



User Manual

Tambic User Manual

The Tampere Bilingual Corpus of Finnish and English user
manual



Feb 2010

Contents

| | |
|--|----|
| Table of figures | 2 |
| Basic Search Procedure | 3 |
| How to start..... | 3 |
| Some basic facts | 3 |
| How to read the main page..... | 4 |
| How to read the search results | 5 |
| More search help | 7 |
| Wild cards (*)..... | 7 |
| At the end of the word | 7 |
| At the beginning of the word | 8 |
| At the beginning and at the end of the word..... | 8 |
| Within the word | 9 |
| Operators (AND, OR, NOT) | 10 |
| AND..... | 10 |
| OR | 11 |
| NOT..... | 12 |
| Combination of the operators..... | 13 |
| Distinguish upper- and lowercase letters..... | 13 |
| New search within search results..... | 14 |
| How to do a refined search | 14 |
| Features of refined search..... | 15 |

Table of figures

| | |
|---|----|
| Figure 1: start page..... | 3 |
| Figure 2: page description | 4 |
| Figure 3: search results for word cucumber | 5 |
| Figure 4: format of results..... | 5 |
| Figure 5: tab..... | 6 |
| Figure 6: search within search results | 6 |
| Figure 7: navigation area..... | 6 |
| Figure 8: more context button | 7 |
| Figure 9: wildcard at the end..... | 7 |
| Figure 10: excerpt of search result with wildcard at the end | 7 |
| Figure 11: wildcard at the beginning..... | 8 |
| Figure 12: excerpt of search result with wildcard in the beginning..... | 8 |
| Figure 13: wildcard at the beginning and at the end | 8 |
| Figure 14: excerpt of search result with wildcard at the beginning and at the end of the word | 9 |
| Figure 15: wildcard within | 9 |
| Figure 16: excerpt of search result with wildcard within the word | 10 |
| Figure 17: search with AND operator..... | 10 |
| Figure 18: excerpt of search result with AND operator | 11 |
| Figure 19: search with OR operator | 11 |
| Figure 20: excerpt of search result with OR operator | 12 |
| Figure 21: search result with NOT operator | 12 |
| Figure 22: search with combined operators | 13 |
| Figure 23: search result with combined operators | 13 |
| Figure 24: search with an uppercase letter..... | 13 |
| Figure 25: excerpt of search result with an uppercase letter | 14 |
| Figure 26: area of refined search | 14 |
| Figure 27: refined search..... | 15 |
| Figure 28: result of refined search | 15 |



Basic Search Procedure

A Search within Tambic is simple. In the following paragraphs you will learn how to use Tambic for simple and complex search queries.

How to start

Following the link <http://www.cs.uta.fi/~tambic/> you will see the welcome page of the Tambic. See Figure 1. To start using Tambic please follow the link “Start using TamBiC!”.



Figure 1: start page

Some basic facts

- Every word matters. Generally, all the words you put in the query will be used. With exception of the operators.
- Search is always case insensitive. Searching for dog is the same as searching for Dog .If not otherwise stated. (See Distinguish upper- and lowercase letters)
- Exceptions are: punctuation is ignored (@#\$\$%^&*()=+[]\ and other special characters).

How to read the main page

The diagram below points out the features that are important to understanding the main page, and that you will use while performing a search:



The screenshot shows the TAMBiC search interface. At the top left is the TAMBiC logo with the text 'The Tampere Bilingual Corpus of Finnish and English'. To the right are two buttons: 'Help' (annotated with 4) and 'About' (annotated with 5). Below the logo, there are two dropdown menus: 'Select corpus:' (annotated with 1) showing 'English to Finnish' and 'Select language:' showing 'English'. Below these is a large text input box (annotated with 2). Under the input box is a checkbox (annotated with 3) labeled 'Distinguish upper- and lowercase letters.'. At the bottom is a 'New Search' button.

Figure 2: page description

1. **The corpus selection:** The first line determines the original language of the corpus, the second line determines the line inside the corpus you will be searching for. By clicking either of this the selection changes and is possible to select other language.
2. **The search box:** The area where you type the query you are searching for.
3. **The case sensitivity:** This will allow you to have case sensitivity, so searching for dog is NOT the same as searching for Dog when this is selected.
4. **Help button:** A link to a help file that will solve your questions of how to use the corpus. Click here if the feature that you wanted isn't available or clear for you.
5. **About button:** A link to the information file that will tell you more about the Tambic corpus, the authors and the texts used in the corpus. Click here to know more about us!



How to read the search results

Tambic's goal is to provide you with results which are clear and easy to read.

In our example we want to find any entry which contains the word “cucumber”. No matter how it is written. So first we select the corpus **English to Finnish** and after that we select the language of the word, we just typed in, which is **English**. After clicking the Button **New Search**, we will get the search results shown in Figure 3.

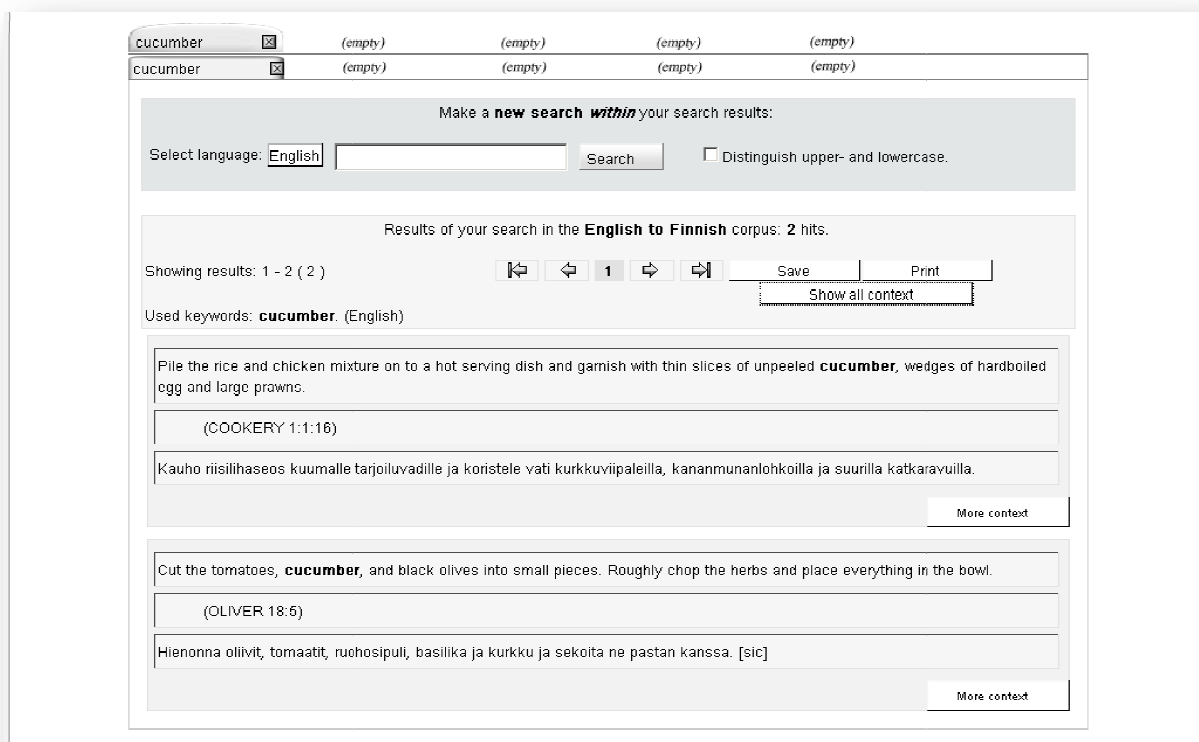


Figure 3: search results for word cucumber

- **Format of results:** The first line of any search result is the original language of the corpus you selected, followed by a code that identifies the text and finally the translated version of the text. For each hit.

| | |
|--|------------------------|
| Pile the rice and chicken mixture on to a hot serving dish and garnish with thin slices of unpeeled cucumber , wedges of hardboiled egg and large prawns. | Original language text |
| (COOKERY 1:1:16) | ID code |
| Kauho riisilihaseos kuumalle tarjoiluvadille ja koristele vati kurkkuviipaleilla, kananmunanlohkoilla ja suurilla katkaravuilla. | Translated text |
| More context | |

Figure 4: format of results

The tabs: Help you to navigate through the searches. To close a tab click on the X of the tab you wish to close, for closing a whole group close the main search tab. A maximum of 5 tabs can be open for the basic search and for the refined search (See Distinguish upper- and lowercase letters) as well.

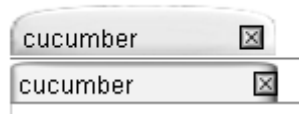


Figure 5: tab

- **The search within search area:** This allows you to perform a new search based on the results from the first search. In addition it gives the information of the corpus language where the search was performed.

Make a **new search** *within* your search results:

Select language: ☐ Distinguish upper- and lowercase.

Figure 6: search within search results

- **Search box:** The area where you type the query you are searching for.
 - **Language selection:** This determines the line inside the corpus you will be searching for. By clicking in it the selection changes and is possible to select other language.
 - **Upper- and lowercase letters:** This will allow you to have case sensitivity in the refined results, so searching for dog is NOT the same as searching for Dog when this is selected.
- **The navigation area:** This allows you to go through the search results shown below. In addition it gives the information of the word which the query was performed.

Results of your search in the **English to Finnish** corpus: 2 hits.

Showing results: 1 - 2 (2)

Used keywords: **cucumber.** (English)

Figure 7: navigation area

- **Show All Context button:** Allows you to display the previous and next sentence from all the hits.
- **Navigation buttons:** Click here to navigate to the next and previous page of the results.
- **Print:** For printing using the format of the website.
- **Save:** Is possible to save the results in .rtf directly in your computer for further analysis.



- More contexts: Allows you to show the previous and next line from a specific hit.

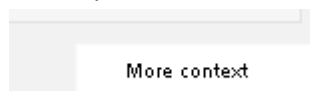


Figure 8: more context button

More search help

The Basic search help covers all the most common issues, but sometimes you need more power. Here we will highlight the more advanced features of Tambic Search Corpus.

Wild cards (*)

This is a very powerful feature. If you include * within a query, it tells the search to try to treat the star as a placeholder for any unknown letter(s) and then find various other forms of the search word (inflections, prefixes, suffixes) in addition to the basic search word.

There are 4 possibilities to use the wildcard *:

At the end of the word

Select corpus: Select language:

☐ Distinguish upper- and lowercase letters.

Figure 9: wildcard at the end

This will find also words like shown in Figure 10

| |
|---|
| Perhaps the girl was part of it! |
| (ORWELL 2:1:66) |
| Kukaties tyttö kuuluikin siihen! |
| More context |

| |
|---|
| He was particularly enthusiastic about a papier mâché model of Big Brother's head, two metres wide, which was being made for the occasion by his daughter's troop of Spies. |
| (ORWELL 2:1:87) |
| Hän oli erityisen innostunut paperimassasta valettuun Ison Veljen päähän, josta tulisi kaksi metriä leveä ja joka oli tuota juhlaa varten tekeillä hänen tyttärensä Vakooja-ryhmässä. |
| More context |

| |
|---|
| in such a way as to cast discredit on a prominent member of the Inner Party , who was now under a cloud. |
| (ORWELL 2:1:96) |
| sillä tavalla, että epäilyksen varjo lankeaisi erääseen Sisärenkaan huomattavaan jäseneseen, joka oli joutunut epäsuosioon. |
| More context |

Figure 10: excerpt of search result with wildcard at the end

At the beginning of the word

Select corpus: Select language:

☐ Distinguish upper- and lowercase letters.

Figure 11: wildcard at the beginning

This will find also words like shown in Figure 12.

| |
|---|
| part of the pedestal had been broken off and lay in the middle of the road - the mule went round it. |
| (GREENE 2:1c:66) |
| Jalustasta oli lohjennut osa, joka virui keskellä tietä - muuli kiersi sen. |
| More context |

| |
|--|
| They were perhaps four metres apart when the girl stumbled and fell almost flat on her face. |
| (ORWELL 2:1:10) |
| He olivat kenties neljän metrin päässä toisistaan, kun tyttö kompastui ja kaatui pitkäkseen lattialle. |
| More context |

Figure 12: excerpt of search result with wildcard in the beginning

At the beginning and at the end of the word

Select corpus: Select language:

☐ Distinguish upper- and lowercase letters.

Figure 13: wildcard at the beginning and at the end

The result can look like shown in Figure 12.

| |
|--|
| It was a common accident in the Fiction Department . |
| (ORWELL 2:1:9) |
| Sellaiset tapaturmat olivat Kirjallisuusosastossa jokapäiväisiä. |
| More context |

| |
|--|
| They were perhaps four metres apart when the girl stumbled and fell almost flat on her face. |
| (ORWELL 2:1:10) |
| He olivat kenties neljän metrin päässä toisistaan, kun tyttö kompastui ja kaatui pitkäkseen lattialle. |
| More context |

| |
|---|
| Perhaps the girl was part of it! |
| (ORWELL 2:1:66) |
| Kukaties tyttö kuuluikin siihen! |
| More context |

Figure 14: excerpt of search result with wildcard at the beginning and at the end of the word

Within the word

Select corpus: Select language:

☐ Distinguish upper- and lowercase letters.

Figure 15: wildcard within

An excerpt of the results is shown in Figure 16.

| |
|--|
| He sighed: it had been quite a good case: one more relic of the quiet past ... |
| (GREENE 2:1b:461) |
| Hän huokasi, sillä se oli ollut oikein hyvä laukku, sekin oli rauhallisen menneisyyden pyhäinjäännöksiä... |
| More context |

| |
|--|
| "Mula, Mula," the priest said, urging the mule on, past the inevitable bandstand and a statue in florid taste of a woman in a toga waving a wreath: |
| (GREENE 2:1c:64) |
| - Mula. Mula, pappi sanoi hoputtaen muulia edelleen, ohi välttämättömän soittolavan ja runsaskoristeiseen tyyliin tehdyn kuvapatsaan, joka esitti togaan puettua, seppelettä heiluttavaa naista. |
| More context |

| |
|---|
| part of the pedestal had been broken off and lay in the middle of the road - the mule went round it. |
| (GREENE 2:1c:65) |
| Jalustasta oli lohjennut osa, joka virui keskellä tietä - muuli kiersi sen. |
| More context |

Figure 16: excerpt of search result with wildcard within the word

Operators (AND, OR, NOT)

In Tambic you can use three operators to make a more precisely search. You have to type the operator in, in capital letters, so that Tambic recognize it as an operator. These 3 operators are

AND

Combining two words with an AND operator will tell Tambic to find phrases just containing these two words.

Select corpus: Select language:

☐ Distinguish upper- and lowercase letters.

Figure 17: search with AND operator

In Figure 18 you can find an excerpt of the search result with an AND operator.

| |
|--|
| They were so unexpected that he grinned at them - with the shy, evasive, untrustworthy smile of an man caught out. |
| (GREENE 2:1a:15) |
| näky oli niin odottamaton, että hän naurahti sille - ilmi tulleen miehen arkaa välttelevää epäluotettavaa hymyä. |
| More context |

| |
|---|
| It was as if he had descended by means of his sin into the human struggle to learn other things besides despair and love, that a man can be unwelcome even in his own home. |
| (GREENE 2:1a:100) |
| Oli kuin hän syntinsä kautta olisi laskeutunut keskelle ihmisten taistelua oppimaan muutakin kuin epätoivoa ja rakkautta, oppimaan, ettei ihminen aina ole tervetullut omaan kotiinsa. |
| More context |

| |
|--|
| he was going to say "my children", but then it seemed to him that only the childless man has the right to call strangers his children. |
| (GREENE 2:1a:116) |
| hän oli sanoa 'lapseni', mutta sitten hänestä tuntui, että vain lapsettomalla ihmisellä on oikeus sanoa vieraita lapsikseen. |
| More context |

Figure 18: excerpt of search result with AND operator

OR

Combining two words with the OR operator will find all phrases which contain one of the word or both.

Select corpus: Select language:

☐ Distinguish upper- and lowercase letters.

Figure 19: search with OR operator

Result may look like in Figure 20

| |
|---|
| He has avoided it for six years, but now it wasn't his fault - it was his duty to go there - it couldn't count as sin. |
| (GREENE 2:1a:23) |
| Hän oli vältellyt sitä kuuden vuoden ajan, mutta nyt tämä ei ollut hänen syynsä - hänen velvollisuutensa oli mennä sinne - sitä ei voitu pitää syntinä. |
| More context |

| |
|---|
| He went back to his mule and kicked it gently, |
| (GREENE 2:1a:24) |
| Hän palasi muulinsa luokse ja potkaisi sitä ystävällisesti. |
| More context |

Figure 20: excerpt of search result with OR operator

NOT

Using the NOT operator before a search word, will exclude it.

Select corpus: Select language:

☐ Distinguish upper- and lowercase letters.

Figure 21 shows the search results.

| |
|--|
| While they finished their preparations to depart they stepped now and then across him, |
| (FAULKNER 10:7) |
| Noiden muiden saattaessa päätökseen matkavalmistelujaan he astuivat muutamia kertoja hänen ylitseen, |
| More context |

| |
|--|
| But the voles are beginning to disappear below ground to hibernate and most of the small birds are starting to depart southwards. |
| (BIRDS 2:312) |
| Mutta maaoravat ja muutamat muutkin jyräjät alkavat hakeutua talvilevolle, päivän valoisa saalistusaika lyhenee ja useimmat linnut aloittelevat muuttomatkaansa etelään. |
| More context |

Figure 21: search result with NOT operator



Combination of the operators

Of course you can also combine the operators if you want or if you have to.

Select corpus: Select language:

☐ Distinguish upper- and lowercase letters.

Figure 22: search with combined operators

This query gives us one result which is shown in Figure 23.

| |
|--|
| She had a weak bladder and wet her knickers a lot which made her unpopular at school - a graffitied, run-down place - with both teacher and pupils. |
| (WELDON 17:216) |
| Hänellä oli huono rakko ja hän kasteli usein housujaan mikä teki hänet syrjityksi koulussa - seinäraapustusten sotkemassa, rapistuneessa paikassa - niin opettajien kuin oppilaittenkin keskuudessa. |
| More context |

Figure 23: search result with combined operators

Distinguish upper- and lowercase letters

The default setting of Tambic is to search the words case insensitive. This means that no matter how you enter the words (capital letters, small letters, mixed capital and small letters) it will find them all. But maybe in some special search reason you want just to find the words which start with a capital letter. In this case you can check the ☒ Distinguish upper- and lowercase letters. and Tambic will return just the phrases containing the word with capital letters.

Let us do one example to make it clearer.

Select corpus: Select language:

☒ Distinguish upper- and lowercase letters.

Figure 24: search with an uppercase letter

We will get just phrases which contain the word "School". See Figure 25. Phrases which contain the word "school" will not be returned in the result.

| |
|---|
| helping to produce booklets in sealed packets with titles like Spanking Stories or One Night in a Girls' School , to be bought furtively by proletarian youths |
| (ORWELL 2:3:96) |
| valmistamassa sinetöidyissä pakkauksissa lähetettäviä kirjoja, joiden niminä oli sellaista kuin "Mainioita Juttuja" tai "Yö tyttö-koulussa" |
| More context |

| |
|--|
| He had Wednesday afternoon off to go to the Art School - Miss Jordan's provision - returning in the evening. |
| (LAWRENCE B:391) |
| Hän sai keskiviikkoiltapäivät lomaa käydäkseen taidekoulua - tämä oli neiti Jordanin aikaansaannos - ja palasi illalla työhön. |
| More context |

| |
|--|
| and had won all the prizes that could be won at the Royal College of Organists and at the Royal School of Church Music. |
| (MUSIC 2:12) |
| ja oli voittanut kaikki mahdolliset palkinnot. [sic] |
| More context |

Figure 25: excerpt of search result with an uppercase letter

New search within search results

This enables you to carry out further searches within your search results.

How to do a refined search

The refined search takes place in the area marked with a red rectangle in Figure 26.

Select corpus: Select language:

☒ Distinguish upper- and lowercase letters.

| | | | | | | |
|------------|-------------------------------------|---------|-------------------------------------|---------|---------|---------|
| he AND man | <input checked="" type="checkbox"/> | School | <input checked="" type="checkbox"/> | (empty) | (empty) | (empty) |
| School | <input checked="" type="checkbox"/> | (empty) | (empty) | (empty) | (empty) | (empty) |

Make a **new search within** your search results:

Select language: ☐ Distinguish upper- and lowercase.

Figure 26: area of refined search

For every basic search a tab will be created in level 1. For every refined search a tab will be created as well in level 2.



Let us do an example. In the basic search we were searching for “he AND man”. Now we want to search within the “he AND man”-results make another search. We want to find all phrases which also contain the word “childless”. So we type in the word “childless” in the refined search field and press the Button Search.

Figure 27: refined search

Figure 28: result of refined search

Following the red arrows in Figure 28 we can see that our basic search was “he AND man” and the refined search “childless” appears in a tab in level 2.

Features of refined search

The following features can be used for the refined search:

- Language selection of the search word
- Distinguish upper- and lowercase (See also Distinguish upper- and lowercase letters)
- Wildcard (*) search (See also Wild cards (*))
- Operators AND, OR, NOT (See also Operators (AND, OR, NOT))