NHS inherited the existing stock of hospital provision.

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.

Lines Frequencies Collocations

10:24 PM Mon Feb 14

Collocations

Lines +

Lines

Cards

Wider Context

Sentence Only

Show Citations

Hide Citations

Normal Spacing

Extra Spacing On

No Mask

Next Mask

Text to the right of

role in the provision

role in provincial Brit

role prior to observat

role but other reports

part of the essence c

part of the political s

part of the legitimacy

part of child care.It is

element of Britain's c

part of family welfare

role to play in integr

ingredient in the mal

element of any GIS.I

element in the syste

component of culture

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

The Card view shows important information about each concordance line.

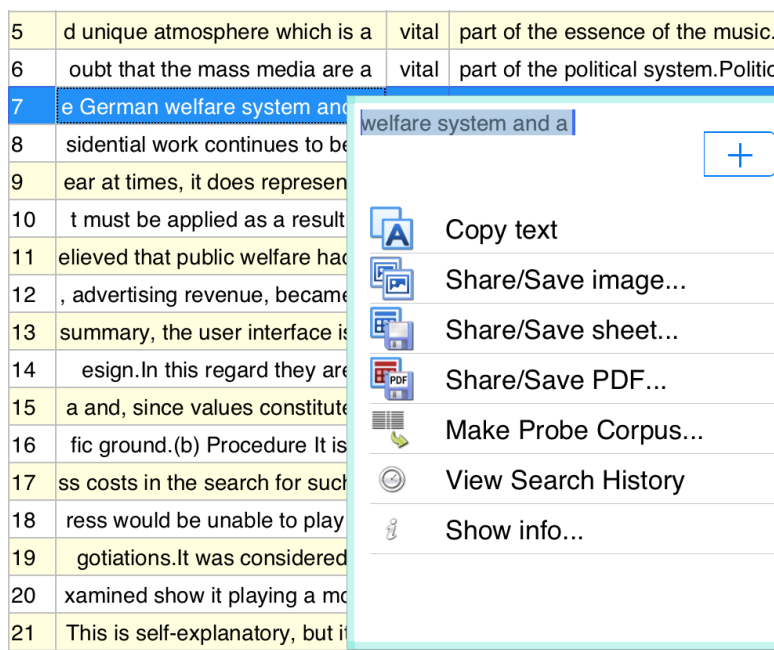
The diagram shows a sample concordance card. A light green callout bubble on the left points to the top section of the card, stating: "Major Text Category and source". Another light green callout bubble on the right points to the top line of the card, stating: "The top of each card shows collocations found in each example". The card itself is a white rectangle with a black border, set against a light purple background. At the top of the card, the phrase "play a vital role" is highlighted in a light blue box. Below this, the text "Social Science" is in bold. Underneath, the source information is listed: "Health and health care in later life. Victor, C Open University Press Milton Keynes 1991 14-156". A horizontal line separates this from the next section, which is titled "Utilization of hospital services" in bold. The main body of the card contains the following text: "Acute non-psychiatric hospitals play a vital role in the provision of medical care for all members of the community but especially older people. At its creation the NHS inherited the existing stock of hospital provision. ...". The word "vital" in the first sentence of this paragraph is highlighted in a light blue box.

More about Lines

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

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

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



The default file format for saving as a sheet is as a Microsoft Excel Workbook (.XLSX 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.

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, share or save a card, you can long tap, right click or double-click on it.

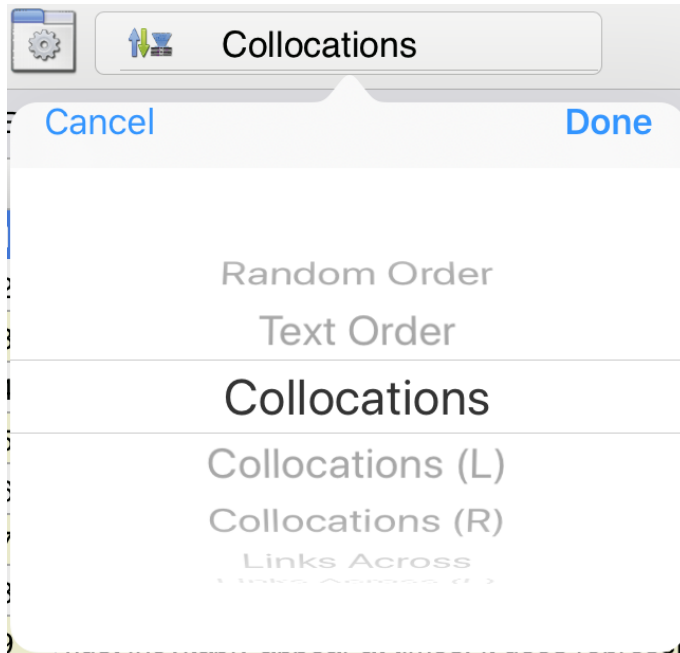


The **Text Category** and **source information** may be useful, but it does make the cards quite large. If you want to hide this information, you can choose to hide citations from the View Settings drop-down menu.

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.

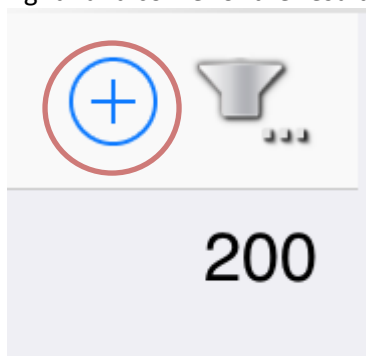


Here is a list of the different ways in which concordance lines can be sorted.

	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 page**. Bookmark not defined..

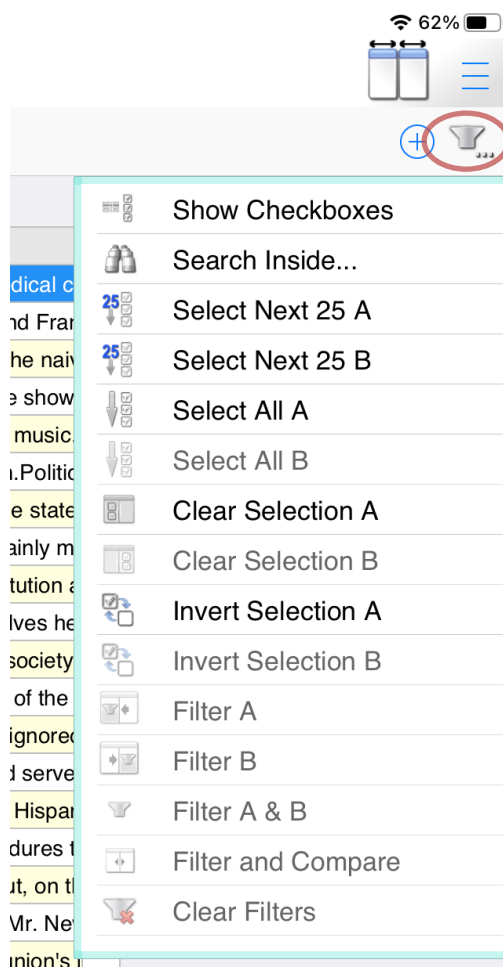
Filtering results

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 under view settings to see the results in a table.

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.

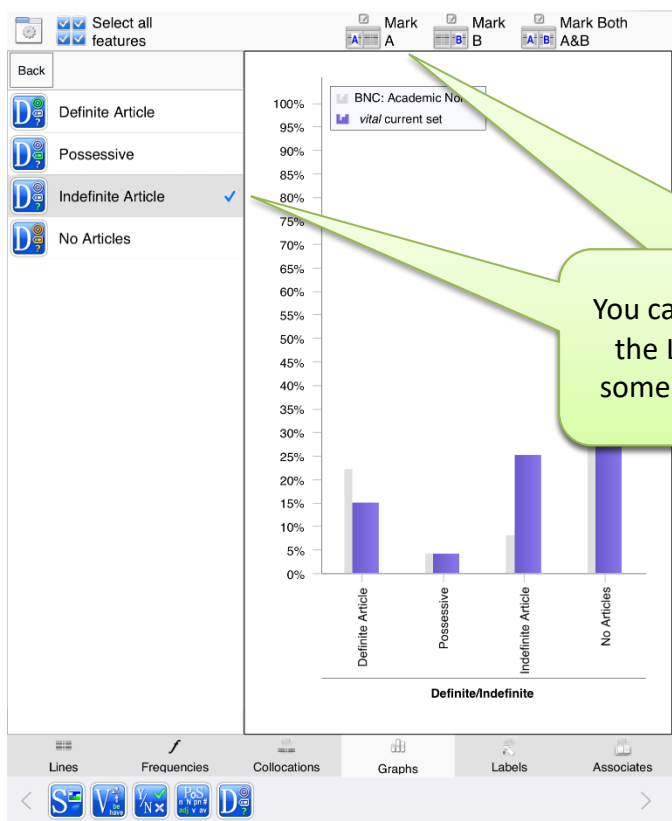
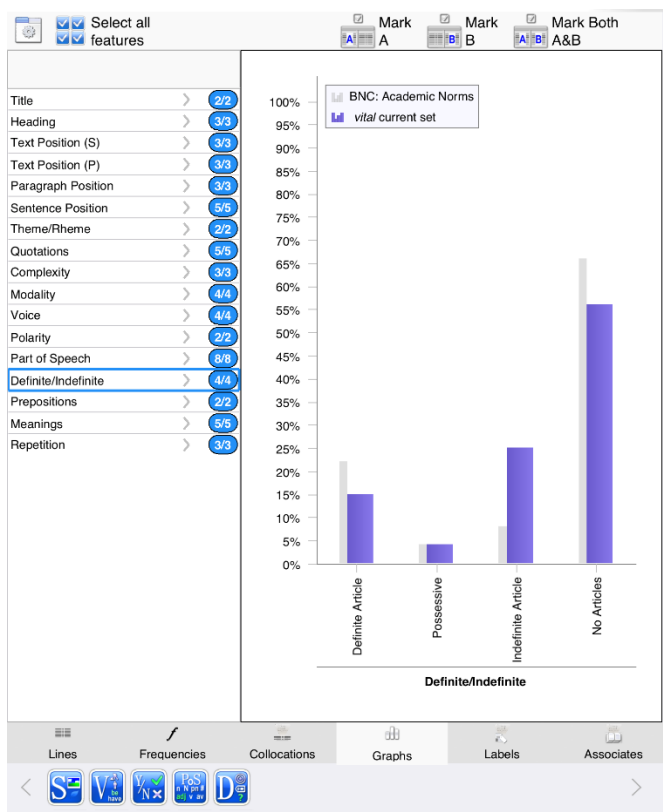
Graphs

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.

For example, the graph below shows how the word *vital* in the BNC: Academic corpus occurs very frequently with the indefinite article. 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.

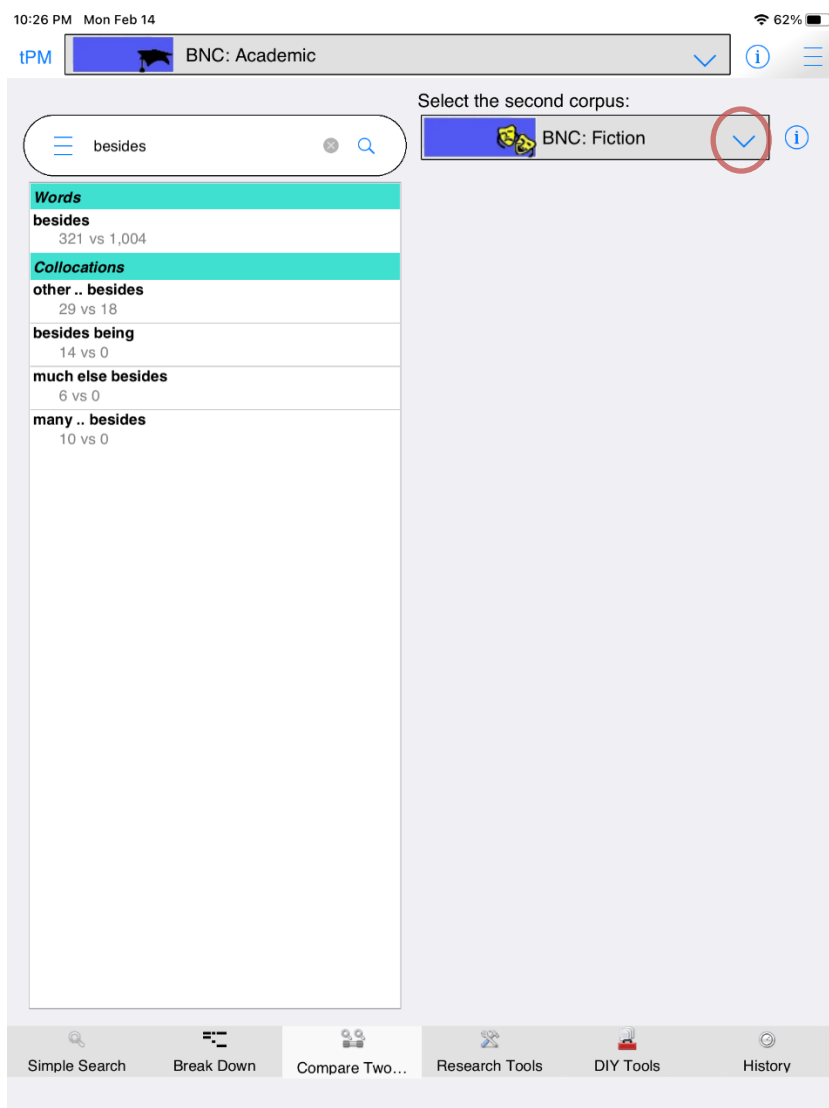
Suggested uses:

- 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**.

Support

For further information see www.theprimemachine.net/help.html

Last Updated: 14 February 2022