

# **CATMA User Manual**

for CATMA 3.0

**by Lena Schüch (B.A.)**

Hamburg University, June 2010

[www.slm.uni-hamburg.de/catma](http://www.slm.uni-hamburg.de/catma)

<b>A brief non-technical introduction to CATMA .....</b>	<b>3</b>
<b>1. Start .....</b>	<b>5</b>
1.1 Load texts .....	5
1.2 Reload texts .....	10
<b>2. Tagging .....</b>	<b>11</b>
2.1 Create Tags .....	11
2.2 Tag all occurrences .....	14
2.3 Edit Tags .....	16
2.4 Remove tags.....	17
2.5. Insert property .....	19
2.6 Remove and edit property.....	21
2.7 Create new tagsets.....	22
2.8 Remove or edit tagsets.....	24
<b>3. Wordlist.....</b>	<b>25</b>
3.1 Display selected words in context .....	27
3.2 Compute Distribution Analysis.....	28
3.3 Compute Collocation Analysis .....	30
<b>4. Queries .....</b>	<b>32</b>
4.1 Execute queries .....	32
4.2 The QueryBuilder .....	33
4.2.1 Search “by word or phrase” .....	34
4.2.2 Search “by grade of similarity” .....	36

4.2.3 Search “by tag” .....	37
4.2.4 Search “by property” .....	38
4.2.5 Search “by collocation” .....	39
4.2.6 Search “by frequency” .....	40
4.2.7 Search “by group” .....	41
4.3 Complex Queries with the QueryBuilder.....	42
4.3.1 Bracket particularities of complex queries.....	44
4.4 Query syntax.....	46
4.4.1 Basic Queries .....	46
4.4.3 Queries with Groups .....	47
4.4.4 Queries with Phrases .....	48
4.4.5 Queries by Frequency .....	49
4.4.6 Queries by Similarity .....	49
4.4.7 Queries by Tag.....	50
4.4.8 Complex Queries .....	50
4.4.9 Combining Results .....	50
4.4.10 Searching by Collocation .....	50
4.4.11 Exclusion .....	51
4.4.12 Adjacency .....	51
4.4.13 Queries with Refinements .....	52
4.4.14 Summary of Query Operators .....	53
<b>5. Groups.....</b>	<b>54</b>
5.1 Create groups.....	54
5.2 Delete or edit groups.....	55

## A brief non-technical introduction to CATMA

CATMA (Computer Aided Textual Markup and Analysis) is an open source software with a focus on textual markup and analysis. If you know how to open a file on your computer and point and click with a mouse, that's all you need to use CATMA. And if you come from a traditional literary research background, all the better: CATMA's workflow tries to emulate that of a traditional literary scholar who reads a text, makes some annotations, compares and analyses the notes and then uses these results to interpret the text. In other words, our main goal in developing CATMA was to create a practical and intuitive tool for scholars, students and other parties with an interest in literary research.

With this user group in mind we realized that CATMA must also pay tribute to one of the key methodological tenets of literary scholarship: the principal open-endedness of any research into the semantics of literary texts. This required designing a tool that, while based on digital technology which models its objects in terms of well-defined sequences of 1s and 0s, will nevertheless respect and support the ambiguity, variance, the loops and communal as well as historical nature of how our human mind interprets the more complex features of our world.

Literary texts can certainly be counted among these – in fact, they themselves are arguably the most complex representations of what we experience as “the world”. And it is this complexity that makes literature particularly relevant to us. But wherein lies this relevance? As ordinary readers, literary texts interest us mainly because we consider them *meaningful* – that is, because the information contained in, as well as the images and emotional reactions triggered by them can be related to our individual life experience. This experience includes not only what we witness with our senses, but also what we construct in our minds, such as hopes, fears, beliefs etc. And in the end it is this subjective experience that stimulates our need to interpret a text.

As literary scholars, however, we accept that there is no single and clear-cut definition as to what a given text's meaning might be. Rather, what an individual reader identifies as „the meaning“ which a text has for him or her will depend on a number of factors, and not just on the content, i.e. on „what the text is about“. These factors are in part contextual, such as the historical circumstances of text production, tradition, and reception; the relevant social dimension; aesthetic conventions and cross-references among texts and genres; motivational and psychological determinants that govern what and how the individuals and groups involved in the exchange of and discourse about literature will act, etc.. In addition a broad variety of text-immanent factors has to be accounted for, such as stylistic and rhetorical features or structural patterns, among other. What we do when we interpret a text as literary scholars is to correlate these two domains in a theory guided, methodical way: that of the empirical textual phenomena, and that of the contextual contingencies to which these phenomena might plausibly point.

The major function of markup and analysis tools such as CATMA is to support us in this process. However, where literary (non-pragmatic) texts are concerned, textual markup and analysis more specifically requires tools that *will allow us to generate, exchange and exploit complementary and even competing descriptions of the textual phenomena which we observe and analyse*. These phenomena, which are often

embedded at the surface level of words and sentences, constitute our starting point. Surface phenomena in a text present as reliable and objective a material base as one could wish for – but they, too, must in the end be kept open to (re-)interpretation.

In CATMA these general considerations have translated into three major functional characteristics.

1. **Keep the source text and its associated interpretations separate:** CATMA applies so-called *stand-off markup*. This means that your description or definition of what you find in the literary text is written into a separate file, and not into the source material itself. A key advantage of this is that one source text can have numerous associated markup-files. Because all markup files reference back to an identical and unchanged source text, the markup itself is no longer considered *the* description, but *a* description of the object at hand. In CATMA we actually generate not one particular, authoritative markup for a given text, but rather *one to n instances of markup* that may be compared, amalgamated, edited, extended, etc.
2. **Integrate textual markup and analysis:** In traditional literary scholarship we constantly oscillate between the description of textual phenomena, and their interpretation. Often a new interpretation will result in a need for more detailed analysis and description, and vice versa. This is why CATMA was designed to allow for seamlessly switching back and forth between description (markup) and interpretation (analysis).
3. **Keep everything open and reusable:** CATMA is XML-based and, from its overall JAVA based programming approach down to the use of the very flexible TEI concept of so-called *feature structure tags*, is built in such a way that whatever you generate in CATMA can be exported to other standard conformant applications – from an individual tag right up to an entire set of markup documents. So even if you do not know what the abbreviations XML (Extended Markup Language), JAVA (a programming language) and TEI (Text Encoding Initiative) actually stand for, their importance for you as a user of CATMA is that you are not locked into a technology that will make your work obsolete once you decide to move on to something else.

The current manual describes how to use CATMA 2.01 which can be run on almost any modern computer with a Windows, Mac or Linux operating system. While the development of CATMA continues (a big ‘Thank you!’ to all the users who have given us valuable feed back!) we will make sure that whatever markup you have generated with the current version will remain usable in later versions of the software.

We hope you will find CATMA useful – please do not hesitate to contact us with any comments, requests, bug alerts etc.. For contact details and more information on CATMA please visit our website at [www.slm.uni-hamburg.de/catma](http://www.slm.uni-hamburg.de/catma)

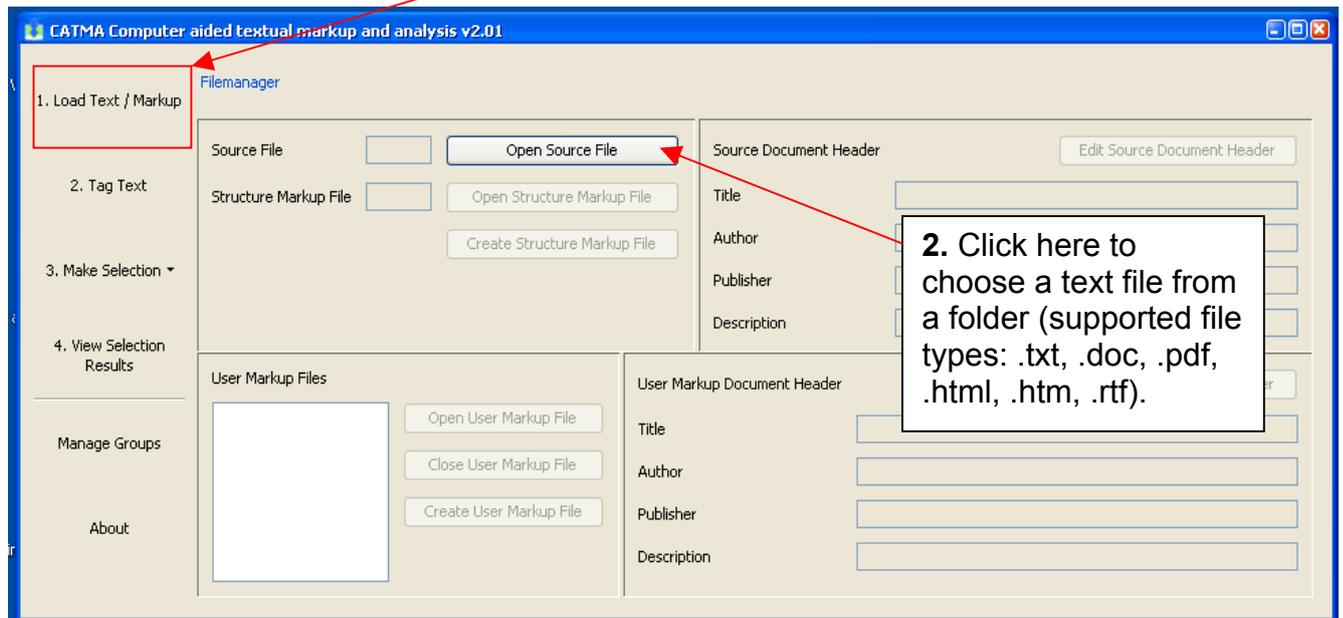
Jan Christoph Meister  
for CATMA development team

Hamburg, 26 February 2010

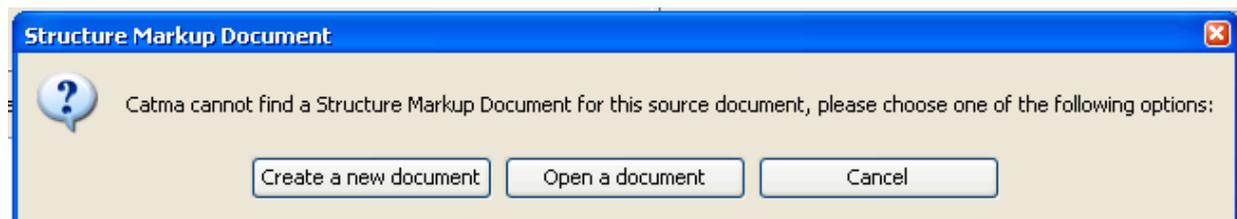
# 1. Start

## 1.1 Load texts

After having started the program, click “1. Load Text/ Markup” in the menu on the left-hand side.

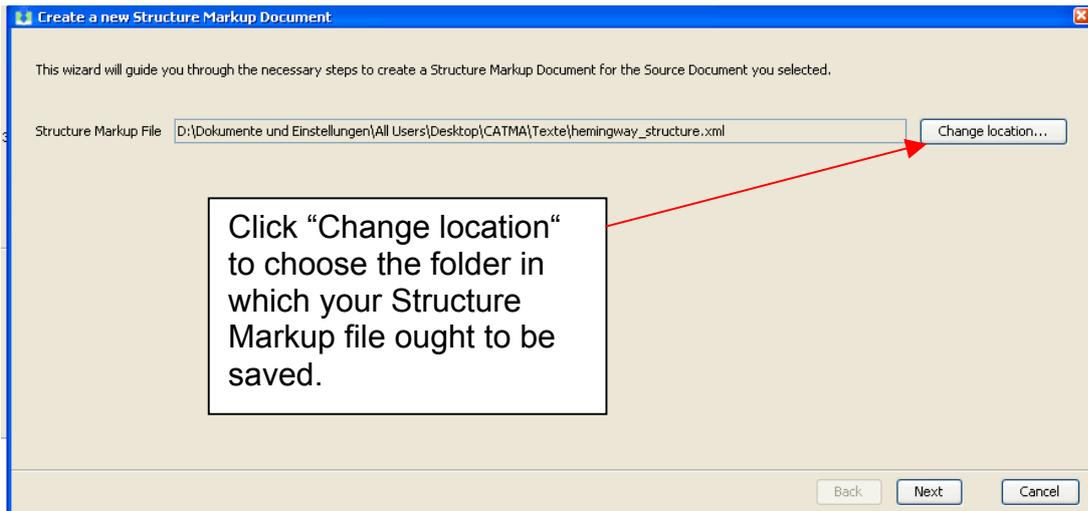


After having chosen a text file, a dialogue window opens:



The Structure Markup Document is the file in which information about your text, such as the file type and the text structure will be saved.

If you have not worked on your chosen text file in CATMA before, please select “Create a new document”.



In the dialogue window that opens (shown above), you can choose the folder in which your Structure Document will be saved. It is recommended to save the associated text files in the same folder. Please click “Next” to continue.

CATMA tries to determine the file type you have chosen so that it displays the characters of your text correctly. If it does not choose the correct encoding automatically, you can try choosing a different encoding as shown below.

1. Determine the file type (TEXT, DOC, PDF, HTML, HTM or RTF).
2. Check if your text is displayed correctly in the “Preview” below.
3. If your text is not displayed correctly, select an encoding here.
4. Click “Refresh Preview” to check if your text is displayed correctly. If not, please try a different encoding.

In the next section you may specify conditions for the creation of your word list. For example, if your source text is in French you may not want contractions that contain an apostrophe (like “j’ai”) to appear as one single entry in the word list.

**1.** Tick off this box if you would like to display all of those words separated that contain an apostrophe (e.g. you’ll will appear as “you” and “ll” in the word list).

**2.** Please select the predominant language of your source text so that the word list can be created in correspondence to the peculiarities of this language.

**3.** If there are contractions or character sequences you do not want to separate, you can enter them here. Please click “Add entry” to extend your list.

**3a.** You can also save the list of words you have entered or load lists you have created before.

In this dialogue window, you may enter information about the text you have chosen – but this is not obligatory. Please click “Next” to continue.

**Create a new Structure Markup Document**

Source Document Header

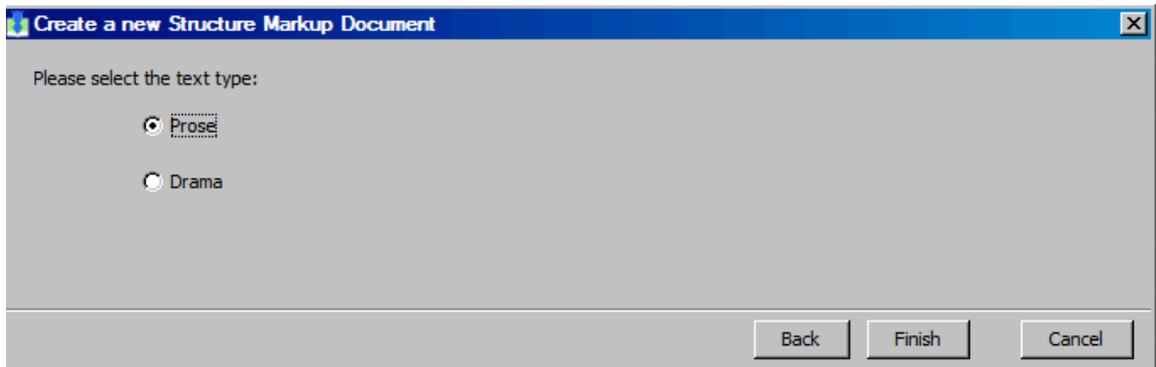
Title: Ernest Hemingway

Author: Hills Like White Elephants

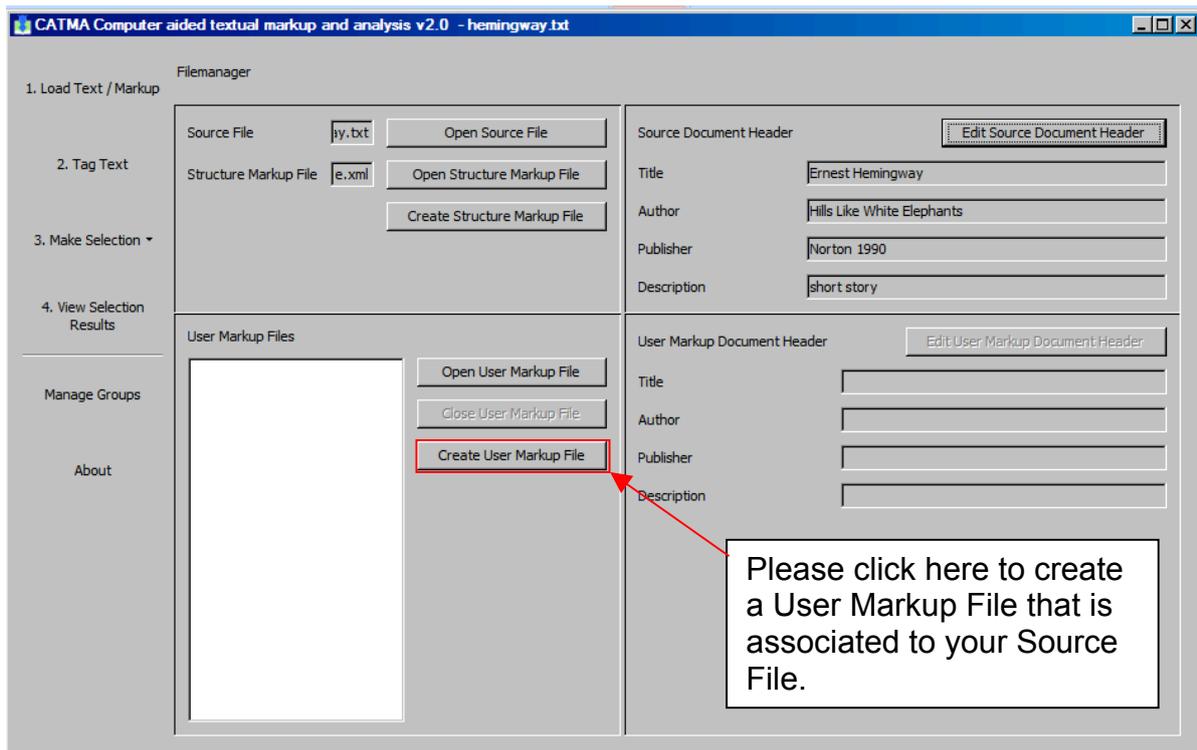
Publisher:

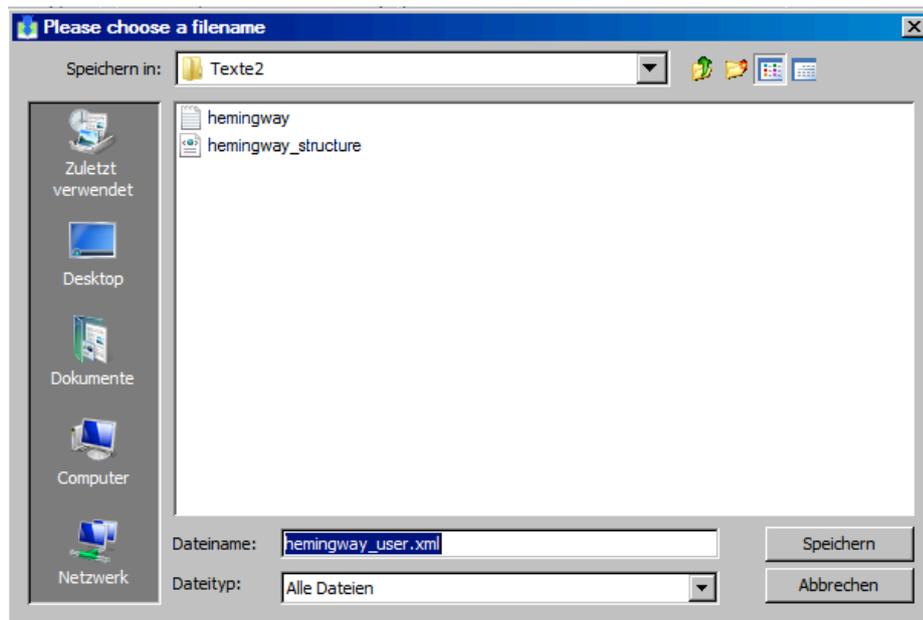
Description:

Back Next Cancel



Please select the type of the text you are loading and click “Finish”.  
The next step is to create a User Markup Document. The User Markup Document stores your workings, such as tagged sequences and the corresponding tags that you create.



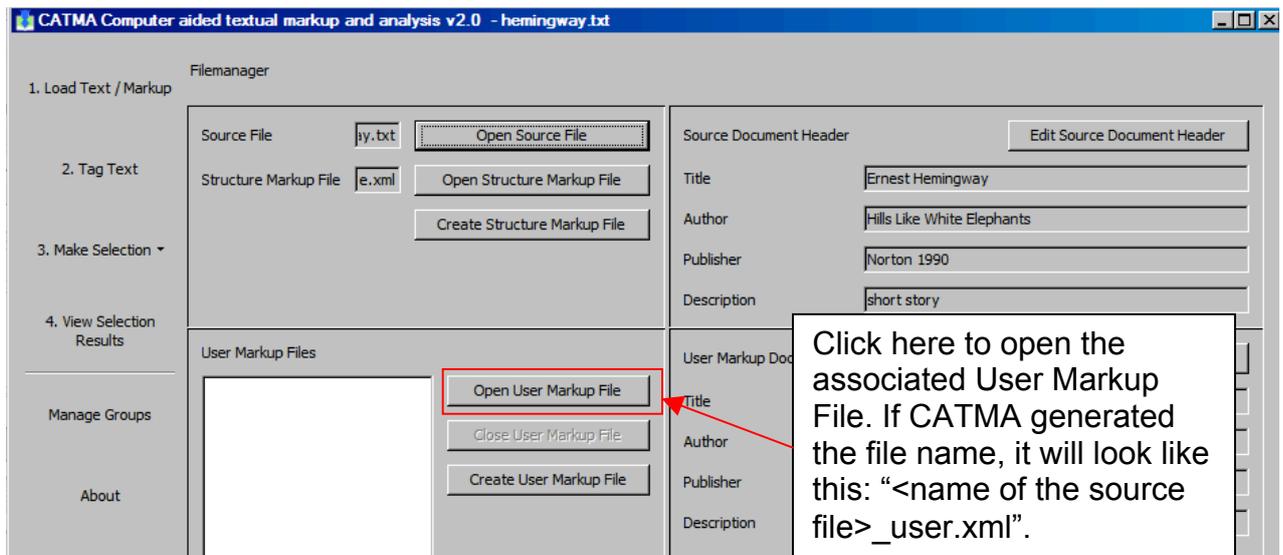


Please choose a filename for your User Markup File. CATMA generates a name for you that is constructed like this: “<file name of your source text>\_user.xml”. It is recommended to retain the name of the source text file, as well as the additional “user” in the file name, so that you can easily assign it to your source text when reloading it for the next session ([see section 1.2](#)). (Please note: You can create several User Markup Files for one source text file.)

After having saved your User Markup file, you can start to tag and analyse your text in CATMA. For further steps, see sections 2, 3 and 4.

## 1.2 Reload texts

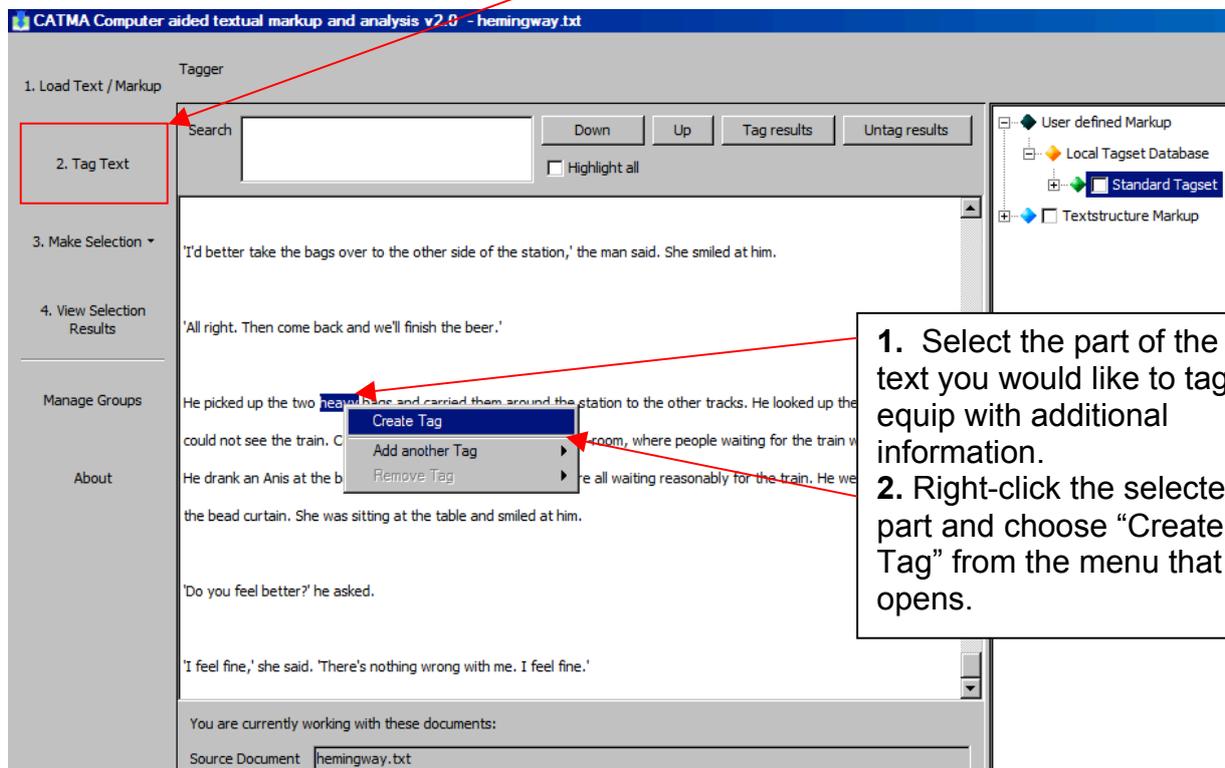
If you would like to reload a text that you have already worked on in CATMA, please click the “1. Load Text/ Markup” button as before. As described in the second step of section 1.1, click “Open Source File” and choose the desired text file from a folder. Since you have already created the associated Structure Markup File for your source text file, the Structure Markup File automatically loads. You only have to choose the associated User Markup File that you would like to use:



## 2. Tagging

### 2.1 Create Tags

To tag or highlight certain elements of your text, please click “2. Tag Text” in the menu on the left. In this section, it is possible to create tags by simply selecting a word or several character strings:



The screenshot shows the CATMA software interface. The title bar reads "CATMA Computer aided textual markup and analysis v2.0 - hemingway.txt". The main window is titled "Tagger" and has a menu on the left with items: "1. Load Text / Markup", "2. Tag Text" (highlighted with a red box), "3. Make Selection", "4. View Selection Results", "Manage Groups", and "About". The main text area contains several paragraphs of text from Hemingway's "The Sun Also Rises". A search bar is at the top with a "Search" field and buttons for "Down", "Up", "Tag results", and "Untag results". A "Highlight all" checkbox is also present. A right-click context menu is open over the text "near the station", with "Create Tag" selected. A text box on the right contains the following instructions:

1. Select the part of the text you would like to tag or equip with additional information.
2. Right-click the selected part and choose “Create Tag” from the menu that opens.

At the bottom, a status bar indicates "You are currently working with these documents:" and lists "Source Document hemingway.txt".

2. Tag Text

3. Make Selection

4. View Selection Results

Manage Groups

About

Source Document | hemingway.txt

User Markup Document - CATMA will store new Markup in this Document

computer aided textual markup and analysis v2.0 - hemingway.txt

Tagger

Search

Down Up Tag results Untag results

Highlight all

Text

'I'd better take the bags over to the other side of the station,' the man said. She smiled at him.

'All right. Then come back and we'll finish the beer.'

He picked up the two **heavy** bags and carried them around the station to the other tracks. He looked up the tracks but could not see the train. Coming back, he walked through the bar-room, where people waiting for the train were drinking. He drank an Anis at the bar and looked at the people. They were all waiting reasonably for the train. He went out through the bead curtain. She was sitting at the table and smiled at him.

Do you feel better?' he asked.

'I feel fine,' she said. 'There's nothing wrong with me. I feel fine.'

You are currently working with these documents:

Source Document | hemingway.txt

User Markup Document - CATMA will store new Markup in this Document

User defined Markup

- Local Tagset Database
  - Standard Tagset
    - qualification
  - Textstructure Markup

Tag

In

Rem

Ed

Imp

Exp

Tag

In

Ret

E

Prope

Inse

Ed

Remo

3. Enter a name for your tag.

4. Choose the colour in which your tag ought to appear in the text.

5. Choose to which tagset your tag ought to belong (you can create different tagsets to organize your tags, cf. 2.7).

6. Save your tag.

The tags you have created are displayed in your tagset.

The tags you have ticked off are also visible in your text.

Once you have created one kind of tag, it is very easy to use it again: Just select and right-click or shift-click the part of the text you would like to tag and choose the favoured tag from the menu:

The screenshot shows the Tagger application interface. At the top, there is a search bar and buttons for 'Down', 'Up', 'Tag results', and 'Untag results'. The main text area contains several lines of text, with the sentence "I've never seen one." highlighted in blue. A context menu is open over this text, showing options: 'Create Tag', 'qualification' (with a red square icon), 'direct-discourse' (with a blue square icon and highlighted), 'Add another Tag', and 'Remove Tag'. Two callout boxes provide instructions: the first points to the highlighted text, and the second points to the 'direct-discourse' option in the menu. On the right side, there is a 'User defined Markup' panel with a tree view showing 'Local Tagset Data' and 'Standard 1' with sub-items 'qualification' and 'direct-discourse' checked. At the bottom, there is a 'Source Document' field containing 'hemingway.txt'.

1. Select the part of the text you would like to tag, right-click or shift-click it to open the menu.

2. Choose the desired tag from the menu.

## 2.2 Tag all occurrences

To tag all occurrences of a token, a word or a phrase, you may use the search function of the tagger. (Please note: You can only tag 'all occurrences' with a tag that has been defined before).

The screenshot displays the CATMA interface with the following components and annotations:

- Search Bar:** Contains the text "he".
- Buttons:** "Down", "Up", "Tag results", and "Untag results".
- Checkbox:** "Highlight all" is checked.
- Text Area:** Displays a sample text with the word "he" highlighted in blue.
- Annotations:** Five numbered boxes with red arrows pointing to specific UI elements:
  - 1:** Points to the search bar, containing the text: "1. Enter the word/phrase you are searching for/you would like to tag (please use inverted commas; mind the use of capital and small initial letters)." (Note: The original image contains a typo "commas" which has been corrected to "commas").
  - 2:** Points to the "Highlight all" checkbox, containing the text: "2. Tick off 'Highlight all' to show all occurrences in the text."
  - 3:** Points to the "Down" and "Up" buttons, containing the text: "3. Click 'Down' or 'Up' to search the text down or up."
  - 4:** Points to the highlighted word "he" in the text area, containing the text: "4. All occurrences of the word/phrase are highlighted."
  - 5:** Points to the "Tag results" button, containing the text: "5. To tag the highlighted parts of the text, click 'Tag results'."

The interface also shows a sidebar on the left with options like "1. Load Text / Markup", "2. Tag Text", "3. Make Selection", "4. View Selection Results", "Manage Groups", and "About". At the bottom, it displays document information: "You are currently working with these documents:", "Source Document: hemingway.txt", and "User Markup Document - CATMA will store new Markup in this Document: hemingway\_user.xml".

After having clicked “Tag results”, you can choose if you would really like to tag all occurrences or only selected results:

The screenshot shows a 'Tag selected results' dialog box with a table of results. The table has two columns: 'Selected' and 'Keyword in context'. The 'Selected' column contains checkboxes, with two checked. The 'Keyword in context' column contains text snippets. Below the table are buttons for 'Select all', 'Select none', 'Total selected' (with a value of 2), and 'Total' (with a value of 5). There are also 'Back', 'Next', and 'Cancel' buttons.

Annotations include:

- A box pointing to the 'Selected' column: "1. Tick off the results you would like to tag."
- A box pointing to the 'Select all' and 'Select none' buttons: "1a. Click 'Select all' to tag all results. Click 'Select none' to undo your selection."
- A box pointing to the 'Next' button: "2. Click 'Next'."

The second dialog box, 'Standard Tagset', shows a tree structure under 'Standard Tagset' with three items: 'character' (highlighted in blue), 'direct-discourse' (blue square), and 'qualification' (red square). A box points to the 'character' item: "Choose the tag you would like to use. (If the tags are not displayed here, please click the 'plus' in front of the tagset to open the tree structure). Click 'Finish'".

## 2.3 Edit Tags

To edit an existing tag (e.g. change name), you can use the menu on the right-hand side:

The screenshot shows a software interface for managing tags. On the left, a tree view displays a hierarchy: 'User defined Markup' contains 'Local Tagset Database', which contains 'Standard Tagset'. Under 'Standard Tagset', there are three items: 'qualification' (red square), 'direct-discourse' (blue square, selected), and 'direct-discourse-girl' (green square). Below this is 'Textstructure Markup'. On the right, a 'Tagset' menu contains buttons for 'Insert Tagset', 'Remove Tagset', 'Edit Tagset', 'Import Tagset', and 'Export Tagset'. Below that is a 'Tag' section with buttons for 'Insert Tag', 'Remove Tag', and 'Edit Tag'. At the bottom is a 'Property' section with buttons for 'Insert Property', 'Edit Property', and 'Remove Property'. Four red arrows point from the instructions on the left to the corresponding buttons in the right-hand menu.

**1.** To change an existing tag, select it.

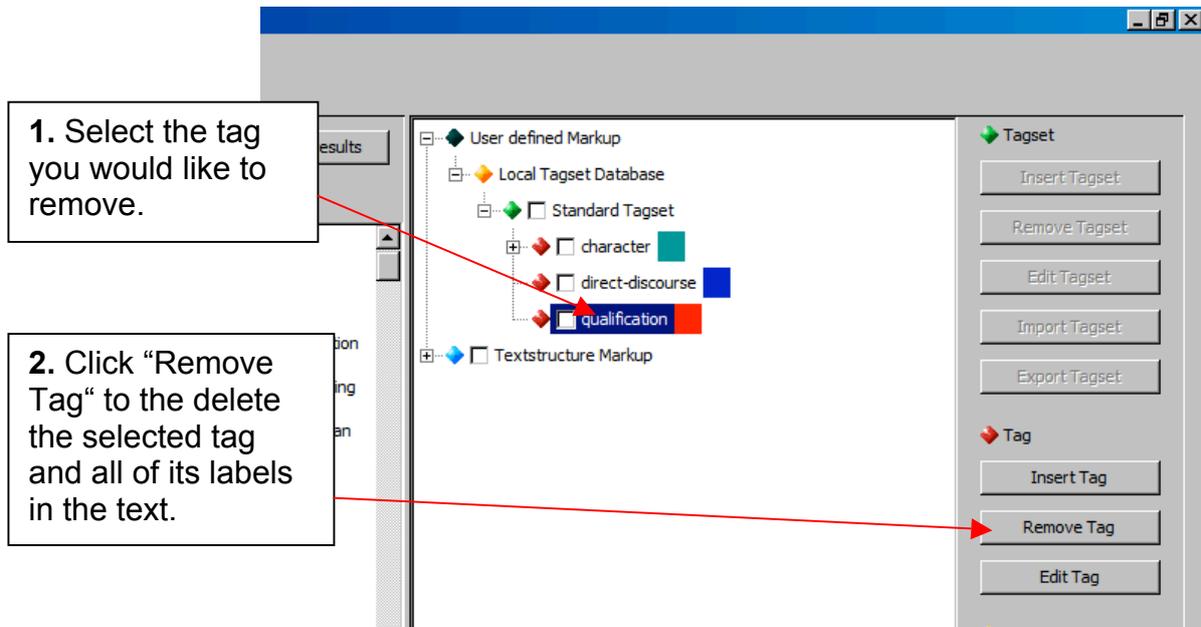
**2.** With “Insert Tag“, you can create a new tag that is subordinate to the tag you have selected. In the tagset, it is organized in a tree structure.

**3.** With “Remove Tag“, you delete the tag and its labels in the text.

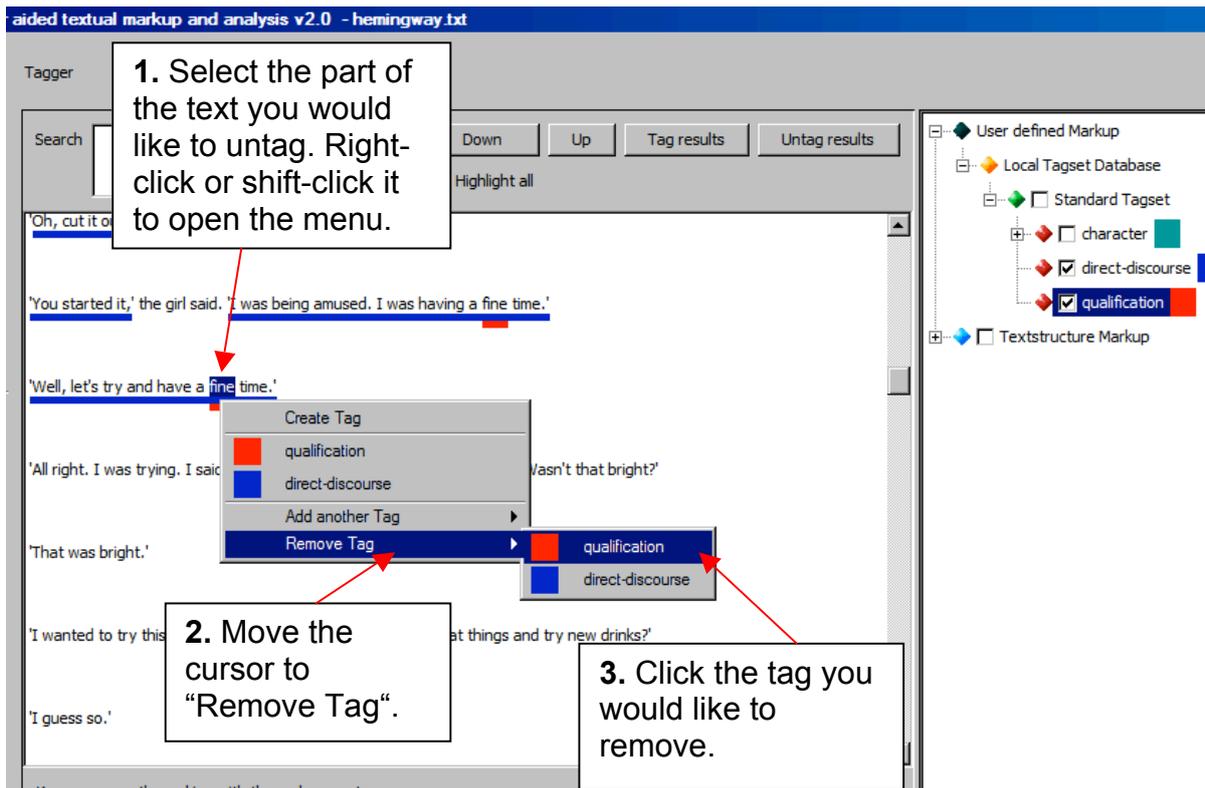
**4.** With “Edit Tag“ you can change the colour and the name of the selected tag.

## 2.4 Remove tags

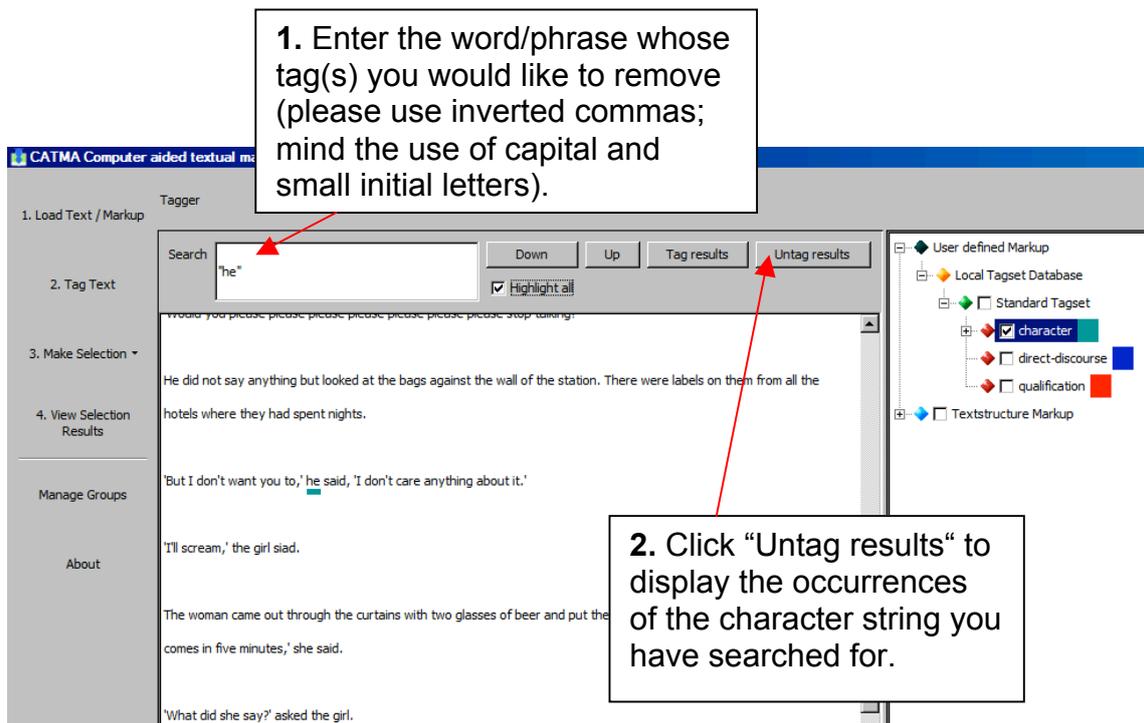
There are two ways to remove tags in CATMA: You can delete a tag and all of its labels in the text, or you can just remove one label of a tag in the text. For the first option, please follow the instructions below:



Second option: Removing a tag or label in the text is similar to the creation of a tag (cf. [2.1 Create Tags](#)):



It is also possible to remove several tags at once:



The following dialogue window displays all occurrences of the character string you have searched for:

Tick off the occurrences whose tags you would like to remove.

Selected	Keyword in context
<input type="checkbox"/>	in the shade,' he said. 'You mustn't
<input checked="" type="checkbox"/>	got to realize,' he said, ' that I
<input checked="" type="checkbox"/>	want you to,' he said, 'I don't
<input type="checkbox"/>	train. Coming back, he walked through the bar-room,
<input type="checkbox"/>	you feel better?' he asked. 'I feel

Select all    Select none    Total selected: 2    Total: 5

Click "Select all" to tick off all occurrences. Click "Select none" to undo your selection.

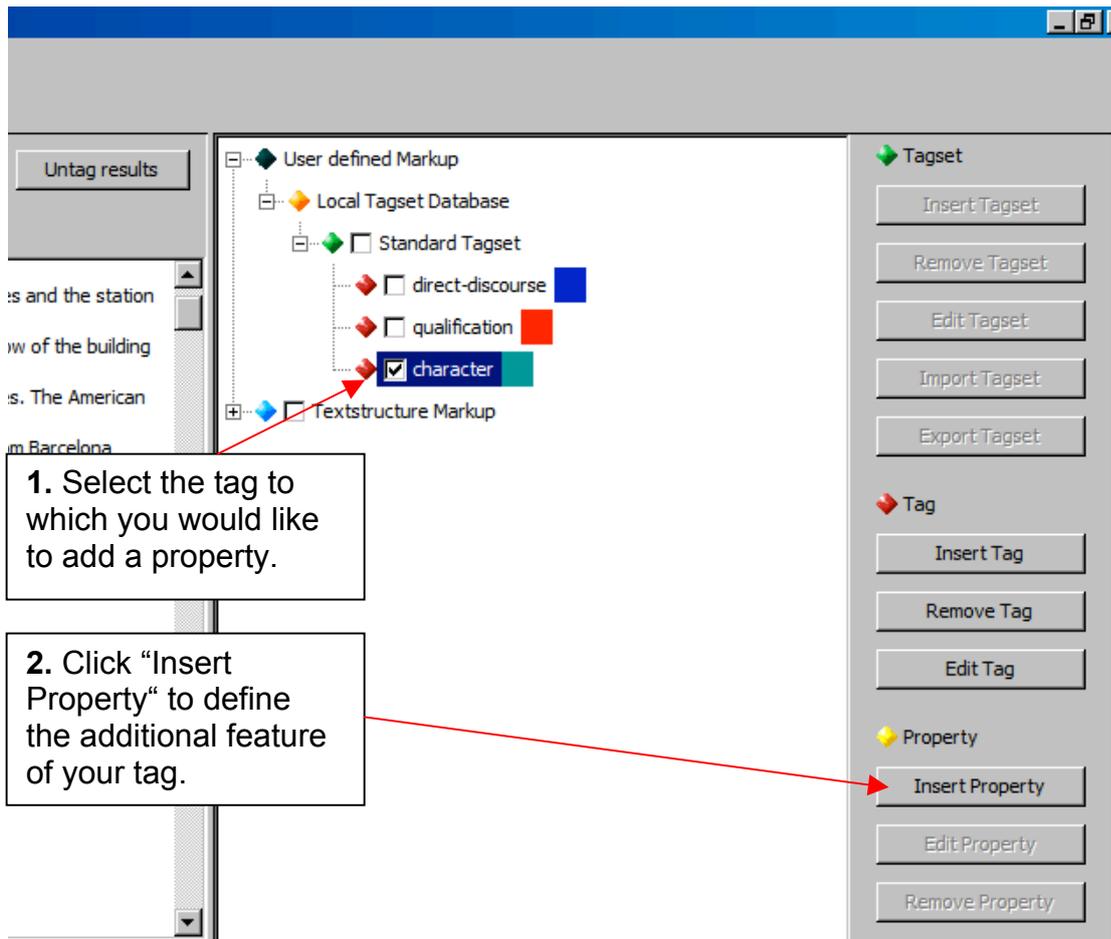
After having clicked "Next", you may choose the tag that is to be removed:

Choose the tag you would like to remove. (If the tags are not displayed here, please click the 'plus' in front of the tagset to open the tree structure). Click "Finish".

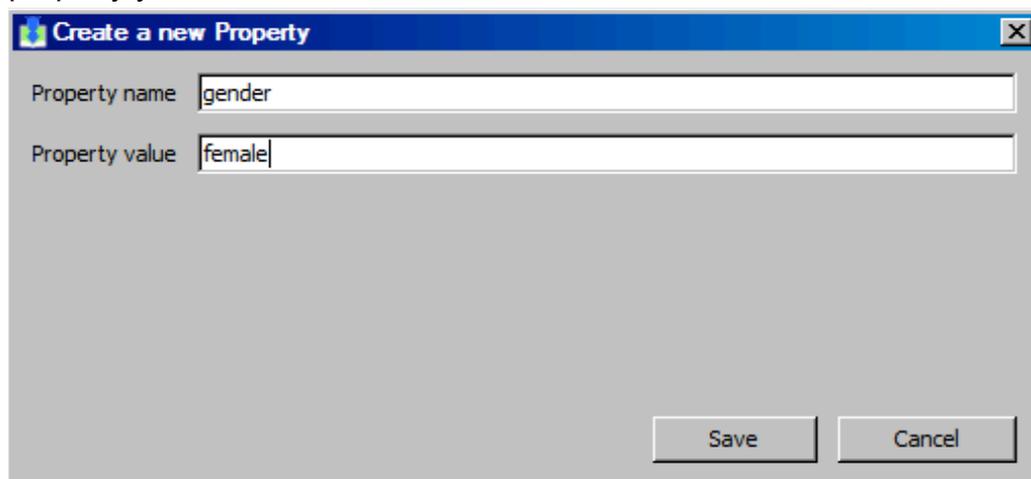
- Standard Tagset
  - character
  - direct-discourse
  - qualification

## 2.5. Insert property

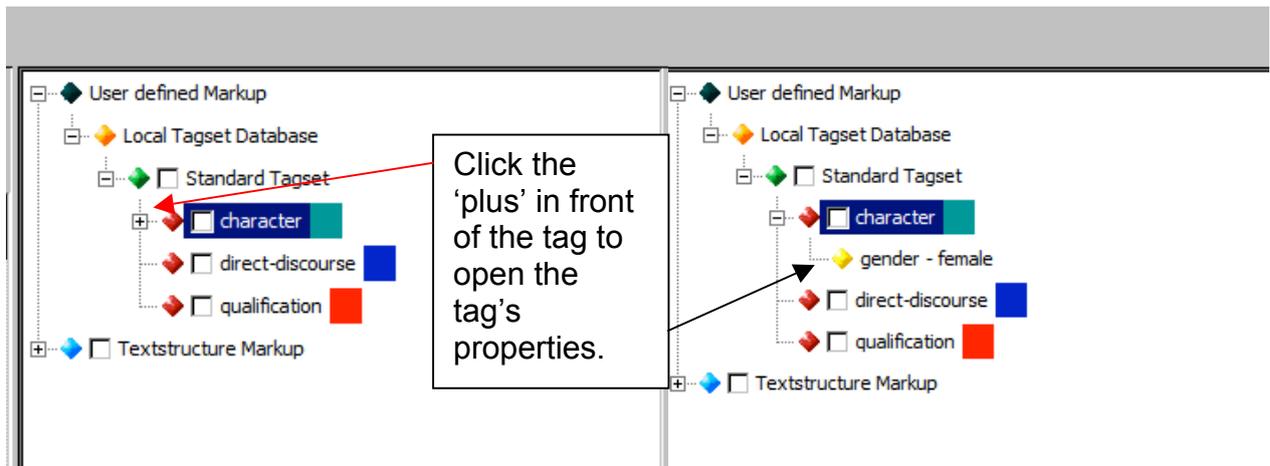
You can also create subunits of an existing tag to differentiate or refine its meaning. For example: You have defined a tag called "character" that you use to highlight every character appearing in the text. As may be the case, you would like to distinguish between male and female characters within your tag; this means you have to "insert property". The operation is similar to editing a tag:



In the following dialogue window, you can enter the name and the value of the property you would like to define. Please click "Save":



You can also display the properties belonging to a tag in the tree structure:



## 2.6 Remove and edit property

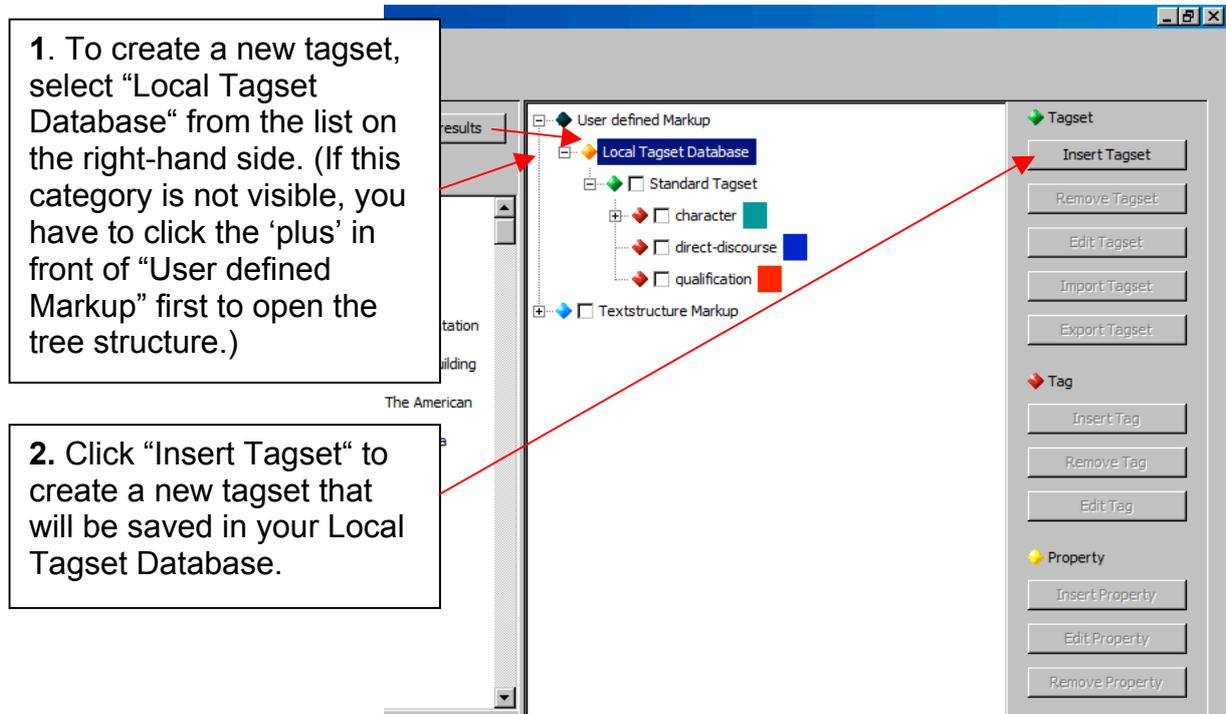
1. To remove a property from a tag, select it. (If the property/ies is/are not displayed here, please click the 'plus' in front of the tag to open the tree structure).

2a. Click "Edit Property" to change the name and value of the property you have selected.

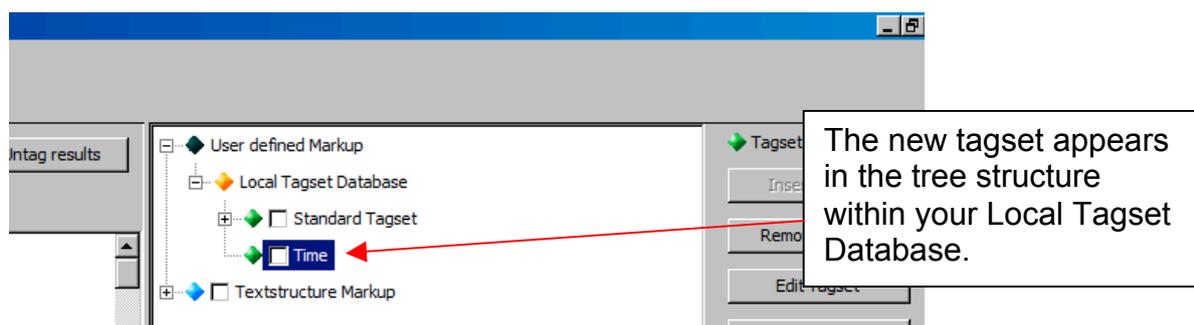
2b. Click "Remove Property" to delete the property you have selected.

## 2.7 Create new tagsets

It is possible to define different tagsets to structure tags thematically, for example. You may also import tagsets you have defined for one text to use them with another.



A dialogue window opens; please enter the name of your new tagset. After having saved the new tagset, it appears in the tree structure.



To save tags in the new tagset, you have to determine the tag position when tagging an element of the text. First, you select the part of the text you would like to tag and right-click or shift-click it, as described in section 2.1. Click "Create new tag". In the dialogue window that opens, you can determine the tag position:

by of the Ebro were long and white. On this side there was no shade and no trees and the station of rails in the sun. Close against the side of the station there was the warm shadow of the building strings of bamboo beads, hung across the open door into the bar, to keep out flies. The American : at a table in the shade, outside the building. It was very hot and the express from Barcelona utes. It stopped at this junction for two minutes and went to Madrid. the girl asked. She had taken off her hat and put it on the table.

1 said.

ing with these documents:  
mingway.txt  
t - CATMA will store new Markup in this

1. Enter the tag name and determine the tag colour as usual.

2. Determine the tag position by selecting the new tag set. Click "Save".

You may also create a tag for the new tagset without first selecting a piece of text:

1. Select the tagset in which the new tag ought to be saved.

2. Click "Insert Tag".

3. A dialogue window opens in which you can determine tag name, tag colour and tag position (see previous screenshot).

## 2.8 Remove or edit tagsets

To edit the name of the new tagset, please select the tagset and click “Edit Tagset” - the tags of the tagset will not be changed by this operation. But if you remove your tagset, all of the tags in the tagset and its labels in the text will be deleted, too.

The screenshot displays a software interface for managing tagsets. On the left, a text snippet is visible: "s. The American" and "m Barcelona". The main area is divided into two panes. The left pane shows a tree structure under "User defined Markup". It includes a "Local Tagset Database" containing a "Standard Tagset" with sub-items: "character" (teal), "direct-discourse" (blue), "qualification" (red), and "Time" (blue, selected). Below it is "time\_span" (yellow). At the bottom is "Textstructure Markup". The right pane, titled "Tagset", contains buttons: "Insert Tagset", "Remove Tagset", "Edit Tagset", "Import Tagset", and "Export Tagset". Below these is a "Tag" section with an "Insert Tag" button. Three callout boxes provide instructions: "1. Select the tagset you would like to edit or delete." points to the "Time" tagset in the tree; "2a. Click 'Remove Tagset' to delete your tagset and all the tags belonging to it." points to the "Remove Tagset" button; "2b. Or, click 'Edit Tagset' to change the name of your tagset." points to the "Edit Tagset" button.

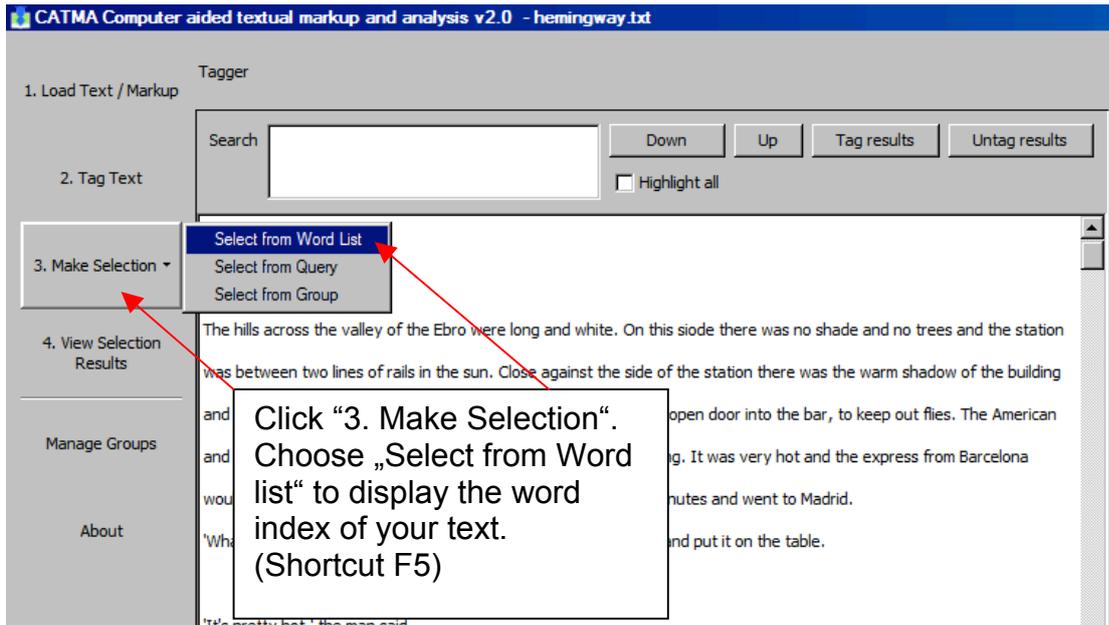
**1.** Select the tagset you would like to edit or delete.

**2a.** Click “Remove Tagset” to delete your tagset and all the tags belonging to it.

**2b.** Or, click “Edit Tagset” to change the name of your tagset.

### 3. Wordlist

The word list allows you to choose one or more words from your text in order to display its/their context or its/their distribution, for example.



(Please note: You can easily jump between the "Tag Text" section and the "Make Selection" section without losing your tags or labels in the text.)

CATMA Computer aided textual markup and analysis v2.0 - hemingway.txt

1. Load Text / Markup Select from Wordlist

Search a word  Search

Selected	Hit	Frequency
<input type="checkbox"/>	And	9
<input type="checkbox"/>	Anis	3
<input type="checkbox"/>	another	2
<input type="checkbox"/>	any	2
<input checked="" type="checkbox"/>	anybody	1
<input checked="" type="checkbox"/>	anyone	1
<input type="checkbox"/>	anything	9
<input type="checkbox"/>	are	1
<input type="checkbox"/>	around	1
<input type="checkbox"/>	asked	5
<input type="checkbox"/>	at	21
<input type="checkbox"/>	away	3
<input type="checkbox"/>	awfully	1
<input type="checkbox"/>	back	4
<input type="checkbox"/>	bags	3
<input type="checkbox"/>	bamboo	1
<input type="checkbox"/>	banks	1
<input type="checkbox"/>	bar	3
<input type="checkbox"/>	Barcelona	1
<input type="checkbox"/>	bar-room	1
<input type="checkbox"/>	be	8
<input type="checkbox"/>	bead	4
<input type="checkbox"/>	beads	2

Manage Groups

Selected Total  Selected Frequencies Total  Freq

compute Distribution Analysis  compute Collocate Analysis

Update Selection and View Results

Select all Select none

Click "Hit" to change the displayed order of words from A-Z to Z-A.

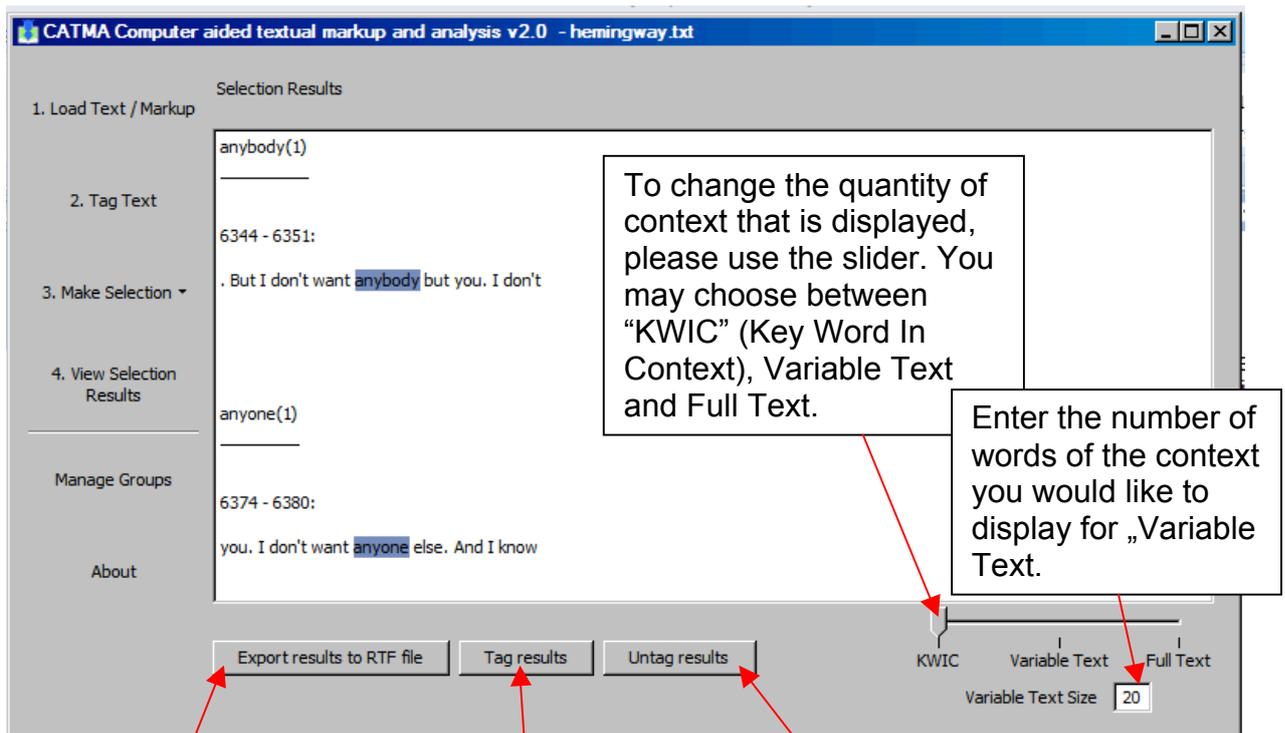
Tick off words you would like to use for further operations. Click "Select all" to tick off all of the displayed words; click "Select none" to undo your selection.

Click "Frequency" to change the displayed order of words from "lowest frequency of occurrences in the text" to "highest frequency of occurrences in the text"

After you have selected words from the word list, you may use them for three basic operations: displaying selected words in context (cf. [3.1 Display selected words in context](#)), compute Distribution Analysis (cf. [3.2 Compute Distribution Analysis](#)) and compute Collocate Analysis (cf. [3.3 Compute Collocation Analysis](#)).

### 3.1 Display selected words in context

To show the context of the words you have selected from the word list, please click “Update Selection and View Results” at the bottom of the list.



To change the quantity of context that is displayed, please use the slider. You may choose between “KWIC” (Key Word In Context), Variable Text and Full Text.

Enter the number of words of the context you would like to display for „Variable Text.

Click „Export results to RTF file“ to save your results in a Rich Text Format file.

Click „Tag results“ to tag selected occurrences (cf. 2.2 “Tag all occurrences”). Please note: The tag must have been created before.

Click „Untag results“ to delete the labels of selected occurrences. (cf. 2.4).

## 3.2 Compute Distribution Analysis

You may also display the distribution over the text of one or more words you have selected from the word list:

**1.** Tick off the occurrences whose distribution over the text you would like to display. Click “Select all” to tick off all of the displayed words; click “Select none” to undo your selection.

Selected	Hit	Frequency
<input type="checkbox"/>	fields	1
<input type="checkbox"/>	fine	6
<input type="checkbox"/>	finish	1
<input type="checkbox"/>	five	2
<input type="checkbox"/>	flies	1
<input type="checkbox"/>	for	8
<input type="checkbox"/>	forty	1
<input type="checkbox"/>	from	4
<input type="checkbox"/>	get	3
<input checked="" type="checkbox"/>	girl	22

Selected Total  Selected Frequencies Total  Frequency Total

compute Distribution Analysis   
 compute Collocation Analysis

Update Selection and View results

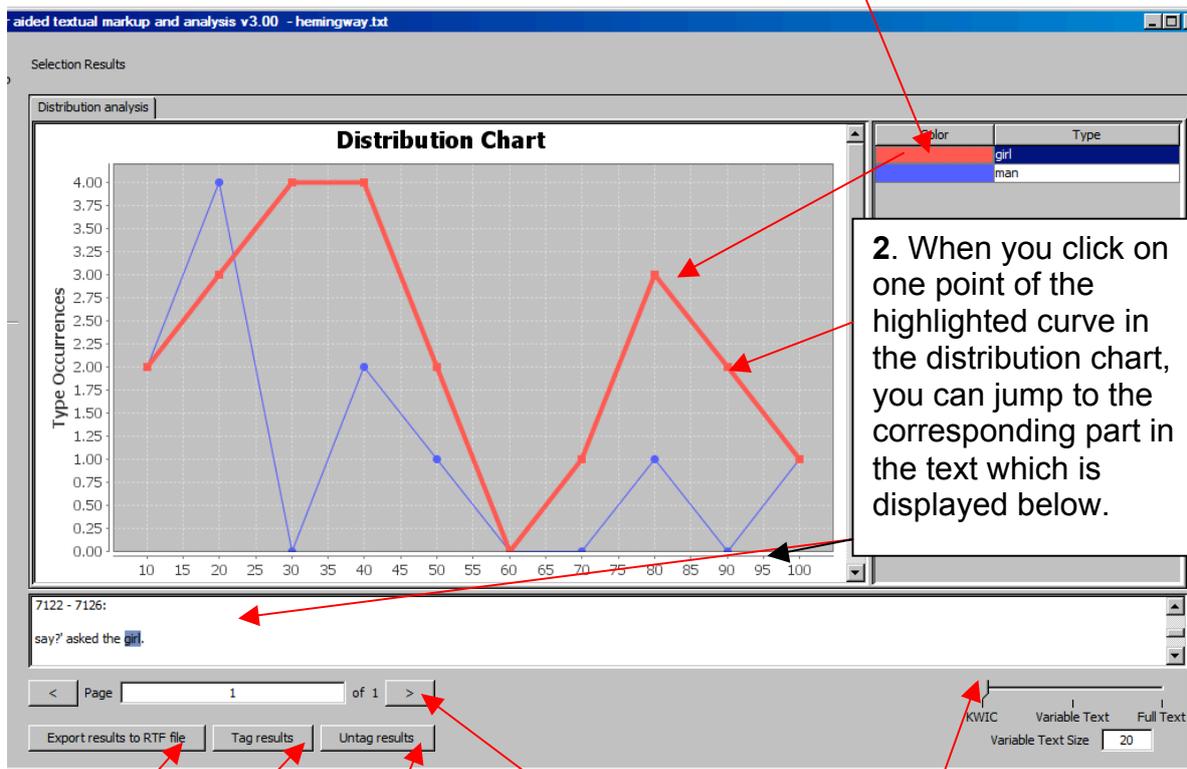
**2.** Tick off “compute Distribution Analysis”.

**3.** Open the drop box to adjust the granularity of the distribution display. The smaller the chunks, the more precise are the results.

To change to the distribution chart, please click “Update Selection and View results” right at the bottom of the window.

The Distribution Chart shows you the frequency of the occurrences on the y-axis, their distribution over the text (in %) on the x-axis.

1. Click on the type of occurrence to highlight its curve shape in the distribution chart.



2. When you click on one point of the highlighted curve in the distribution chart, you can jump to the corresponding part in the text which is displayed below.

In this section you are also able to save the results in Rich Text Format, and to tag or untag the results (cf. 3.1).

3. With the arrow buttons you may jump to the next or previous page of the results, with the slider you can change the quantity of context that is displayed, please use the slider. You may choose between "KWIC" (Key Word In Context), Variable Text and Full Text.

Please note: It is possible to display the distribution of occurrences of every kind of query (words, tags, collocations, etc.).

### 3.3 Compute Collocation Analysis

You may also display the collocations of one or more words you have selected from the word list. To compute a Collocation Analysis means: you search for the words that are in the vicinity of the word(s) you have selected; moreover, CATMA shows you with which statistical probability the selected word(s) occur(s) with other words in the text. As shown in section 3.1 and 3.2, you have to select words from the word list first:

**1.** Tick off the words whose collocations you would like to display. Click "Select all" to tick off all of the displayed words; click "Select none" to undo your selection.

Selected	Hit	Frequency
<input type="checkbox"/>	It's	11
<input type="checkbox"/>	looked	11
<input checked="" type="checkbox"/>	man	11
<input type="checkbox"/>	was	11
<input type="checkbox"/>	with	11
<input type="checkbox"/>	that	10
<input type="checkbox"/>	were	10
<input type="checkbox"/>	And	9
<input type="checkbox"/>	anything	9
<input type="checkbox"/>	if	9
<input type="checkbox"/>	in	9
<input type="checkbox"/>	like	9

**2.** Tick off „compute Collocation Analysis“. Click "Update Selection and View Results".

Selected Total: 2    Selected Frequencies Total: 33    Frequency Total: 1936    Type Total: 93

compute Distribution Analysis     compute Collocate Analysis

Update Selection and View Results

Click "Collocation" to change the order of the displayed collocations from A-Z to Z-A.

Click "Collocation frequency" to change the order of the displayed collocations from "lowest collocate frequency" to "highest collocate frequency".

Click "Type frequency" to change the order from "lowest type frequency" to "highest type frequency".

Click "z-score" to change the order from "lowest z-score" to "highest z-score".

Collocation	Collocation frequency	Type frequency	z-score
at	15	21	6.069
The	13	18	5.697
the	41	109	5.354
said	15	27	4.881
looking	2	1	4.433
looked	7	11	3.753
lovely	2	2	2.843
!	19	61	2.711
smiled	2	3	2.083
cervezas	1	1	2.01
cool	1	1	2.01
prove	1	1	2.01
American	1	1	2.01
might	1	1	2.01
ground	1	1	2.01

Page 1 of 1

Export results to RTF file | Tag results | Untag results

KWIC | Variable Text | Full Text

Variable Text Size: 20

In this section you are also able to save the results in Rich Text Format, and to tag or untag the results (cf. 3.1).

3. With the arrow buttons you may jump to the next or previous page of the results, with the slider you can change the quantity of context that is displayed, please use the slider. You may choose between "KWIC" (Key Word In Context), Variable Text and Full Text.

In the column "Collocation", you will find the words that are in the vicinity of the word(s) you have selected.

In the column "Collocation frequency", you will find the frequency with which the word from the column "Collocation" occurs together with the word(s) you have selected from the word list.

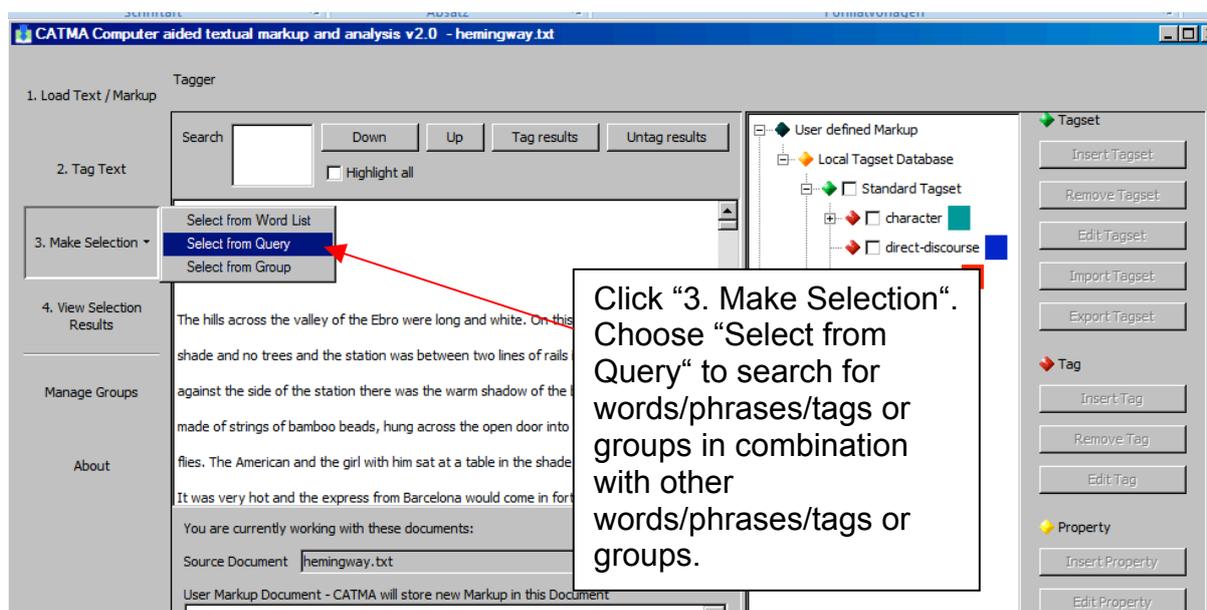
The column "Type frequency" shows the total frequency of each collocation for the entire text.

The column "z-score" shows with which statistical probability the collocation occurs in the text. Or, in other words, how likely it is that two words did not occur together at random. The higher the z-score either positive or negative, the more probable is the collocation.

## 4. Queries

### 4.1 Execute queries

You may also search for combinations of words and phrases, or combinations of tags and groups (c.f. section 5), for example. For these operations, please select “Select from Query” from the “Make Selection” menu.



1. In this input field, you can either enter your query manually (you can find the operators and the correct syntax of queries in section 4.3.), or – which is much easier – use the QueryBuilder which compiles the correct query syntax for you (cf. 4.2).

1a. Click here to use the QueryBuilder (cf. 4.2).

2. Click „Execute Query“.

Query: ("girl", "woman") & tag="direct-discourse" 20

Selected	Hit	Frequency
<input checked="" type="checkbox"/>	girl	10
<input type="checkbox"/>	woman	4

Selected Total: 1   Selected Frequencies Total: 10   Frequency Total: 14   Total: 2

compute Distribution Analysis   10% chunks

compute Collocation Analysis

Update Selection and View Results

3. Tick off the results you would like to use for further operations.

You may execute the same operations with the results of a query as described in chapter 3: Display selected results in its context (cf. [3.1 Display selected words in context](#)), Compute Distribution Analysis (cf. [3.2 Compute Distribution Analysis](#)), Compute Collocate Analysis (cf. [3.3 Compute Collocation Analysis](#)).

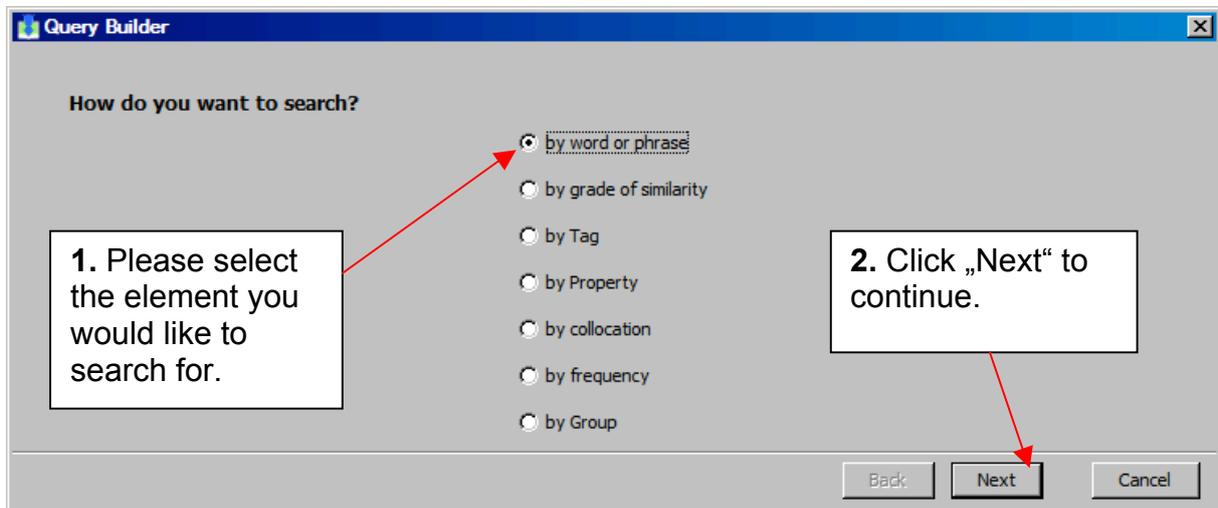
## 4.2 The QueryBuilder

The QueryBuilder helps you to compile the correct syntax for your query. (Please note: The syntax structure the QueryBuilder constructs sometimes differs from the examples in [section 4.4](#) (Query syntax) because in some cases the QueryBuilder compiles a standardized, more complex form – so don't be confused.)

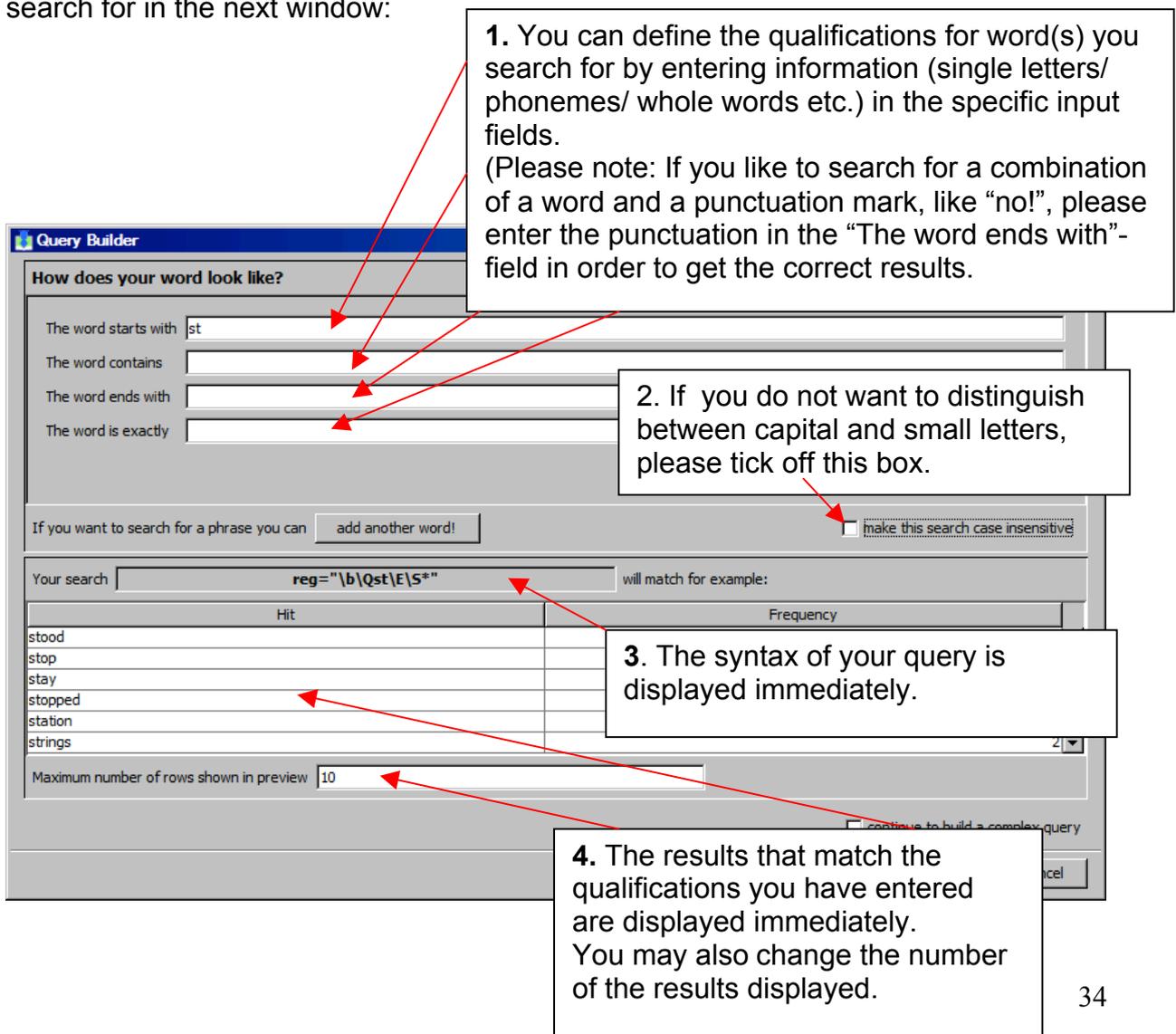
After having clicked "Make Selection" → "Select from Query" → "QueryBuilder", you can select the elements (tags, words, phrases, collocations, etc.) you would like to

search for. The different operations for the use of different elements will be introduced in the following section, one after another.

#### 4.2.1 Search “by word or phrase”



If you have selected “by word or phrase”, you can enter the words or phrases you search for in the next window:



1. If you would like to search for a phrase, please click „add another word!“.  
(You can repeat this and add as many words as you need for your query.)

How does your word look like?

The word starts with

The word contains

The word ends with

The word is exactly

Next word starts with

Next word contains

Next word ends with

Next word is exactly

The position of this word is

If you want to search for a phrase you can

make this search case insensitive

Your search  will match for example:

Hit	F
love	

Maximum number of rows shown in preview

continue to build a complex query

3. The query syntax and the matching results are displayed immediately.

3a. You may also change the number of the results displayed.

4. You can finish building the syntax for your query by clicking “Finish”.

(Please note: You can also add further elements to your query. For example, if you like to search the words/phrases you have selected in combination with other elements such as tags or collocations, please tick off “continue to build a complex query” in order to extend your query. The next steps are described in [section 4.4](#)).

1. Load Text / Markup

2. Tag Text

3. Make Selection

4. View Selection Results

Manage Groups

About

Help

Select from Query

Query

Selected	Hit
<input type="checkbox"/>	feel that
<input type="checkbox"/>	for two
<input type="checkbox"/>	for the
<input type="checkbox"/>	finish the
<input type="checkbox"/>	from the
<input type="checkbox"/>	fine time.'

Selected Total  Selected Frequencies Total

compute Distribution

compute Collocation

Search finished!

After having clicked „Finish“, the correct query syntax appears in the input field .  
Now you can select the occurrences you would like to continue to work with (e.g., compute distribution analysis cf. 3.2, compute collocate analysis cf. 3.3).

#### 4.2.2 Search “by grade of similarity”

If you have selected from the QueryBuilder menu that you would like to search “by grade of similarity”, you may use the following options:

The screenshot shows the 'Query Builder' window. It includes a text input field for the word to search, a slider for the similarity grade, a preview table of results, and a 'Finish' button. Red arrows point from callout boxes to these specific UI elements.

**1.** Enter a word to search other occurrences that are similar to the word you have chosen.

**2.** Adjust the grade of similarity (in percent) by moving the slider.

**3.** The query syntax and the matching results are displayed immediately.

**3a.** You may also change the number of the results displayed.

**4.** You can finish building the syntax for your query by clicking “Finish”.

Hit	Frequency
bar-room	
to	
no	
you've	
Could	
you'll	
ours	
course	

(Please note: You can also add further elements to your query. For example, if you like to search the words/phrases you have selected in combination with other elements such as tags or collocations, please tick off “continue to build a complex query” in order to extend your query. The next steps are described in section 4.4: Complex queries with the QueryBuilder).

After having clicked „Finish“, the correct query syntax appears in the input field of the “Select from Query” section.

Now you can select the occurrences you would like to continue to work with (e.g., compute distribution analysis cf. 3.2, compute collocate analysis cf. 3.3).

### 4.2.3 Search “by tag”

If you have selected from the QueryBuilder menu that you would like to search “by tag”, you may use the following options:

The screenshot shows the 'Query Builder' window. At the top, it says 'Please select a Tag. Only Tags that are used in one of the currently loaded User Markup Documents are shown:'. Below this is a tree view of tagsets. The 'Standard Tagset' is expanded to show 'Perspective' and 'Time'. Under 'Perspective', 'speech-show' is selected. Under 'Time', 'present' is selected. A callout box points to the 'speech-show' tag with the text: '1. Select the tag from one of you tagsets you would like to search. (If the your tags are not displayed in a tree structure, please click on the “plus” next to the name of your tagset to open the tree.)'. Below the tree view is a search input field containing 'tag="speech-show"'. A callout box points to this field with the text: '2. The query syntax and the matching results are displayed immediately.'. Below the search field is a table with two columns: 'Hit' and 'Freq'. The table contains five rows of text. A callout box points to the table with the text: '2a. You may also change the number of the results displayed.'. Below the table is a text box for 'Maximum number of rows shown in preview' with the value '10'. At the bottom right, there are three buttons: 'Back', 'Finish', and 'Cancel'. A callout box points to the 'Finish' button with the text: '3. You can finish building the syntax for your query by clicking “Finish”.'.

(Please note: You can also add further elements to your query. For example, if you like to search the words/phrases you have selected in combination with other elements such as tags or collocations, please tick off “continue to build a complex query” in order to extend your query. The next steps are described in [section 4.3](#)).

After having clicked „Finish“, the correct query syntax appears in the input field of the “Select from Query” section.

Now you can select the occurrences you would like to continue to work with (e.g., compute distribution analysis cf. 3.2, compute collocate analysis cf. 3.3).

### 4.2.4 Search “by property”

If you have selected from the QueryBuilder menu that you would like to search “by property”, you may use the following options:

The screenshot shows the Query Builder window with the following components and callouts:

- 1.** A callout box pointing to the 'Please select the Property you want to search for:' field, which contains a list with 'preposition' and 'character' selected.
- 2.** A callout box pointing to the 'You can optionally specify the value of the Property:' field, which contains a list with 'Man' and 'girl' selected.
- 3.** A callout box pointing to the 'Tags containing the selected Property/value pair:' field, which contains 'speech-show'.
- 4.** A callout box pointing to the 'Your search' field, which displays the query 'property = "character" value = "Man"', and to the 'Hit' column of the results table.
- 4a.** A callout box pointing to the 'Maximum number of rows shown in preview' input field, which is set to '10'.
- 5.** A callout box pointing to the 'Finish' button at the bottom right of the window.

Hit	Freq
They've painted something on it	
They look like white elephants	
'Four reales.' 'We want two Anis del Toro.' 'With water?' 'Do you want it with w...	
'It tastes like liquorice	
Oh, cut it out	
That's the way with everything	
'It's all right.' 'You want them with water?	1
Yes	1
What does it say?	1
It's pretty hot,	1

(Please note: You can also add further elements to your query. For example, if you like to search the words/phrases you have selected in combination with other

elements such as tags or collocations, please tick off “continue to build a complex query” in order to extend your query. The next steps are described in section 4.3: Complex queries with the QueryBuilder).

After having clicked „Finish“, the correct query syntax appears in the input field of the “Select from Query” section.

Now you can select the occurrences you would like to continue to work with (e.g., compute distribution analysis cf. 3.2, compute collocate analysis cf. 3.3).

#### 4.2.5 Search “by collocation”

If you have selected from the QueryBuilder menu that you would like to search “by collocation”, you may use the following options:

The screenshot shows the Query Builder window with the following fields and controls:

- Search for all occurrences of:
- that appear near:
- within a span of:
- Your search:
- will match for example:
- Hit: 

Hit	Frequency
man	
- Maximum number of rows shown in preview:
- continue to build a complex query
- Buttons: Back, Finish, Cancel

Four callout boxes provide instructions:

1. In the first input field, please enter a word of which you would like to search all occurrences that appear next to the word you enter in the second field.
2. Adjust the span (the number of words that is allowed to stand between the two words you have chosen) by using the up and down-buttons.
3. The query syntax and the matching results are displayed immediately.
- 3a. You may also change the number of the results displayed.
4. You can finish building the syntax for your query by clicking “Finish”.

(Please note: You can also add further elements to your query. For example, if you like to search the words/phrases you have selected in combination with other elements such as tags or collocations, please tick off “continue to build a complex query” in order to extend your query. The next steps are described in section 4.3: Complex queries with the QueryBuilder).

After having clicked „Finish“, the correct query syntax appears in the input field of the “Select from Query” section.

Now you can select the occurrences you would like to continue to work with (e.g., compute distribution analysis cf. 3.2, compute collocate analysis cf. 3.3).

#### 4.2.6 Search “by frequency”

If you have selected from the QueryBuilder menu that you would like to search “by frequency”, you may use the following options:

**1.** Please select the restrictions for the word(s) you search for: You can choose whether the number of occurrences of one type ought to correspond exactly to the value you have entered in the field in the middle, if the number of occurrences ought to exceed or fall below this value or ought to be between the first and the second value you have entered in the two input fields.

The screenshot shows the 'Query Builder' window. At the top, there is a dropdown menu for 'The word shall appear' with options: 'between', 'exactly', 'more than', 'less than', 'more than or exactly', 'less than or exactly', and 'between'. The 'between' option is selected. To the right of this dropdown are two input fields containing the numbers '5' and '10', with 'and' between them, and the text 'times.' to the right. Below this is a section 'Your search' with the text 'freq = 5-10' and 'will match for example:'. Below that is a table with the following data:

Hit	
hills	5
they	5
but	5
for	8
What	6
love	5
were	10
beer	7
please	7
woman	6

At the bottom of the window, there is a text box 'Maximum number of rows shown in preview' with the value '10'. To the right of this is a checkbox labeled 'continue to build a complex query' which is unchecked. At the bottom right are three buttons: 'Back', 'Finish', and 'Cancel'.

**2a.** You may also change the number of the results displayed.

**2.** The query syntax and the matching results are displayed immediately.

**3.** You can finish building the syntax for your query by clicking “Finish”.

(Please note: You can also add further elements to your query. For example, if you like to search the words/phrases you have selected in combination with other elements such as tags or collocations, please tick off “continue to build a complex query” in order to extend your query. The next steps are described in section 4.3: Complex queries with the QueryBuilder).

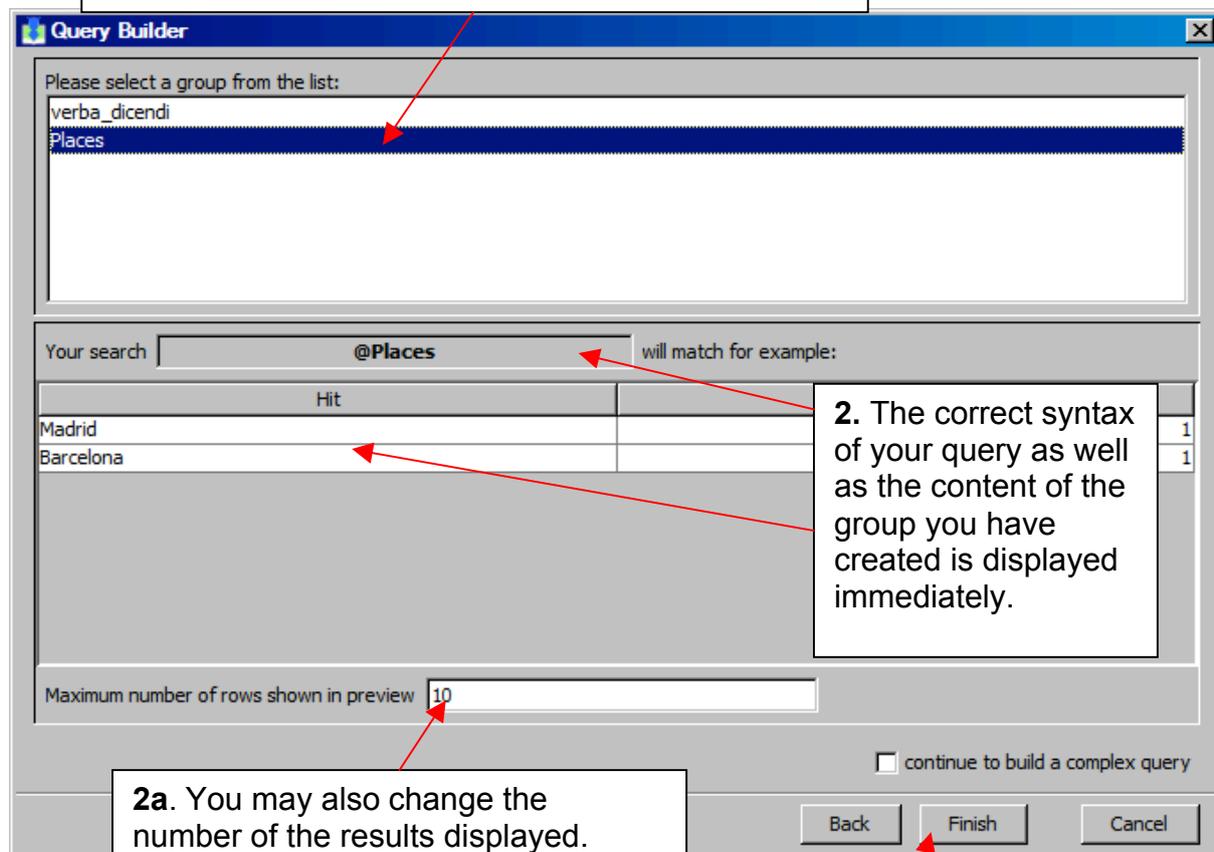
After having clicked „Finish“, the correct query syntax appears in the input field of the “Select from Query” section.

Now you can select the occurrences you would like to continue to work with (e.g., compute distribution analysis cf. 3.2, compute collocate analysis cf. 3.3).

#### 4.2.7 Search “by group”

If you have selected from the QueryBuilder menu that you would like to search “by group”, you may use the following options:

1. Please select one of the groups you have created before.



The screenshot shows the 'Query Builder' window. At the top, it says 'Please select a group from the list:' followed by a list box containing 'verba\_dicendi' and 'Places'. The 'Places' group is selected. Below this, the search syntax '@Places' is shown in a text field, with a note 'will match for example:'. Underneath is a table with two columns: 'Hit' and a column with numbers. The table contains two rows: 'Madrid' with '1' and 'Barcelona' with '1'. At the bottom, there is a text input field for 'Maximum number of rows shown in preview' with the value '10'. A checkbox labeled 'continue to build a complex query' is present, and at the bottom right are 'Back', 'Finish', and 'Cancel' buttons.

2. The correct syntax of your query as well as the content of the group you have created is displayed immediately.

2a. You may also change the number of the results displayed.

3. You can finish building the syntax for your query by clicking “Finish”.

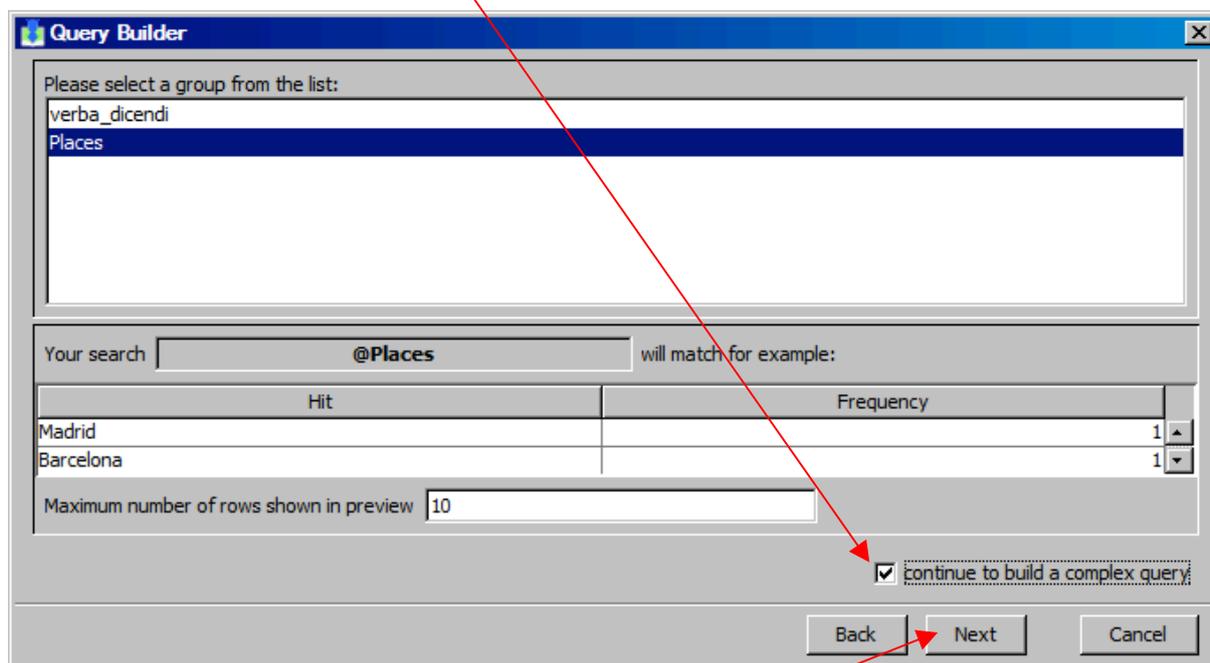
(Please note: You can also add further elements to your query. For example, if you like to search the words/phrases you have selected in combination with other elements such as tags or collocations, please tick off “continue to build a complex query” in order to extend your query. The next steps are described in section 4.3: Complex queries with the QueryBuilder).

After having clicked „Finish“, the correct query syntax appears in the input field of the “Select from Query” section.

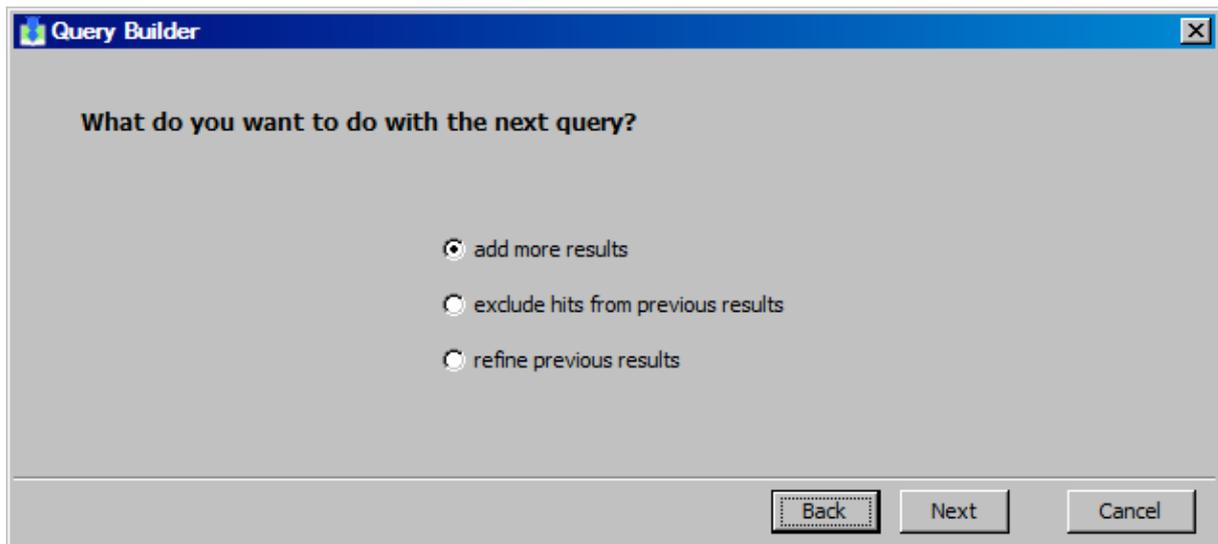
Now you can select the occurrences you would like to continue to work with (e.g., compute distribution analysis cf. 3.2, compute collocate analysis cf. 3.3).

### 4.3 Complex Queries with the QueryBuilder

To compile more complex queries that consist of one than more element, you can tick off “continue to build a complex query” after you have chosen the first element (e.g. group) of your query:



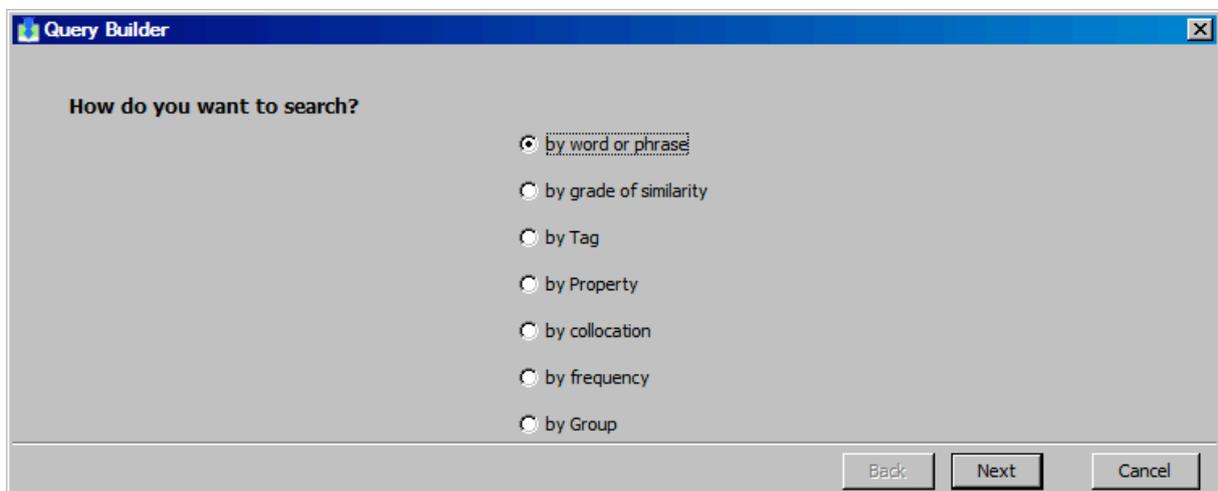
Please click „Next“ to continue.



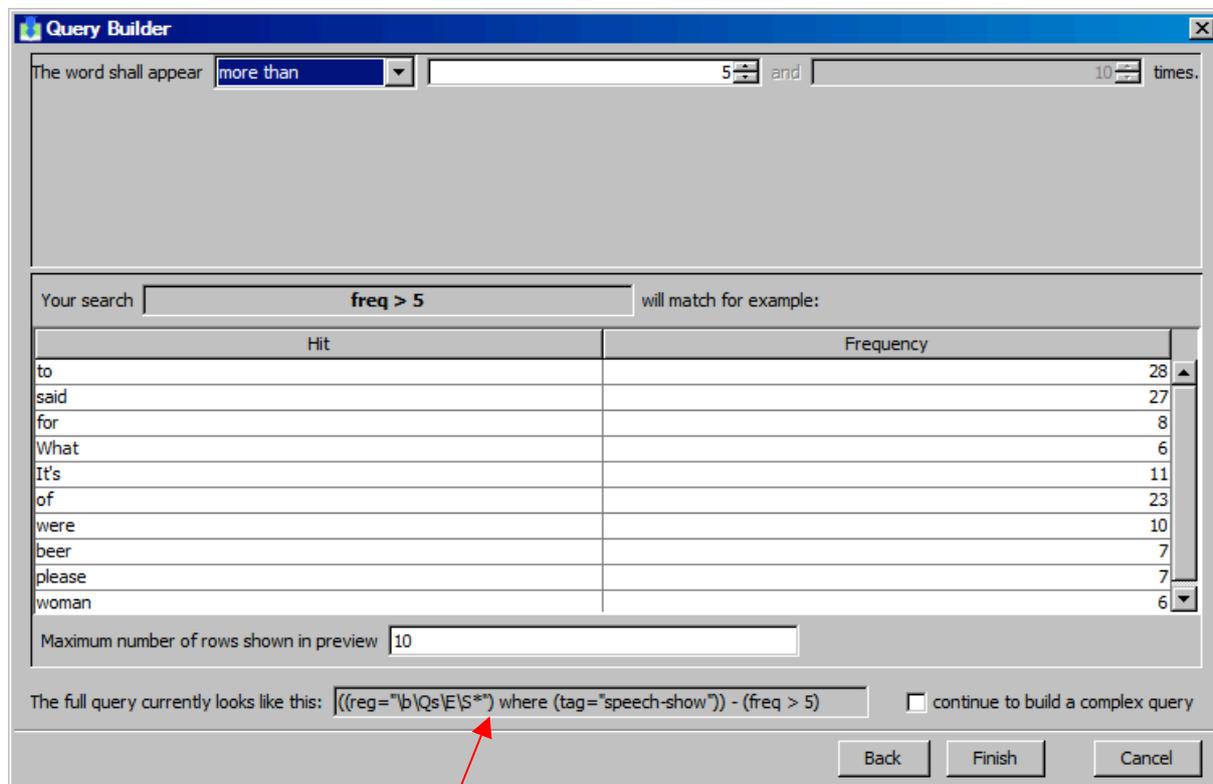
Within the next window, you can decide whether

1. You would like to add another element to your query, such as a tag, group, collocation, property, grade of similarity or frequency which will be searched in addition to the element you have already chosen ("add more results"). For example: You have searched and displayed all occurrences for the word "thought" in the preceding part of the query but you also would like to display all occurrences of the tag "psychological-verbs" at the same time.
2. You would like to exclude results from the preceding part of your query by adding elements that restrict the results of the previously chosen elements ("exclude hits from previous results"). For example: You have searched and displayed all occurrences of the group "flowers" in the previous part of your query but now you would like to exclude all occurrences of "tulip" (which is part of the "flowers"-group) from the query.
3. You would like to refine your query by defining further qualifications for the preceding part of your query. For example: You have searched for the collocations of "girl" in the preceding part of your query, now you would like to restrict this result by adding the condition that you would like to display the collocations which are tagged as "direct-speech" only.

After you have selected one of these three options, please click "Next". Now, you can choose which element you would like to integrate in your query.



No matter which element you choose now, the operations are identical to those described in section 4.2.1 (Search “by word or phrase”) to 4.2.7 (Search “by group”). You can add as many elements to your query as you wish.



When compiling a complex query, the QueryBuilder always shows you how your query looks like at the specific point of creation.

### 4.3.1 Bracket particularities of complex queries

When you compile complex queries, the QueryBuilder also places brackets for you in order to regulate in which sequence the information must be processed (the query syntax functions like a mathematical formula: The information in brackets is processed at first). As it may be the case, the QueryBuilder does not place brackets in the way you would like it to do (because it would change the result of your query), then you have to change the brackets manually. For example:

If you start the query builder with the “search by word/phrase option”, enter an “s” in the field “the word starts with”, refine your query by using the tag “speech show”, select “exclude hits from previous results” and exclude all words with more than five occurrences and “add more results” by using the “search by word/phrase option” again, but this time you enter a “t” in the field “the word ends with”, then you have compiled this query:

**((reg="\b\Qs\E\S\*" where (tag="speech-show")) - (freq > 5) ,  
(reg="\b\S\*\Qt\E(?:\W)"))**

This query would search for all occurrences of words that start with the letter “s” ((reg="\b\Qs\E\S\*") and appear within text parts that are tagged as “speech-show” (where (tag="speech-show")) ) but all of those results are excluded that occur more than five times (- (freq > 5)) and additionally, you search for all occurrences that end with the letter “t” ((reg="\b\S\*\Qt\E(?:\W)")).

As it may be the case, you would like search for all occurrences of words that start with the letter “s” ((reg="\b\Qs\E\S\*") and appear within text parts that are tagged as “speech-show” (where (tag="speech-show")) ) but you do not only want to exclude all words that do occur more than five times but also all words that end with a ‘t’, then you have to change the brackets manually:

**(reg="\b\Qs\E\S\*" where tag="speech-show") - (freq > 5 ,  
reg="\b\S\*\Qt\E(?:\W)"))**

For this query, it is important that you make clear that the first part (“I search for words starting with the letter ‘s’ that appear within the text parts that are tagged as “speech show”) is one unit – that is why, there are brackets around it ((reg="\b\Qs\E\S\*" where tag="speech-show")) – and that you would like to exclude another unit (the words that appear more than five times and the words that end with the letter ‘t’) from it. So, you also put “(freq > 5 , reg="\b\S\*\Qt\E(?:\W)"))” in brackets.

## 4.4 Query syntax

This is a selection of query examples you can enter manually when working with CATMA. If you prefer to compile your queries with the QueryBuilder, you may skip this section.

### 4.4.1 Basic Queries

The most basic type of query that you can run consists of just a single word. Please use inverted commas:

Example: `"bicycle"`

This query would select all occurrences of the word “bicycle” from the text that you have loaded. If you had to use the word list to search for and select “bicycle”, the results would be identical.

Since inverted commas are of special meaning to the query parser, they cannot be used in queries unless they are escaped. To escape them, place a backslash just in front of them:

Example: `"and then he said: \"hello\""`

Regular expressions allow you to search for words, as well as parts of words. Let’s take a look at the syntax for regular expressions:

### 4.4.2 Queries with Regular Expressions

Operator	Description
.	Period/dot – represents any single character.
*	Asterisk – zero or more consecutive occurrences of the preceding character (closure). Usually paired with the period operator, for example: <code>.*</code>
?	Question mark – <a href="#">Reluctant quantification</a> . Usually paired with the period and asterisk operators to prevent these from consuming word boundaries, for example: <code>.*?</code>
<code>\b</code>	Word boundary marker.
<code>\S</code>	Any non whitespace character.
<code>[]</code>	Square brackets – Represents a set of characters (character class). For example, <code>[aeiou]</code> matches any vowel.

^	Circumflex accent – negate. Used to negate a character class. For example, [^aeiou] matches any non-vowel.
-	Hyphen/minus – range. Used to specify a range in a character class. For example, [a-m] matches any character between a and m in the alphabet (inclusive).

**Note:** Since you can also search for punctuation characters in queries, you have to indicate this when you are using them as operators. In this case, please use 'reg=' to indicate a query with regular expressions, as shown in the following examples:

Example: **reg="\bb\b"**

Would select any two-letter word which starts with the letter 'b'.

Example: **reg="\bbe.\*?\b"**

Would select all occurrences of 'be', as well as any word longer than two characters starting with the letters 'be'.

Example: **reg="\bta[^aeiou]\*\b"**

Would select any word starting with the letters 'ta', followed by zero or more non-vowels.

Example: **reg="\bt[d-j]\*.\*?\b"**

Would select any word starting with the letter 't', followed by zero or more occurrences of any letter from 'd' to 'j' (inclusive), followed by any letter zero or more times.

Example: **reg="\S\*ed\b"**

Would select any word ending in the letters 'ed'.

#### 4.4.3 Queries with Groups

To include a group in your query, type in the '@' symbol followed immediately by the name of the group you would like to include.

Example: @cars

Would select everything from the “cars” group.

#### 4.4.4 Queries with Phrases

If you’re looking for a specific sequence of words, or a phrase, you simply need to put this phrase in quotes.

Example: "the sky is blue"

This will cause the Analyzer to look for any occurrence of the phrase “the sky is blue” in the text. Note that only exact matches are returned.

#### 4.4.5 Queries by Frequency

If you take a look at the word list again, you will see that the Frequency column displays the frequency of occurrence for each of the words. You can use this information in your queries as follows:

Example: **freq = 5**

Would select any word which occurs exactly five times.

Example: **freq < 100**

Would select any word which occurs less than a hundred times.

Example: **freq >= 50**

Would select any word which occurs fifty or more times.

Example: **freq = 5-10**

Would select any word which occurs between five and ten times (inclusive). Note that this is the only way a range can be used in frequency queries.

#### 4.4.6 Queries by Similarity

Sometimes it's useful to look for words which are similar to some other word. The Analyzer uses the Ratcliff/Obershelp Pattern Recognition Algorithm for this purpose. (A discussion of the algorithm is beyond the scope of this user manual, but if you are interested please have a look [here](#).) Here is how similarity queries work:

Example: **simil="friend" 70%**

Would select any word which is 70% similar to the word "friend". Note that the percentage symbol is optional.

#### 4.4.7 Queries by Tag

If you have tagged the text you're working on, then you can use this meta-information in your queries as follows:

Example: **tag = "town"**

Would select any word tagged with the tag named "town"

#### 4.4.8 Complex Queries

Complex queries allow you to combine everything you have learned about queries so far, to create even more powerful queries. Let's take a look at how this is done.

#### 4.4.9 Combining Results

You can combine the results of multiple basic queries by simply creating a comma-separated list.

Example: **"bob", "builder"**

Would select any occurrence of "bob", as well as any occurrence of "builder".

Example: **reg="\S.\*ed\b", freq > 20**

Would select any word ending in the letters 'ed', as well as any word which occurs more than twenty times.

#### 4.4.10 Searching by Collocation

Collocate queries allow you to search for things that appear near other things. We just combined the results of searching for "bob" and "builder". Wouldn't it be nice if we could find out where "bob" actually occurs near "builder"?

Example: **"bob" & "builder" 10**

Would select all occurrences of "bob", where "builder" appears within a pre-determined number of words to either side of "bob". 10 is the span size for the collocation, if you omit the span size the default value is 5 words to either side.

Note also that collocate queries are evaluated from right to left, meaning that the left-most word will always be the one that is selected. If you had to write “bob” & “builder” & “house”, this would search for occurrences of “builder” that appear near “house”, and it would then search and select occurrences of “bob” that appear near those occurrences of “builder”.

#### 4.4.11 Exclusion

Exclusions are helpful when you are searching for something which you know will return a large number of results, but you also know that you’re not interested in some of those results. Suppose you have two groups, one called “cars” and another called “germancars”. You want to find all references to cars that are not made in Germany. Here is how you do it:

Example: `@cars - @germancars`

Would select everything from the group “cars” that is not also in the group “germancars”.

Example: `reg="\bte.*?\b" - "test"`

Would select any word beginning with the letters ‘te’, as long as that word is not “test”.

As with collocate queries, exclusions are evaluated from right to left. So writing “te.\* - test - text” would not work, because the Analyzer would try to first exclude instances of “text” from “test”, of which there are none. In this case you would need to rewrite the query using brackets, as follows: `te.* - (test, text)`

#### 4.4.12 Adjacency

This is basically another way of selecting phrases, except that only the first word in the sequence will be selected. For example:

Example: `"an" ; reg="\bu.*?\b"`

Would select all instances of “an” that are directly followed by any word starting with the letter ‘u’.

Example: `"the" ; reg="\bu.*?\b" ; "request"`

Would select all instances of “the” that are directly followed by any word starting with the letter ‘u’, which in turn is directly followed by the word “request”.

Once again, this type of query is evaluated from right to left. So the Analyzer first searches for words starting with the letter 'u' that are followed by "request". It then searches for instances of "an" that are followed by any of the 'u' words from the previous operation.

Unfortunately the Analyzer is not yet capable of selecting an item other than the first item in the sequence. So at the moment you can't search for something that is preceded by something else, for example.

#### 4.4.13 Queries with Refinements

Up to now the only type of refinement we have looked at is the exclusion operation. If, for example, you wanted to search for a word starting with the letters 'te' that occurs less than 100 times, you could write something like this:

Example: **reg="\bte.\*?\b" - freq >= 100**

But what if you wanted to search for something specific which occurs an exact number of times? That would be quite tedious – there is a better way:

Example: **reg="\bte.\*?\b" where freq < 100**

The where operator allows you to refine the results of your query according to some criteria. To be exact, you can refine by frequency, similarity and tag. Here are a few examples:

Example: **"windows" ; "crashed" where freq < 100**

Would select all instances of "windows" that are followed by "crashed", but only if this resulted in less than a hundred instances of "windows" being selected.

Example: **reg="\ba.\*?\b" where simil="Andy" 50%**

Would select all words starting with the letter a, which are 50% similar to "Andy".

Example: **"Markus" where tag= "villain"**

Would select all instances of "Markus" that are tagged with the tag named "villain".

You can also specify multiple refinements. This works as follows:

Example: **"Markus" where tag= "villain", tag= "heir"**

Would select all instances of “Markus” that are tagged both as “villain” AND as “heir”.

Example: **"Markus" where tag= "villain" | tag= "foe"**

Would select all instances of “Markus” that are tagged as “villain” OR as “foe”.

In the case of multiple refinements the comma operator (AND) takes precedence over the pipe/vertical bar operator (OR). Use brackets to preserve the meaning of your refinement in cases where the operator precedence applies another meaning.

#### 4.4.14 Summary of Query Operators

The basic query operators that have not already been summarized are included here for your reference.

Operator	Description
,	Comma - Combines results when used outside of a refinement. When used within a refinement, it has the effect of a logical “AND” operation.
	Pipe – When used within a refinement, it has the effect of a logical “OR” operation.
&	Ampersand – Used to define collocate queries.
-	Hyphen/minus – Used to define exclusions.
;	Semicolon – Used to define adjacency.
<b>where</b>	Used to specify refinements.
<b>@</b>	At – Used to include groups in queries.
<b>freq</b>	Used to define frequency queries.
<b>simil</b>	Used to define similarity queries.
<b>tag</b>	Used to define tag queries.

## 5. Groups

### 5.1 Create groups

To save associated results (words, phrases, tags) in a bundle, you may create a group (e.g., a group that includes all the words of a text that belong to the semantic field “perception”). Groups can also be part of further queries.

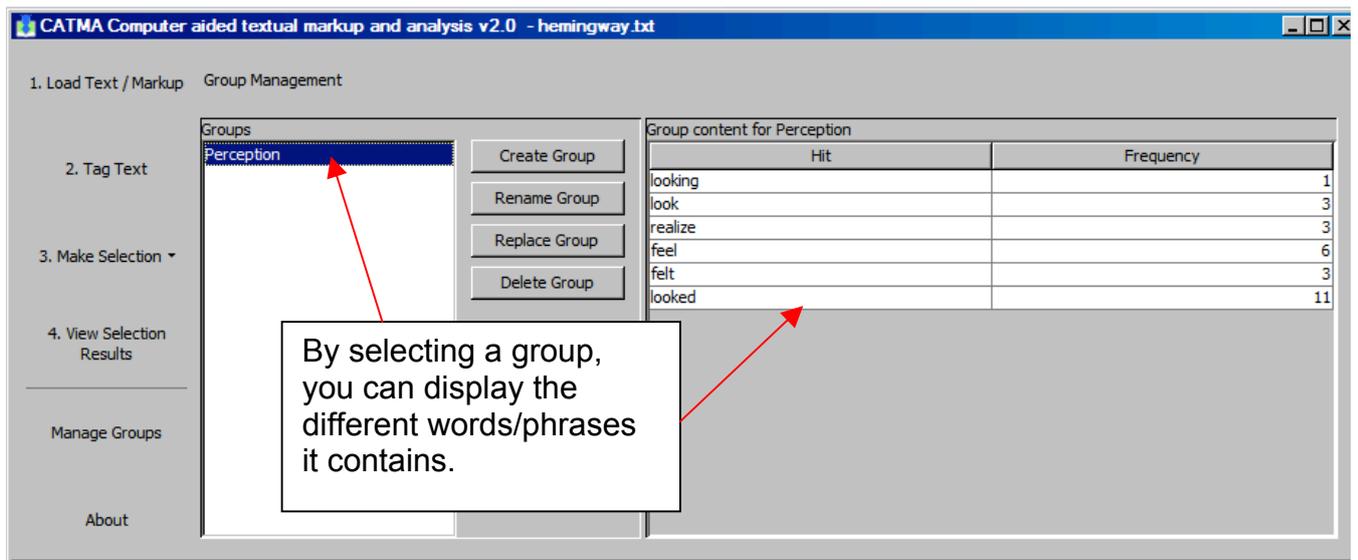
To create a group, you first have to select results from the word list (cf. [3. Wordlist](#)) or a query (cf. [4. Queries](#)). Once you have ticked off all the words/phrases you would like to bundle, change into the “Manage Groups” section to display them in the result view:

The image shows the CATMA v2.0 interface with the 'Group Management' section active. The interface includes a 'Groups' list, a 'Group content' table with columns for 'Hit' and 'Frequency', and a 'Manage Groups' button. A red arrow points from the 'Manage Groups' button to a callout box labeled '1. Click „Manage Groups“'. Another red arrow points from the 'Create Group' button to a callout box labeled '2. Click „Create Group“ to bundle the results you have selected before (from the word list or a query)'. Below the main interface is a 'Group creation' dialog box with the text 'Please enter the new name of the group:' and a text input field containing 'Perception'. A red arrow points from the input field to a callout box labeled '3. Please enter the name of the group. (This is also the name you would use in a query.)'. The dialog box also has 'OK' and 'Abbrechen' buttons.

1. Click „Manage Groups“

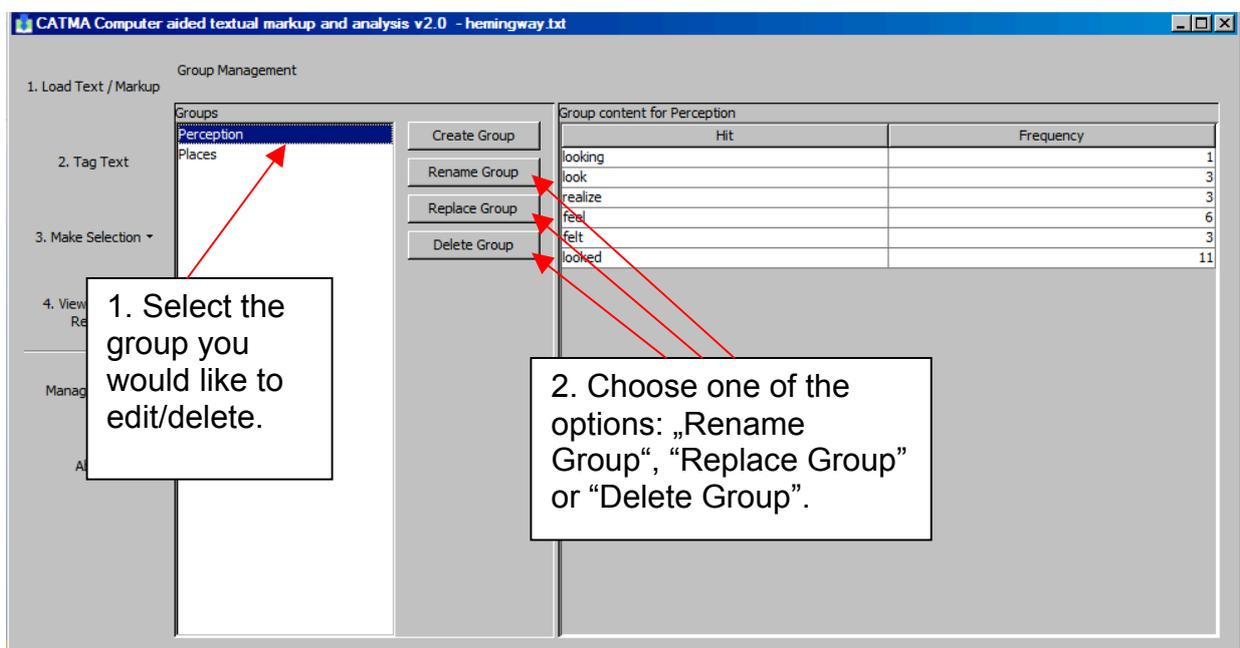
2. Click „Create Group“ to bundle the results you have selected before (from the word list or a query).

3. Please enter the name of the group. (This is also the name you would use in a query.)



## 5.2 Delete or edit groups

Just as with tags (cf. [2.3 Edit Tags](#) and [2.4 Remove tags](#)) or tagsets (cf. [2.8 Remove or edit tagsets](#)), you can also edit or delete groups that you have created. For this operation, please click “Manage Groups” (in the menu on the left-hand side) again:



Please note: The operation “Replace group” is only possible if you have new group content (from the word list or a query) in the result view that is to replace the old content.

Please note: Once you have deleted a group, you will not be able to undo the operation.