



# TamBiC Help

## The Tampere Bilingual Corpus of Finnish and English

### [Basic search procedure](#)

1. [Some basic facts](#)
2. [How to read the main page](#)
3. [How to read search results](#)

### [More search help](#)

1. [Wild cards \(\\*\)](#)
2. [Operands \(OR, AND, THEN, NOT, "..."\)](#)
3. [Distinguish upper and lower case letters](#)
4. [Refining](#)

### [Frequently Asked Questions](#)

1. [Why was the TamBiC corpus created?](#)
2. [What kind of texts can I find in the TamBiC corpus?](#)
3. [How can the TamBiC corpus help me?](#)
4. [Who can use the TamBiC corpus and how much does it cost to use?](#)
5. [What is the format of the texts in the TamBiC corpus?](#)
6. [Where can I find more information about the texts used the TamBiC corpus?](#)
7. [Who compiled the corpus?](#)
8. [Who designed the search engine?](#)

## Basic search procedure in TamBiC

Search is simple:

- Select the corpus and the language you are interested in.
- Type the word you wish to find in the search box.
- Hit Enter or click on the New Search button, and TamBiC will search for sentences containing your query.

Most of the time you might find exactly what you are looking for with just a basic query. However the following tips can help you refine your technique to make the most of your searches.

[Back To Top](#)

### Some basic facts



- Every word matters. Generally, all the words you put in the query box will be searched for.
- The searches are always case **insensitive**. Unless you choose otherwise, the search engine will look for words beginning with both upper and lower case letters. A search for **dog** will also find **Dog**.
- In general, punctuation is ignored and you cannot search for @ # \$ ^ & \* () = + [ ] \ or other special characters). However, apostrophes (**can't**, **John's**) can be included in search strings.

[Back To Top](#)

### How to read the main page

The diagram below indicates the features that are important for understanding the main page, and will help you when performing searches.



① Select corpus: English to Finnish  ② Select language: English 

③

☐ Distinguish upper and lower case letters. ④

1. **Scope of the corpus:** The TamBiC corpus is a bilingual translation corpus and is in two parts: original English texts together with their Finnish translations ("English to Finnish"), and original Finnish texts with their English translations ("Finnish to English"). You can select the part of the corpus you want to search by clicking on the Select Corpus button.
2. **Language selection:** It is possible to search in either the original language or in the translations. Just click on the Select Language button.
3. **Search box:** This is where you type the word(s) you are searching for.
4. **Distinguish upper and lower case letters:** Normally the search engine will look for words beginning with either upper or lower case letters. Click here if you want the search to include (or exclude) words with initial capitals (i.e. so that searching for **dog** is NOT the same as searching for **Dog**).
5. **Help button:** This is a link to the help file. Click here for detailed information about the search engine and the different features that are available.
6. **About TamBiC:** This is a link to the information file. It contains information about the corpus, the authors and texts that are included, and the development of the search engine.

[Back To Top](#)

## How to read search results

TamBiC was designed is to provide you with results that are clear and easy to read. The diagram below points out the features that are important for understanding the search results page:

The screenshot shows the TamBiC search interface. At the top, there are three search tabs labeled 'Search 1', 'Search 2', and 'Search 3', each with a close button (X). Below the tabs is a search bar containing the word 'play'. The main area displays the results of the search in the 'English to Finnish' corpus, showing 115 hits. There are buttons for 'Save' and 'Print'. Below this, it says 'Used keywords: play. (English)'. There are two red numbers, 4.5 and 4.4. Below that, there is a section for refining the search, labeled 'Make a new search within your search results:'. It includes a 'Select language:' dropdown menu with 'English' selected, a 'Search' button, and a checkbox for 'Distinguish upper and lower case.' There are red numbers 3.2, 3.1, and 3.3. Below this, it says 'Showing results: 1 - 10 ( 115 )'. There are navigation buttons (left arrow, right arrow, 1, 2, 3, 4, 5, right arrow) and a 'Show all context' button. Below this, there are two search results. The first result shows the original text 'he said, "My dear, tell me what games you play..."', the code '(GREENE 2:1a:278)', and the translation 'hän sanoi: - Kultaseni, kerro millaisia leikkejä sinulla on...'. The second result shows the original text '"It's a game children play.', the code '(GREENE 2:1a:307)', and the translation '- Se on eräs lasten leikki.'. There are red numbers 4.1, 4.2, 4.3, and 4.4. There are also red circled numbers 1, 2, 3, 4, and 5.

### 1. Format of results:

The first line of any search result is in the original language of the corpus selected. If you have searched for a word in the translations, the original language is still given first.  
The second line is a code that identifies the text (see About TamBiC: Sources).  
The third line is a translation of the first line.

<b>Original language text-</b>	The group immigrated to Western Canada in the late 1800s to escape brutal persecution.
<b>ID code-</b>	(PRESS 1:1:11)
<b>Translated text-</b>	Ryhmä muutti Länsi Kanadaan 1800 luvun lopulla paetakseen julmaa vainoa.

Note that : Sentences omitted in the published translation are indicated by three asterisks: \*\*\*.

### 2. Search results:

The results of each search are stored under their own "tabs". To delete the results of a particular search, simply click on the X of the tab in question. If you delete the main search tab (Search 1, Search 2, etc), this will also delete all the search results stored under it.

### 3. Refining your search:

When you have made a "basic search", it is possible to perform further searches within the results of the first search.

3.1 **Search box:** This is where you type your search query.

3.2 **Language selection:** As with "basic searches", you can search in either the English or the Finnish lines. Simply click on the language shown to select the other language.

3.3 **Distinguish upper and lower case letters:** This will allow you to have case sensitivity in refined searches, as in the basic searches (see above).

It is also possible to further refine the results of a refined search. For tips on the use of refined search see below.

### 4. The navigation area:

This allows you to go through the search results shown bellow. In addition it gives the information of the word which the query was performed.

4.1 **Showing results:** 1-10 (199) This means that the first 10 results (out of 199 altogether) are displayed on the current page. To view more results, use the navigation buttons (see 4.3 below).

4.2 **Show All Context button:** This displays the previous and following sentences for *all* the search results.

4.3 **Navigation buttons:** Click here to navigate to the following or previous page of the results.

4.4 **Print:** For printing (using the on-screen format of the website).

4.5 **Save:** This enables you to save the currently displayed search results on your computer. The saved file

will be in .rtf format, and can be renamed if you so wish.

5. **More context:** This displays the previous and following sentences for a specific search result.

Familiarity with all these features will help you get the most out of the search engine.

[Back To Top](#)

## TamBiC search basics: More search help

The basic search procedure (above) covers all the most common issues, but sometimes you may need more power. Here we highlight the advanced features of TamBiC search engine.

[Back To Top](#)

### Wild cards (\*)

This is a very powerful feature. If you include \* within a query, the asterisk is taken to stand for any other letter(s). This enables you find various forms of the search word (with inflections, prefixes, suffixes) in addition to the basic search word.

For example:

- Word containing something else at the end- **part\*** (**part, parts, parting, parted**)
- Word containing something else at the beginning- **\*part** (**part, depart, impart**)
- Word containing something else at the beginning and at the end- **\*part\*** (**part, parts, etc; depart, impart, etc; departing, compartment, etc**)

Note that if a wild card is typed between spaces, this will be treated as a whole word:

**I was \* working** (finds: **I was not working, I was always working**, etc)

**They \* \* leave** (finds: **they did not leave, they wanted to leave**, etc)

[Back To Top](#)

### Operands (OR, AND, THEN, NOT, "...")

These operands will again allow you to make more powerful searches. Note that you have to type them in CAPITALS.

- **AND** - If you place the **AND** operand between two words, this will find sentences containing both words. The words can appear be in any order.  
he AND man (finds: **he** was a **man** / **he** saw a **man**, etc)
- **OR** - This allows you to find any one of several words. For example, **go OR went** will find sentences containing either one of these words.  
go OR went (finds: they often **go** there / we often **went** there, etc)
- **NOT** - If you place **NOT** immediately before a word, this indicates that you want to exclude sentences containing this word. The **NOT** operand only affects the word which follows it.  
depart\* NOT department (only finds: **depart, departs, departed, departing**)
- **THEN** - If you place **THEN** immediately before a word, TamBiC will find sentences containing both words *in that order*.  
talk **THEN** about (finds: **I shall talk to him about it**)
- **"..."** - Quotation marks can be used to find an exact phrase:  
"as soon as" (finds sentences containing the phrase **as soon as**.)
- **More than one operand** - It is possible to use the same operand twice in the same search (e.g. **dog OR cat OR horse; sing\* NOT singer NOT single**; etc).  
Different operands, on the other hand, should not be used in the same search query. Instead, perform two separate searches.  
"..." Phrases in quotation marks may be combined with the other operands: "in the evening" OR "at night"; "as \* as" NOT "as well as" NOT "as soon as"

[Back To Top](#)

## Distinguish upper and lower case letters

The default is that the search program looks for both upper and lower case (**Then, then**); however, if **Distinguish upper and lower case letters** is selected, words with initial capital letters can be searched for separately (**This... / President**, etc).

[Back To Top](#)

## Refining

This enables you to carry out further searches using the current display results. Note that it is possible to search in either language by selecting the appropriate language (i.e. the original or the translation).

It is also possible to use the same features as in the basic search, i.e. operands, case sensitivity and wild cards. The results will be then be displayed and stored under a separate tab within the basic search. You can further refine the results of a previous refined search, and you may do so as many times as you wish.

A simple example:

To find and organise data on the main English translations of the Finnish noun **taivas**

Basic search:

Select corpus: **Finnish-English**

Select language: **Finnish**

Search word: **taiva\*** (to find **taivas**, **taivaassa**, **taivaalla**, etc)

1st refined search:

Select tab: **taiva\***

Select language in refined search box: **English**

Search word: **sky**

2nd refined search:

Select tab: **taiva\***

Select language in refined search box: **English**

Search word: NOT **sky** OR **heaven**

(Using the **NOT** operand alone removes sentences containing unwanted words and stores the remaining sentences.)

[Back To Top](#)

---

## Frequently Asked Questions

### 1. Why was the TamBiC corpus created?

The corpus was created as a tool for students and scholars interested contrastive linguistics. It is also a useful resource for translators.

[Back To Top](#)

### 2. What kind of texts can I find in the TamBiC corpus?

The corpus is divided in two parts:

- Fiction sub-corpus: containing long extracts (15,000 words/50 pages) from 16 English novels (plus their Finnish translations) and similar extracts from 16 Finnish novels (plus their English translations).
- Non-fiction sub-corpus: consisting of long extracts (and their translations) from the following: books on history, art,

music, science, nature and linguistics; tourist guides, cookery books and instruction manuals; and a large number of newspaper articles.

[Back To Top](#)

### 3. How can TamBiC corpus help me?

If you are interested in any aspect of cross-language research (whether morphological, syntactic, lexical or pragmatic) involving English and Finnish, the corpus provides a useful database which can be searched easily. In addition, translators will find it useful both as a look-up facility and as a resource for researching problems of translation theory.

[Back To Top](#)

### 4. Who can use the TamBiC corpus and how much does it cost to use?

The corpus is freely available to all the students, teachers and researchers in the School of Modern Languages and Translation Studies at the University of Tampere. It has been compiled for the purpose of academic research, not for commercial profit.

[Back To Top](#)

### 5. What is the format of the texts in the TamBiC corpus?

The corpus has the following fixed format:

The Sons of Freedom are a breakaway sect of the Doukhobor faith, an obscure group of Russian Christians whose beliefs include pacifism and total equality.

(PRESS 1:1:10)

Vapauden veljet on Doukhobor uskosta irtautunut lahko, hämäräperäinen ryhmä venäläisiä kristittyjä joiden uskomuksiin kuuluu rauhanaate ja täydellinen tasa arvo.

Searches can be made in either the English or the Finnish lines. The search results contain both lines: the sentence containing the search word and the corresponding sentence in the other language.

[Back To Top](#)

### 6. Where can I find more information about the texts used the TamBiC corpus?

Visit the [About TamBiC](#) section to view detailed source information for the texts used in the corpus.

[Back To Top](#)

### 7. Who compiled the corpus?

The corpus was compiled by [Robert Cooper](#) from the English Department in [the School of Modern Languages and Translation Studies](#) at the [University of Tampere](#).

[Back To Top](#)

### 8. Who designed the search engine?

The search engine was designed by as part of a special project organized by the Department of Computer Sciences at the University of Tampere. The project group members were Reeta Karjalainen, Georgina Monjaraz Gomez, Pekka Ihalainen, Xiaozhou Li, Harri Mustonen, Simone Strasser and Juhani Vainio.

[Back To Top](#)

For information on the background to the corpus see Cooper W. R. (1998) "The Tampere Bilingual Corpus of Finnish and English: Development and Applications." In Cooper, W. R. (ed.) *Compare or Contrast? Current Issues in Cross-Language Research*. Tampere: University of Tampere.

