



## **CATMA 4 Manual**

**Universität Hamburg  
Institut für Germanistik**

**Version: 2013-04-09  
Author: Janina Jacke (M.A.)**

**<http://www.catma.de>**

# Contents

1. Getting Started	2
1.1 Before You Start	2
1.2 Adding Texts	2
1.3 Exporting Texts	4
1.4 Creating a Markup Collection	4
1.5 Exporting and Importing Markup Collections	5
1.6 Creating Corpora	6
2. Tags	7
2.1 Creating a Tag Library	7
2.2 Creating Tags	8
2.3 Tagging	9
2.4 Editing and Removing Tags or Tagsets	12
2.5 Properties	12
2.5.1 Creating Properties	12
2.5.2 Working with Properties	13
2.6 Tagging Results from Wordlists or Queries	14
2.7 Resuming Work on a Text	14
3. Analyzing the Text	15
3.1 Creating Wordlists	16
3.2 Running Queries	17
3.2.1 The Query Builder	17
3.2.1.1 Simple Queries with the Query Builder	17
3.2.1.1.1 Searching by Word or Phrase	18
3.2.1.1.2 Searching by Grade of Similarity	20
3.2.1.1.3 Searching by Tag	21
3.2.1.1.4 Searching by Collocation	22
3.2.1.1.5 Searching by Frequency	23
3.2.1.2 Complex Queries	24
3.2.1.2.1 Adding More Results	25
3.2.1.2.2 Excluding Hits from Results	25
3.2.1.2.3 Refining Results	25
3.2.1.2.4 Bracket Particularities of Complex Queries	26
3.2.2 Query Syntax	27
3.2.2.1 Simple Queries	27
3.2.2.1.1 Queries with Words or Phrases	27
3.2.2.1.2 Queries with Regular Expressions	28
3.2.2.1.3 Queries with Wildcard	30
3.2.2.1.4 Queries with Similarity	31
3.2.2.1.5 Queries with Tags	31
3.2.2.1.6 Queries with Collocation	32
3.2.2.1.7 Queries with Frequency	32
3.2.2.2 Complex Queries	33
3.2.2.2.1 Combining Queries	33
3.2.2.2.2 Queries with Exclusions	34
3.2.2.2.3 Queries with Adjacency	34
3.2.2.2.4 Queries with Refinements	35
3.2.2.2.5 Summary Chart of Operators for Complex Queries	36
3.3 Key Word in Context	37
3.4 Analyzing Corpora	39
4. Visualizing Results	39
4.1 Visualization with a Simple Graph	39

<a href="#">4.2 Combining Results into One Graph</a> .....	40
<a href="#">4.3 Visualization with Multiple Graphs</a> .....	41

## 1. Getting Started

### 1.1 Before You Start

In contrast to the previous CATMA releases, CATMA 4 is implemented as a web application. General information on the application can be found on [www.catma.de](http://www.catma.de). To use CATMA 4, please visit [www.digitalhumaities.it/catma](http://www.digitalhumaities.it/catma). Log in by clicking the *Login* button in the top right corner of the site. A pop-up window will appear, offering you to log in via Google. If you have got a Gmail account already, please select *Login via Google* and enter your email address and password. If you do not have a Gmail account, please go to [mail.google.com](http://mail.google.com) and create a new account using the *Create Account* button in the top right corner.

### 1.2 Adding Texts

When you log in to CATMA, the Repository Manager window will open automatically, displaying the Repositories Overview tab and the CATMA DB Repository, in a further tab. Here is a brief overview over the different sections of the Repository Manager window when displaying the CATMA DB Repository (fig. 1): In the top left corner you will find a list of your corpora. (For information on how to create corpora, please see section 1.6 Creating Corpora.) The section in the middle of the upper half of the Repository Manager shows a listing of the Source Documents you upload and the respective Markup Collection. (The adding of Source Documents is explained in the following paragraphs. For information on Markup Collections, please see section 1.4 Creating a Markup Collection.) In the top right corner of the Repository Manager you find information on your Source Documents or Markup Collections, e. g. their titles and authors. The bottom left section displays your Tag Libraries. (For information on Tag Libraries, please see section 2.1 Creating a Tag Library.) The bottom right section of the Repository Manager shows information on your Tagsets.

Please click the *Add Document* button in the Documents section in the center of the page.

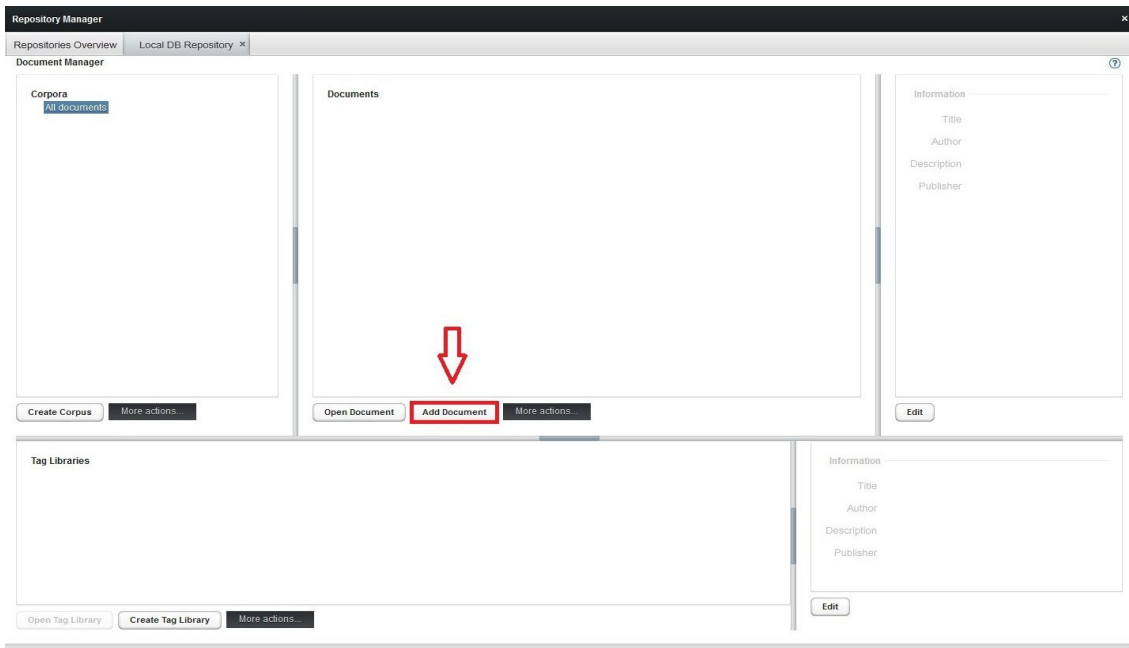


fig. 1

You can choose to either enter an URI to add a text that is accessible via the internet or to upload a file from your computer. If you choose the latter option, please click *Upload local file*. When finished, please click *Next*.

You are now asked to select a file type for your text using the drop-down list on the left-hand side. CATMA tries to determine the file type of the file you have chosen automatically so that it displays the characters of your text correctly. If the wrong file type is preselected, please change the settings to match your source text. The preview of the text according to the current file type selection is shown in the *Preview* box. When finished, please click *Next*.

You have now the possibility to select the language of your text, which CATMA tries to determine automatically as well, and to adjust the Wordlist options. You may, for example, decide that certain character chains shall be treated as either a single word or else as distinct words. These choices may be important when using several of CATMA's automated functions such as counting or search functions. To add an inseparable character chain, please type in the chain in the box at the bottom of the window and click *Add entry*. When finished, please click *Save list* and click *Next*.

You can now enter content details for your text, e. g. its name and author. To finally add your text, please click *Finish*. Your text is now listed in the Documents section of the Repository Manager.

Please note: When you add your first Source Document, CATMA generates a set of example items to get you going: a User Markup Collection to hold your markup and a Tag Library with an example Tagset that contains an example Tag. For more information on these items, please see sections 1.4 Creating a Markup Collection, 2.1 Creating a Tag Library and 2.2 Creating Tags.

If you wish to share a Source Document with somebody, please select the Document, click *More actions...* and click *Share Document*. Enter the email-address of the user with whom you would like to share the Document and select whether they are may modify the Document (*WRITE*) or not (*READ*). Then click *Save*.

If someone shares a Document with you while you are working in CATMA, you may hit the Refresh-button in the top right corner of the Repository Manager to receive it. If someone modifies a shared Document while you are working in CATMA, you also need to hit the Refresh-button in order to receive the latest changes.

### **1.3 Exporting Texts**

It is possible to export documents from CATMA to your computer, e. g. in order to create a backup. To do this, please select a document from the list in the Repository Manager and click *More actions...*, then select *Export Document* and click *Save*. The document is now saved on your computer.

### **1.4 Creating a Markup Collection**

When you add your first text, CATMA automatically creates a Markup Collection for you. This is where your workings, such as your taggings in the text, are stored. If you want to create your own Markup Collection, please select the text for which the Markup Collection is intended in the Documents section. Now click the button *More actions...* in the Documents section and choose *Create User Markup Collection*.

Enter the name of your new Markup Collection and click *Save*. The Markup Collections you created will be displayed if you click the little arrow in front of the according Document in the Documents section of the Repository Manager, then the one in front of *User Markup Collection*.

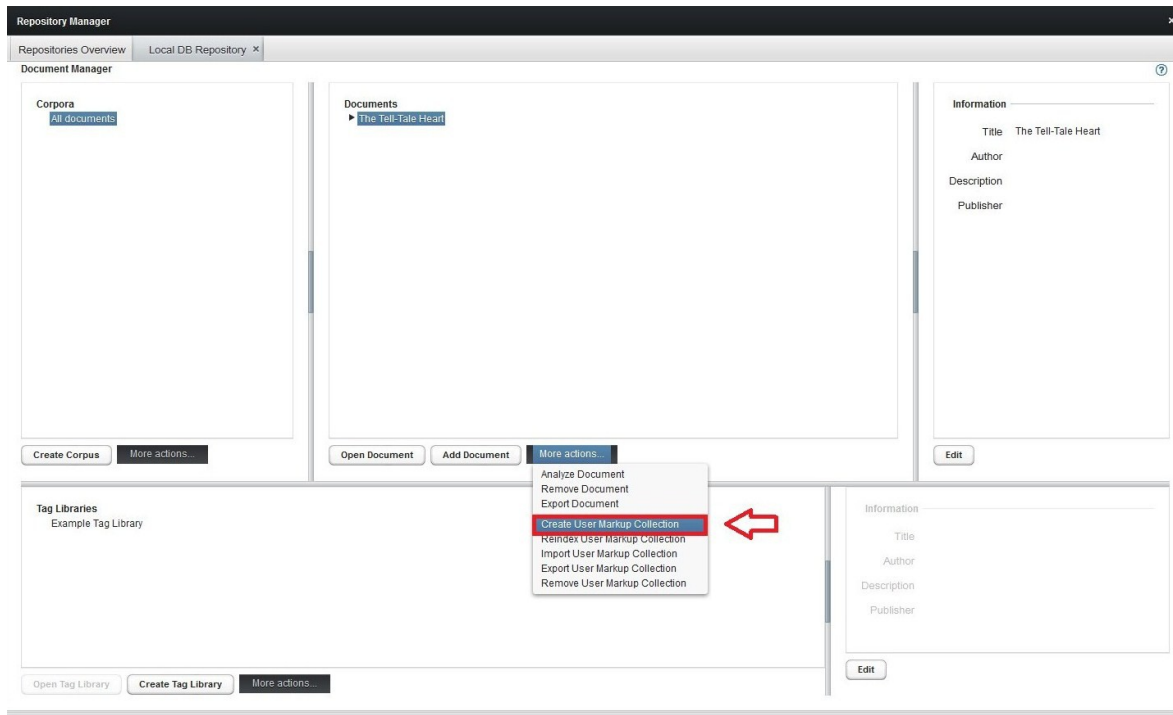


fig. 2

Please note: If you add a text that is already marked up according to TEI or other XML standards, the Markup Collection will automatically be imported to CATMA as Static Markup Collection. More features concerning Static Markup Collections are about to be implemented.

If you wish to share a Markup Collection with somebody, please select the Markup Collection, click *More actions...* and click *Share Markup Collection*. Enter the email-address of the user with whom you would like to share the Markup Collection and select whether they may modify the Markup Collection (*WRITE*) or not (*READ*). Then click *Save*.

If someone shares a Markup Collection with you while you are working in CATMA, you may hit the Refresh-button in the top right corner of the Repository Manager to receive it. If someone modifies a shared Markup Collection while you are working in CATMA, you also need to hit the Refresh-button in order to receive the latest changes.

## 1.5 Exporting and Importing Markup Collections

It is possible to export Markup Collections from CATMA to your computer for backup reasons. To do so, please select a Markup Collection from the Documents section of the Repository Manager. If no Markup Collection is displayed, please click the little arrow in front of the according Document, then the one in front of *User Markup Collection*. Click

*More actions...* and select *Export User Markup Collection*. Please click *Save* now and the Markup Collection will be saved on your computer.

To re-import a Markup Collection, please select the Document to which you wish to attach a Markup Collection from the Documents section of the Repository Manager. Now click *More actions...* and select *Import User Markup Collection*. Please click *Select file* and double-click the Markup Collection you wish to import. CATMA is backward compatible in the sense that all Markup created with prior versions is importable.

## 1.6 Creating Corpora

You have the possibility to pool documents in corpora. The default setting in CATMA pools all the documents you add in the All documents corpus. To create a personal corpus, please click *Create Corpus* in your Repository Manager, enter the name of the corpus and click *Save*.

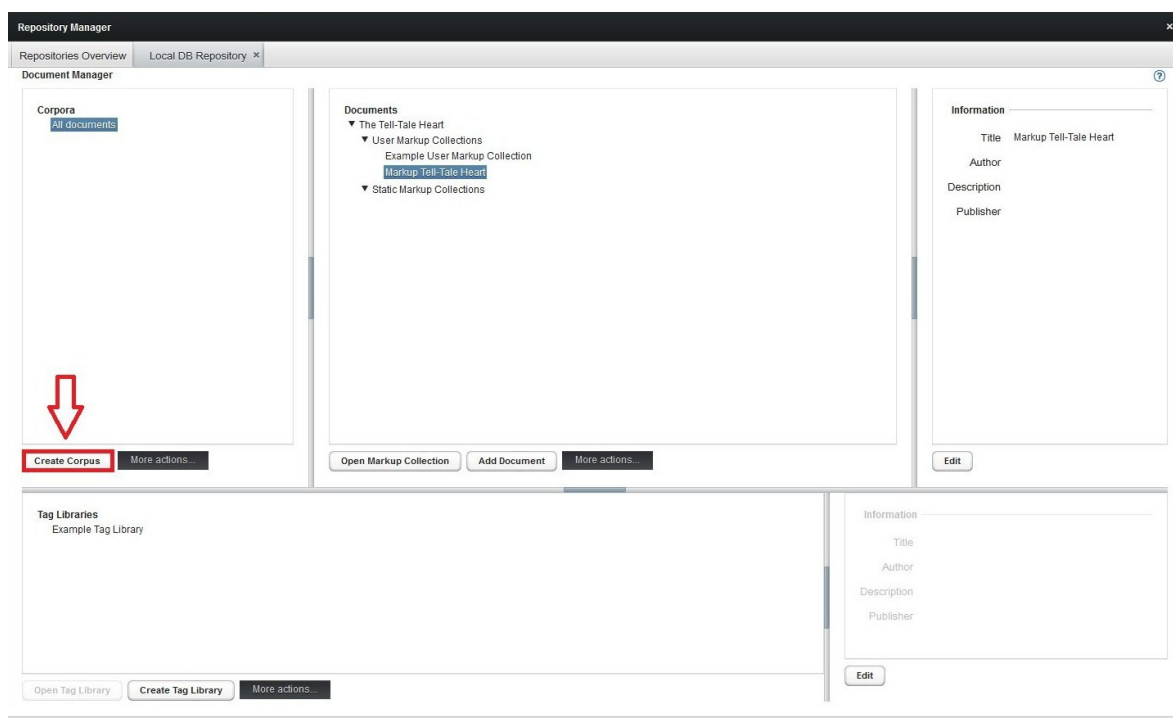


fig. 3

You are now free to select the All documents corpus and drag and drop certain documents to the desired corpus. Every document that is added to a specific corpus is also part of the All documents corpus.

For information on how to analyze whole corpora, please see section 3.4 Analyzing Corpora.

If you wish to share a Corpus with somebody, please select the Corpus, click *More actions...* and click *Share Corpus*. Enter the email-address of the user with whom you would like to share the Corpus and select whether they are may modify the Corpus (*WRITE*) or not (*READ*). Then click *Save*.

If someone shares a Corpus with you while you are working in CATMA, you may hit the Refresh-button in the top right corner of the Repository Manager to receive it. If someone modifies a shared Corpus while you are working in CATMA, you also need to hit the Refresh-button in order to receive the latest changes.

## **2. Tags**

Tags are categories that you wish to apply to the text in order to analyze it. They can be chosen freely so that they meet your purposes and be stored in a hierarchical structure by using the Tag Manager. The parts of the text that you assign to a certain tag are highlighted in the tag's color. All tagging operations are executed using the Tagger module.

### **2.1 Creating a Tag Library**

When you add your first text, CATMA automatically generates a Tag Library for you. This is where your Tags are stored. If you want to create your own Tag Library, please click the *Create Tag Library* button in the Tag Library section in the bottom left corner of your Repository Manager. Enter the name of your Tag Library and click *Save*. The Tag Library is now listed in the Tag Library section of the Repository Manager.



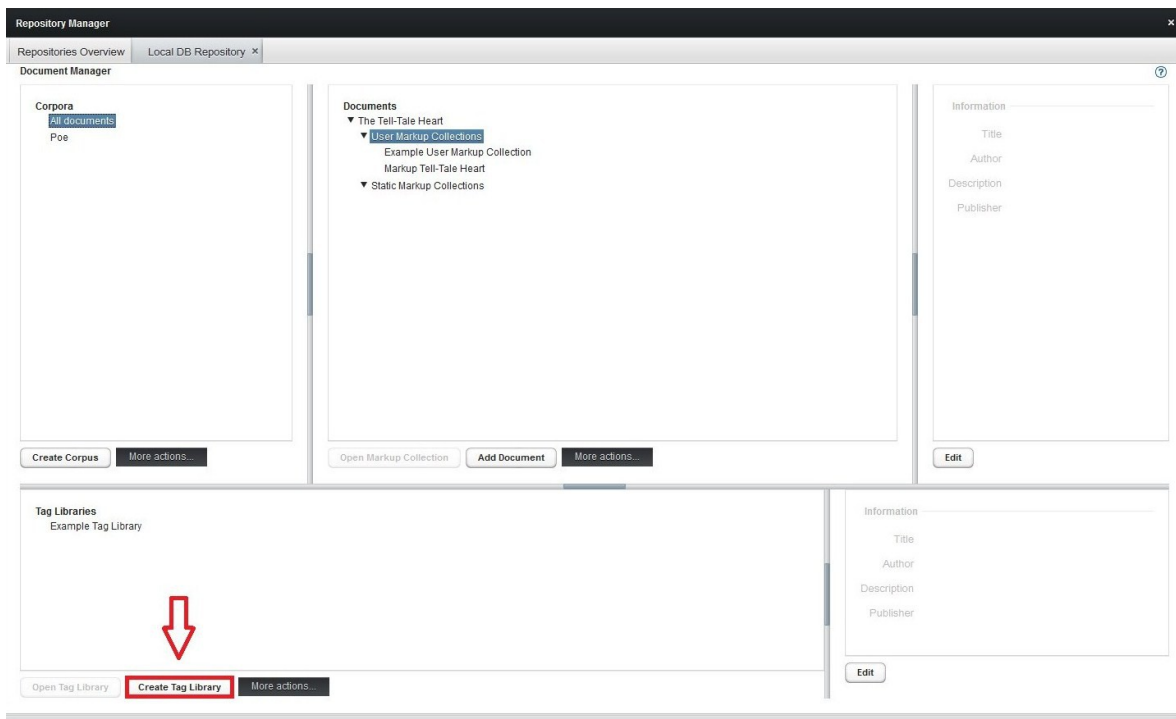


fig. 4

If you wish to share a Tag Library with somebody, please select the Tag Library, click *More actions...* and click *Share Tag Library*. Enter the email-address of the user with whom you would like to share the Tag Library and select whether they may modify the Tag Library (*WRITE*) or not (*READ*). Then click *Save*.

If someone shares a Tag Library with you while you are working in CATMA, you may hit the Refresh-button in the top right corner of the Repository Manager to receive it. If someone modifies a shared Tag Library while you are working in CATMA, you also need to hit the Refresh-button in order to receive the latest changes.

## 2.2 Creating Tags

When you add your first text, CATMA automatically creates an example Tagset containing an example Tag for you. To create your own Tagsets and Tags, please select a Tag Library from the Tag Library section of your Repository Manager and click *Open Tag Library* in order to open the Tag Manager window. (Please note: If you open more than one Tag Library in one working session, the Libraries will be displayed as different tabs in the Tag Manager window. If the window is closed and re-opened in one working session, either via the menu bar or via the *Open Tag Library* button, all tabs that were opened once will be displayed again.)

The Tagger window displays your Tagsets, your Tags and their colors as well as their Properties in the left-hand section. The right-hand section is used to create, remove and edit

Tagsets, Tags and Properties.

Tags can only be created when you create a Tagset first. To do so, click the *Create Tagset* button and enter the name of your Tagset. The Tagset you created is now listed in the Tagset section of your Tag Manager. You can now select a Tagset and click the *Create Tag* button. Please enter the name of your Tag, choose a color and click *Save*. The color you choose will be used to highlight the sections you tagged in the text. The Tag you created will be displayed in the Tagset section of your Tag Manager if you click the little arrow in front of the according Tagset.

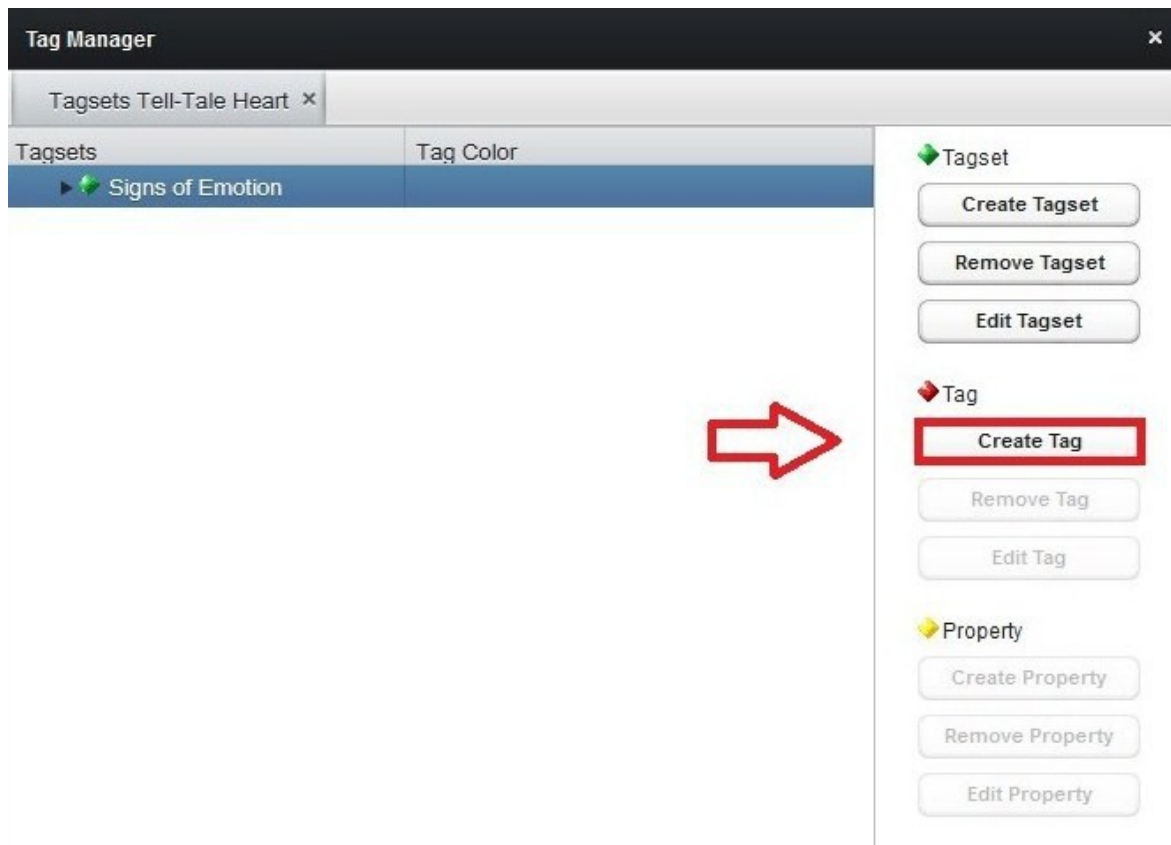


fig. 5

You can also create Subtags for Tags by selecting the relevant Tag in the Tag Manager instead of the Tagset, then click *Create Tag*. The Tags you create in this manner will automatically be ranked as Subtags to the Tag you selected beforehand.

## 2.3 Tagging

To start the actual tagging, click the little arrow in front of the text you want to tag in the Documents section. Then click the arrow in front of *User Markup Collections* and select the Markup Collection in which you would like your workings to be stored. Click *Open Markup Collection* in order to open the Tagger window.

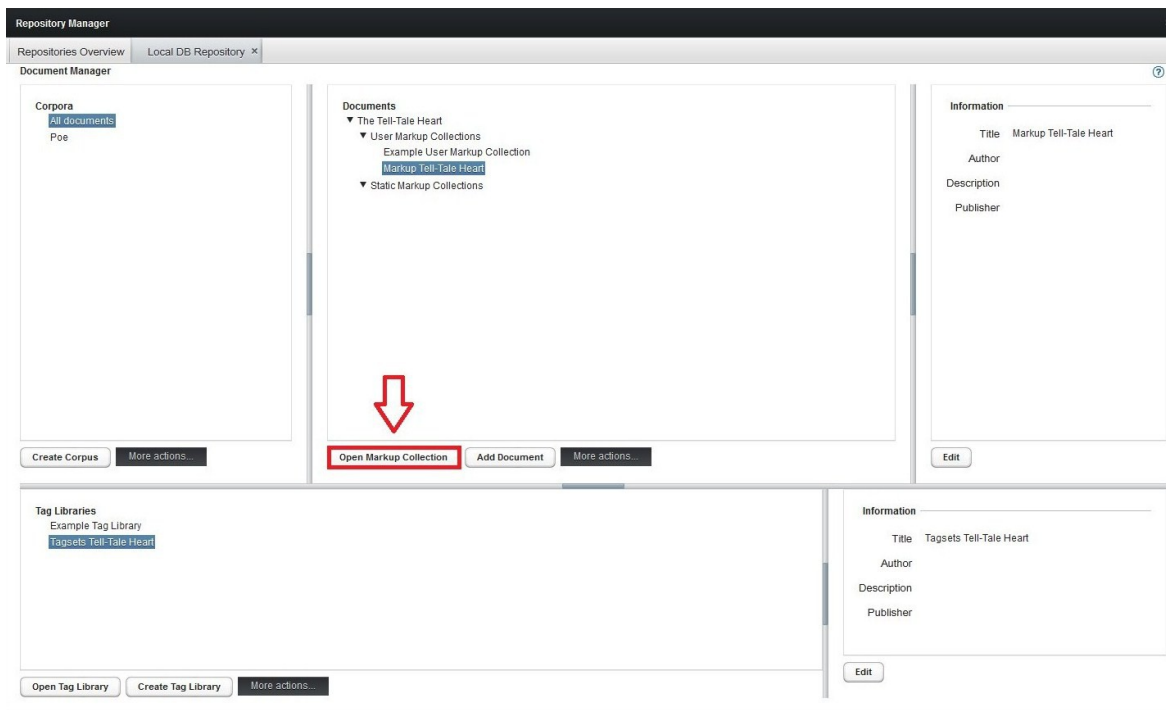


fig. 6

Here is a brief overview over the different sections of the Tagger window (fig. 9): The Tagger window displays your text on the left-hand side. (To adjust how much of your text is displayed on one page, please use the scale below the text.) In the top right section, you find your active Tagsets and active Markup Collections in different tabs. The bottom right section shows your writable Markup Collections and the Tag instances of specific selected parts of the text. (If you click on a part of the text that you have tagged, the Tag instances will be listed in this section, making it possible for you to remove Tag instances or to add Property values to Tags. For information on Properties, please see section 2.5 Properties.) (Please note: If you open more than one document/Markup Collection in one working session, the documents will be displayed as different tabs in the Tagger window. If the window is closed and re-opened in one working session, either via the menu bar or via the *Open Document/Open Markup Collection* button, all tabs that were opened once will be displayed again.)

In addition to the Tagger window, you need to open the Tag Library that contains the Tags you want to use for tagging the text. To do so, please select a Tag Library from the Tag Library section of the Repository Manager and click *Open Tag Library* in order to open the Tag Manager window. Now drag and drop the Tagset you want to use from the Tag Library to the Tagsets section of the Active Tagsets tab of the Tagger.

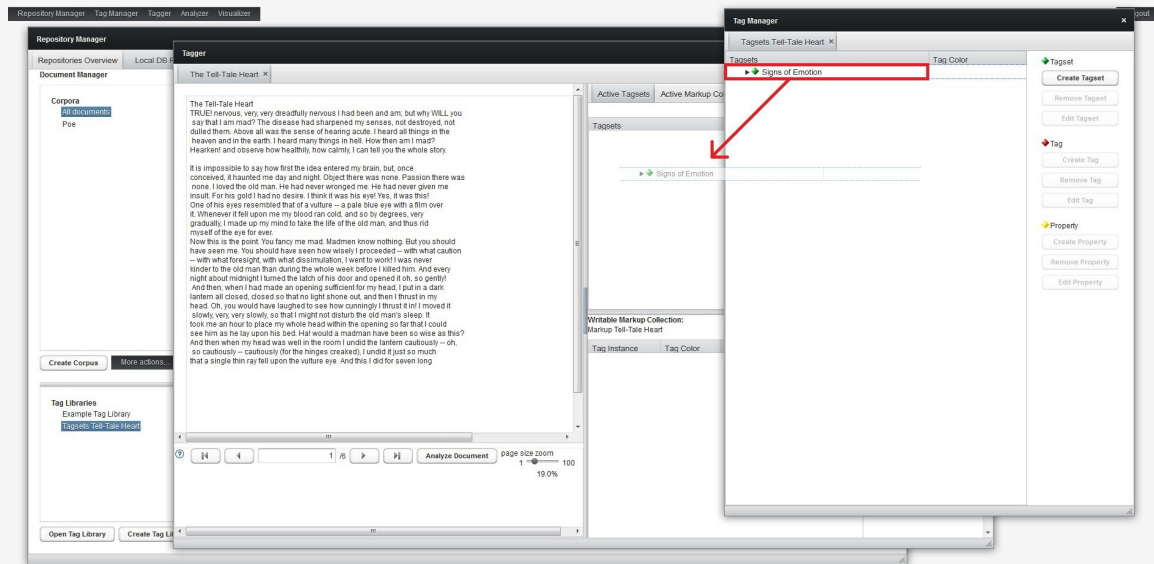


fig. 7

Click the little arrow in front of the Tagset in the Tagger in order to display the Tagsets you created. Now click the Active Markup Collections tab of the Tagger, select the Markup Collection in which you want to store your workings (if not displayed yet, please click the little arrow in front of *User Markup Collection*) and tick off the box in the *writable* column.

You can now select parts of your text on the left-hand side of the Tagger and then click on the Tag Color button in the Active Tagsets tab on the right-hand side of the Tagger in order to highlight and tag the according parts of the text.

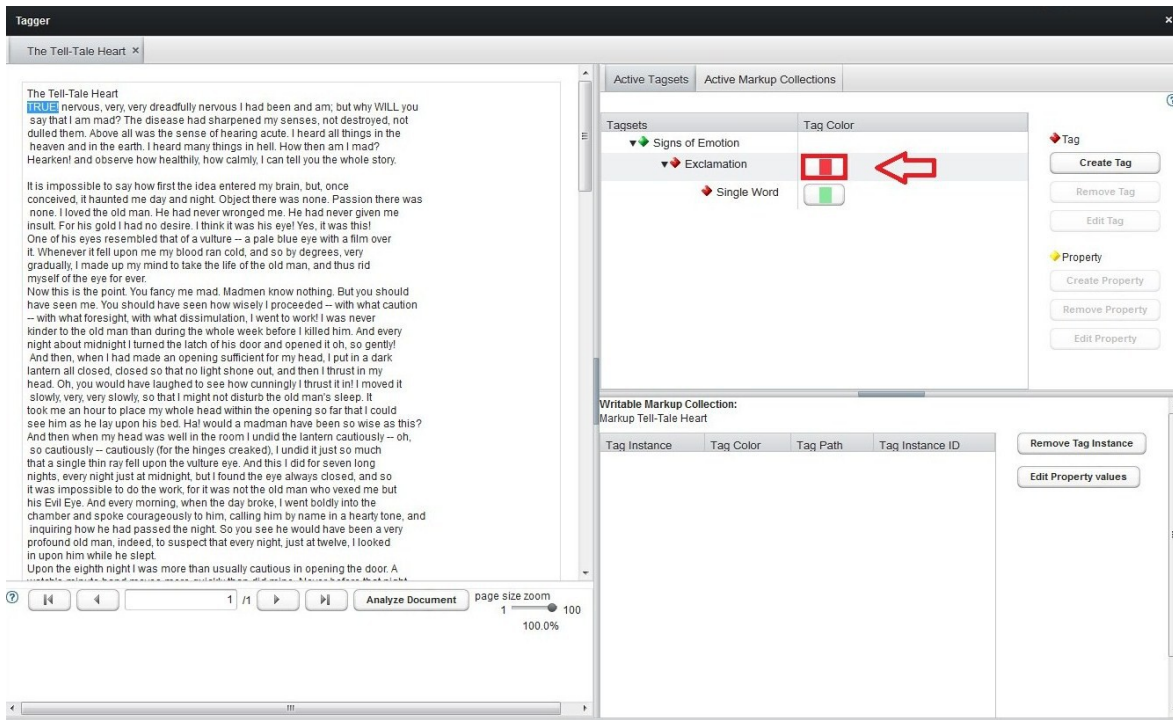


fig. 8

To create a discontinuous tagging in the text, i. e. one Tag instance that consists of two or more parts, please select the first part of the text you would like to tag, then hold Control and select a further part of text. Then click on the Tag Color button in the Active Tagsets tab on the right-hand side of the Tagger.

If you wish to close a Tagset in the Active Tagsets Section or a User Markup Collection in the Active Markup Collections section of the Tagger, please right-click the Tagset or the Markup Collection and select *close*.

## **2.4 Editing and Removing Tags or Tagsets**

To edit a Tag, please select the Tag in the Tag Manager or, if you already dragged and dropped the according Tagset to the Tagger, in the Tagger, click *Edit Tag* and change name or color. To remove a Tag together with all the instances of the Tag in the text, select the Tag in the Tag Manager or Tagger and click *Remove Tag*. If you wish to remove only a Tag instance from the text, please click on the tagged part of the text displayed on the left-hand side of the Tagger, then select the Tag instance you want to remove from the list in the bottom right section of the Tagger and click *Remove Tag Instance*.

To edit the name of a Tagset, please select the Tagset in the Tag Manager or the Tagger and click *Edit Tagset*. You can also edit a Tagset by adding or deleting Tags in the Tag Manager or Tagger. Every change is automatically implemented in both locations. To remove a Tagset, select a Tagset and click *Remove Tagset*.

## **2.5 Properties**

You can also assign Properties to Tags, which enables you to differentiate or refine its meaning. Properties also provide the possibility to structure the Tagsets horizontally in addition to the vertical structuring via Subtags, because you can choose multiple Values for a Property that is assigned to a Tag.

### **2.5.1 Creating Properties**

To add a Property, please open either the Tagger or the Tag Manager via the menu bar and select a Tag. If no Tags are displayed, please click the little arrow in front of a Tagset. Now click the *Create Property* button and enter the name of your Property in the pop-up window. To add a Value, please enter the desired Value in the field at the bottom of the pop-up window and click the + button. The Value is now displayed in the list of possible values. If

you want to add more Values, proceed in the same manner until you are finished. If you want to remove a Value from the list, please select the Value and click the – button. When finished, please click *Save*.

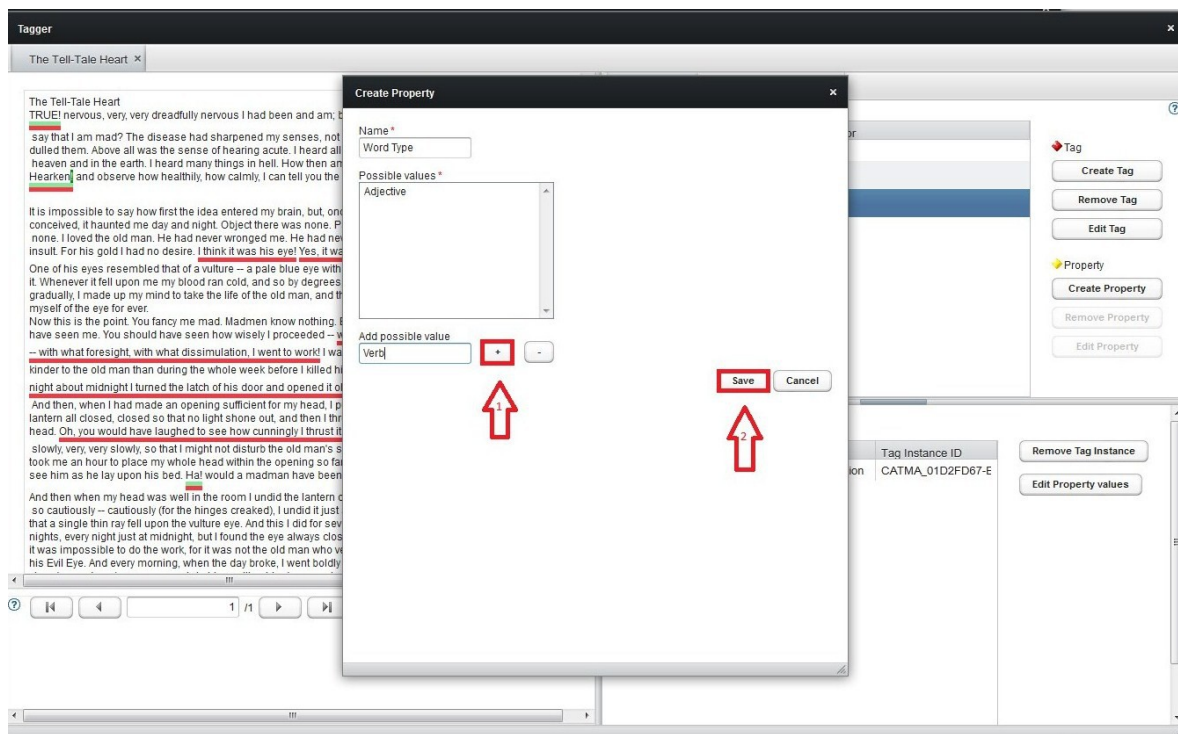


fig. 9

## 2.5.2 Working with Properties

If you would like to assign a Property Value to a Tag instance in the text, please open a Markup Collection and drag and drop a Tagsets that includes a Tag with a Property with different possible Values from the Tag Manager to the Active Tagsets section of the Tagger. If the Tagset does not include a Property yet, the Property and possible Values can also be created in the Active Tagsets section of the Tagger. Please click on the part of the text that contains the Tag instance to which you would like to assign a Property Value. The Tag instances of the part of the that you clicked on are now displayed in the bottom right section of the Tagger. Now click the little arrow in front of the chosen Tag instance and select the chosen Property. Click the *Edit Property* button, select one or more Values from the list and click *Save*.

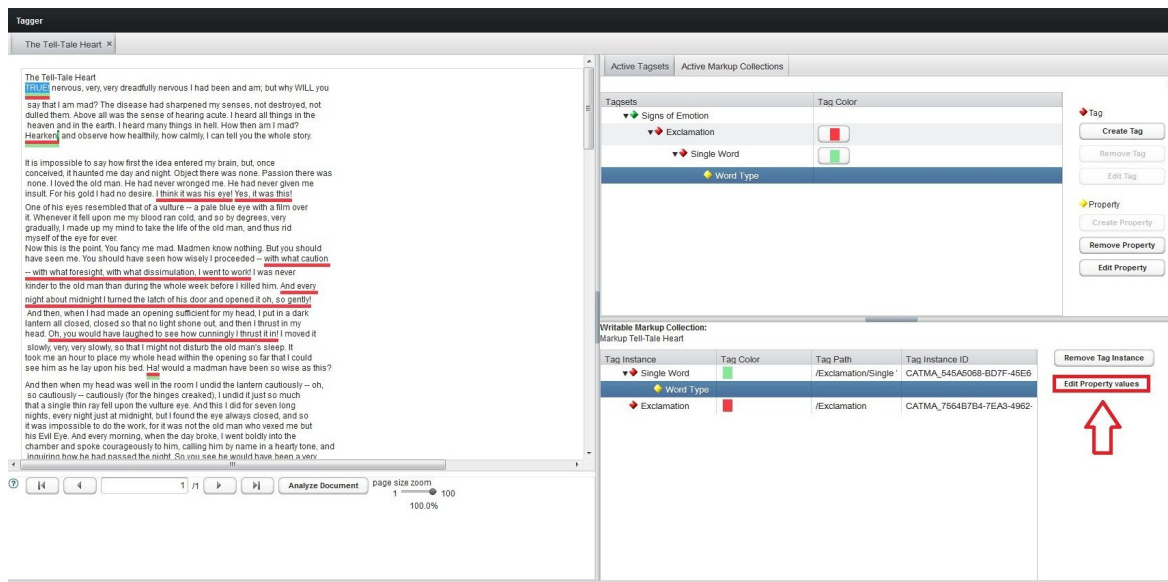


fig. 10

## 2.6 Tagging Results from Wordlists or Queries

For a description of how to tag search results, please see section 3.3 Key Word in Context.

## 2.7 Resuming Work on a Text

If you wish to resume work on a document that you have already started to tag in a previous working session, please select a User Markup Collection from the Documents section of the Repository Manager. If no Markup Collection is displayed, please click the little arrow in front of the according Document, then the one in front of *User Markup Collection*. Please click *Open Markup Collection* in order to open the Tagger window which now displays your text and the active Markup Collections. To display your previous taggings in the text, please tick off the *visible* box behind the Tagsets in the Active Markup Collections section of the Tagger. You also have the possibility to only display the instances of a single Tag or a few chosen Tags in the text. This may be useful when you have already done very many taggings and would like to simplify the overview.

If you want to resume tagging, please select the Active Tagsets tab of your Tagger, which will show no active Tagsets. (Please note: The Tagsets you used to tag your text in the previous working session are not displayed in the Tagger automatically. This is because you might have modified the Tagsets using the Tag Manager after having tagged the text. If the modified Tagsets were now automatically imported to the Tagger, your User Markup Collection would be changed automatically as well, which would make it impossible for you to view the previous working results.)

To import a Tagset, please open the Tag Manager by selecting a Tag Library from the bottom left section of the Repository Manager and clicking *Open Tag Library*. Drag and drop the desired Tagset from the Tag Manager to the Active Tagsets section of the Tagger. In case you have modified the Tagset after the last tagging operation, you will be asked whether you would like to update the attached Markup Collection. If you would, please click *OK*.

### 3. Analyzing the Text

The Analyzer module enables you to apply different analyzing functions to the text. It may, for instance, be used to generate Wordlists of your text, listing the words of the analyzed text in alphabetical order or according to their frequency. Many analyzing functions may also be used in combination with the other modules: You can, for example, use the Analyzer to search for a word or a phrase and then tag your search results. You may also generate a listing of all occurrences of a certain Tag in the text and then use the Visualizer module to show a distribution chart of the Tag in the text (for information on the Visualizer module, please see section 4. Visualizing Results).

To use any analyzing function, you can open the Tagger and click the *Analyze Document* button under the text in order to open the Analyzer window.

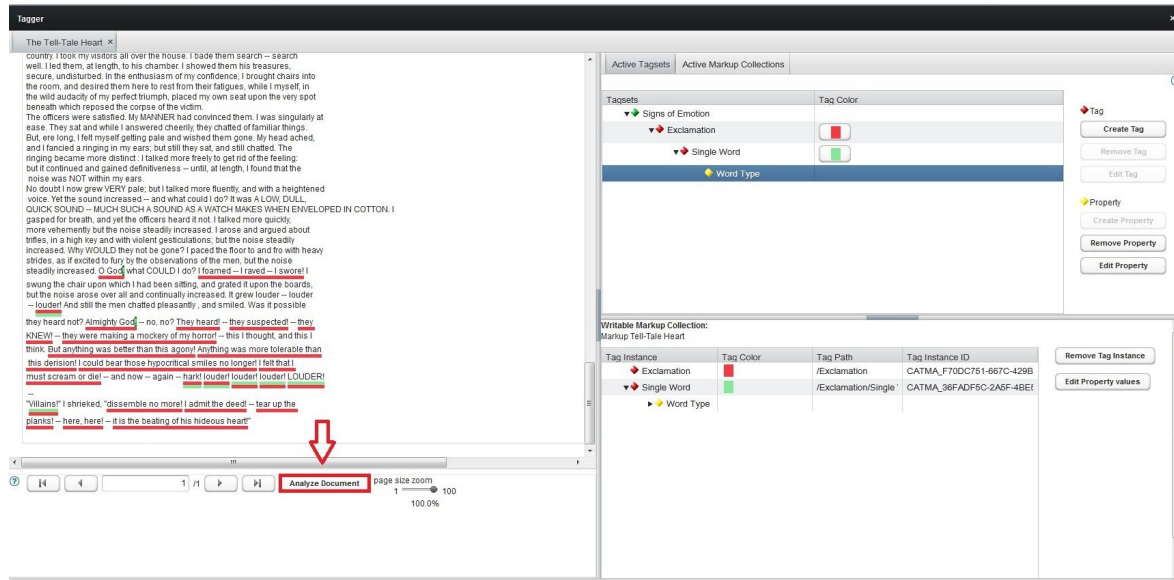


fig. 11

Once you have opened the Analyzer this way, you can also close and re-open it using the menu bar. You can also select the document you wish to analyze in the Documents section of the Repository Manager, click *More actions...* and select *Analyze Document*. You are



now asked which documents (Source Documents and Markup Collections) you wish to include in the analysis. Tick off the boxes behind the chosen documents and click *Ok*. Please note that at least one Source Documents must be selected for the analysis.

Here is a brief overview over the different sections of the Analyzer window (fig. 13): In the top left section, you have the possibility create Wordlists (please see section 3.1 Creating Wordlists) and to run Queries, either by using the Query Builder or by typing in the Queries directly (please see section 3.2 Running Queries). In the top right section, you find an overview over the different Source Documents and Markup Collections that are included in your current analysis. The bottom left section displays the Wordlist or Query results together with their frequencies (please see sections 3.1 Creating Wordlists and 3.2 Running Queries). The bottom right sections lists chosen key words in their contexts (please see section 3.3 Key Word in Context).

### 3.1 Creating Wordlists

To display a Wordlist, please click the *Wordlist* button in the top left section of the Analyzer. The Wordlist, i. e. every word type of your text, is now displayed in the bottom left section of the Analyzer window. The frequency of a word in the text is represented by the figures behind the phrases. Beneath the listing you find the number of word types (total count) on the left-hand side and the number of word occurrences (total frequency) on the right-hand side.

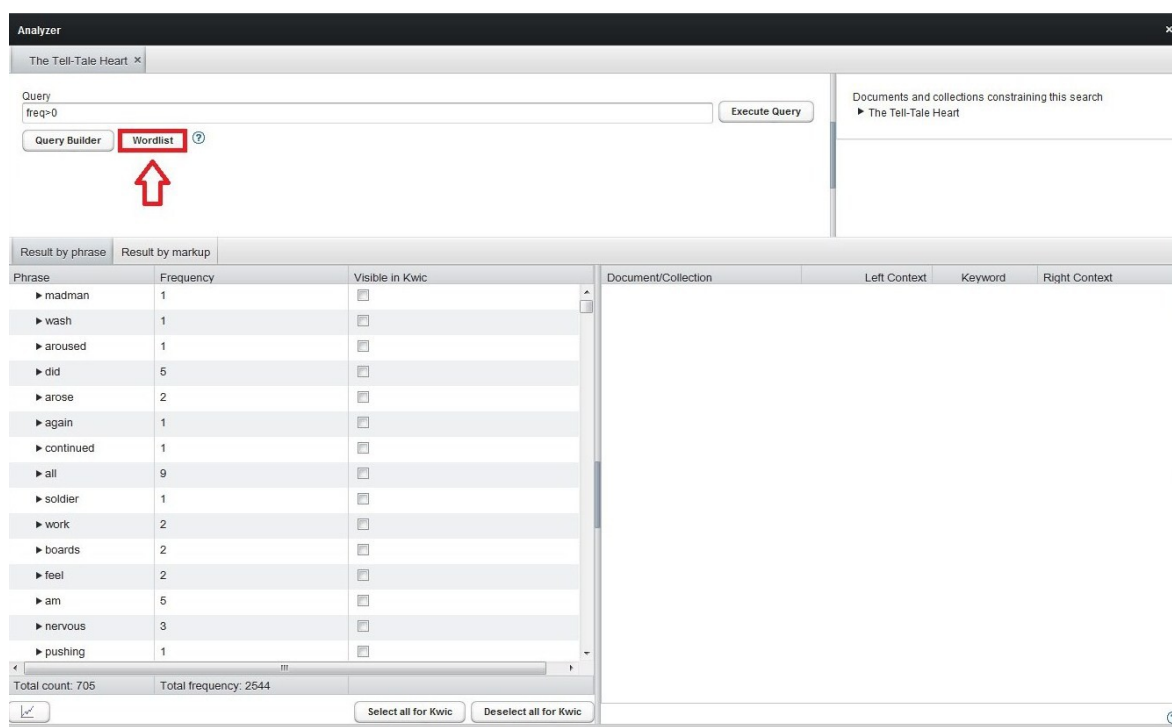


fig. 12

You can now choose the criterion according to which the Wordlist is supposed to be arranged. If you would like the words to be displayed in alphabetical order, please click *Phrase* on top of the first column of the Wordlist. A little arrow pointing upwards appears behind *Phrase* and the Wordlist is generated in alphabetical order. If you want to reverse the listing, click *Phrase* again. The arrow now points downwards and the Wordlist is displayed in reverse alphabetical order. You can also decide to arrange the list according to the frequency of the listed phrases. To do so, please click *Frequency* on top of the second column of the Wordlist. The words are now displayed according to frequency, listing the least frequent words first. If you want to reverse the order, please click *Frequency* again.

For information on how to tag words from the Wordlist, please see section 3.3 Key Word in Context. For information on how to visualize the distribution of words from the Wordlist, please see section 4. Visualizing Results.

### **3.2 Running Queries**

The Analyzer module offers you the possibility to directly search for words, parts of words, word combinations or Tags in the text. The results of these Queries are available for further operations, like tagging or visualization (For information on how to tag Query results, please see section 3.3 Key Word in Context. For information on how to visualize the distribution of Query results, please see section 4. Visualizing Results.)

There are two different ways of running Queries: You can either enter the Queries directly or use the Query Builder. Please note that the use of special operators and a specific syntax structure is required for the Queries. It is, of course, possible to learn these peculiarities—they are described in section 3.2.2 Query Syntax. It may be easier though to use the natural language based Query Builder that helps you build your Query step by step and intuitively—and transforms your Query so that it meets the requirements of the syntax.

## 3.2.1 The Query Builder

### 3.2.1.1 Simple Queries with the Query Builder

To use the Query Builder, please open the Analyzer (either by clicking the *Analyze Document* button in the Tagger or, if you have opened the Tagger already during your current working session, via the menu bar) and click the *Query Builder* button in the top left section of the Analyzer. You can now decide how you want to search: by word or phrase, by grade of similarity, by Tag, by collocation and by frequency.

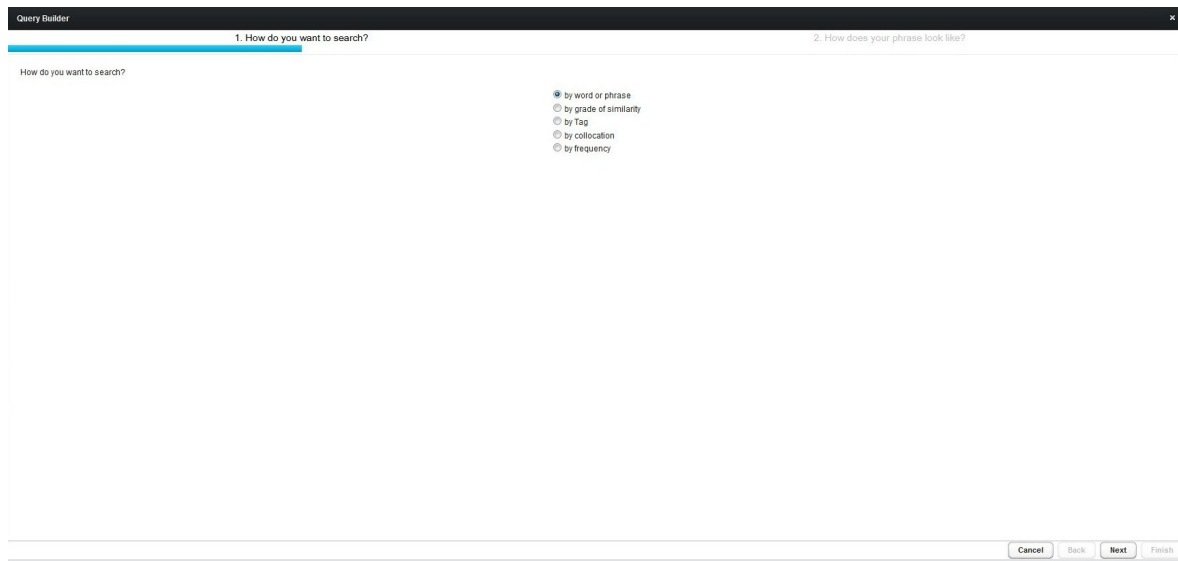


fig. 13

#### 3.2.1.1.1 Searching by Word or Phrase

One option is to search by word or phrase. Please select *by word or phrase* and click *Next*. You can now decide whether you want to search for one exact word or if you want to search for parts of words. The latter option may be useful if you, for example, want to search for every inflected form of a word. You then may want to choose the option to enter only the beginning of the word for the Query, leaving out the inflected part of the word.

For example:

You may want to search for every form of the word “eye”, e. g. “eye” and “eyes”. Just type in that the words you search starts with “eye”. (Note that this Query may also list complex nouns that start with “eye”, e. g.. “eyeball”.)

Please enter the phrase you want to search and click *Show in preview*. The syntax of your current Query and a preview of your current Query results are now displayed in the bottom section of the Query Builder.

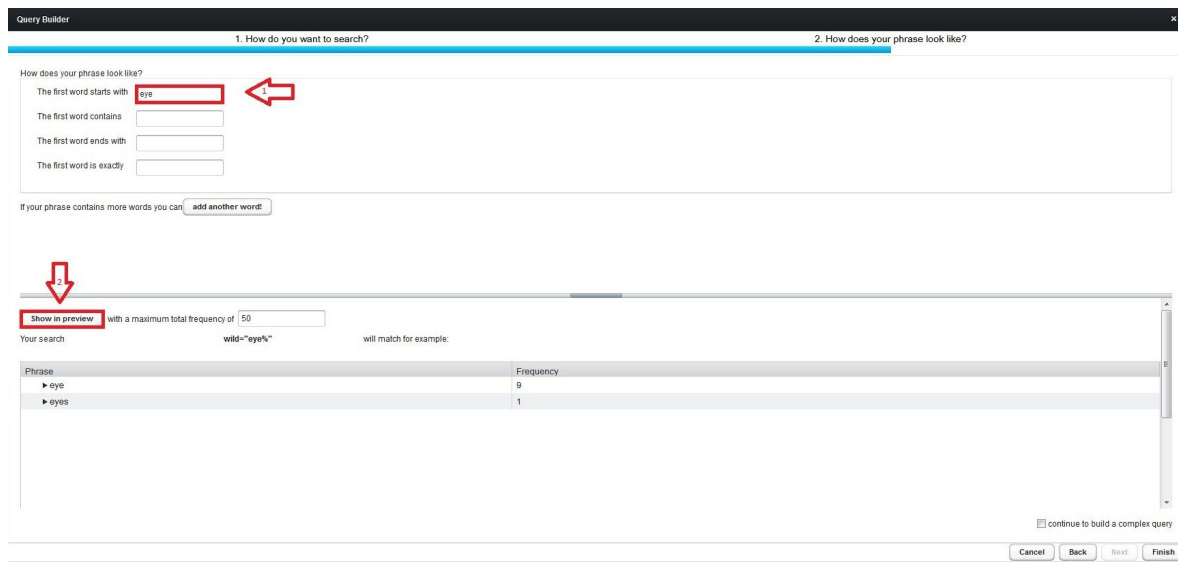


fig. 14

If you would like to change the maximum number of results displayed in the preview, please enter the desired maximum in the bottom section of the Query Builder and click *Show in preview*. If you choose not to enter the exact word but one of the other options (like *The first word starts with*) and if there are actually different word types that meet your Query among the results, different words will be displayed in different rows, together with their respective frequency. At the bottom of the preview section you find the total count of word types that meet your Query on the left-hand side, and the total frequency of occurrences in the text on the right-hand side.

If the phrase you are looking for contains more than one word or item, please click the *add another word* button at the bottom of the upper section of the Query Builder. You can now enter a second word and determine the desired position after the first word.

For example:

You want to search for any form of the verb “eye” that is directly followed by a full stop. Just type in that your first word starts with “eye”, click *add another word*, type in that your second word is exactly “.” and that its position is one word after the previous word.

If your Query contains more than two words, proceed in the same manner.

When you are finished, please click *Finish* to display your Query results in the Analyzer window where they are available for further use. The results are displayed in the same manner as in the preview. For information on how to extend your Query, please see section 3.2.1.2 Complex Queries.

(Please note: If you run more than one Query in one working session, the search results will be displayed in different tabs of the Analyzer window. If the window is closed and re-opened during this working session, all tabs will be displayed again.)

### 3.2.1.1.2 Searching by Grade of Similarity

Another option for your Query is to search by grade of similarity. To do so, please select *by grade of similarity* and click *Next*. You may now enter a word in the top left corner of the Query Builder—the results of your search will include the word you entered plus different words that are, to some degree, similar to the word you entered. The Analyzer uses the Ratcliff/Obershelp Pattern Recognition Algorithm for this purpose. (A discussion of the algorithm is beyond the scope of this manual, but if you are interested, please have a look [here](#).) You can determine the desired grade of similarity for the search results by using the *Grade of similarity* scale on the left-hand side of the Query Builder. The scale works by percentage.

For example:

You may want to search for words that show 70 percent of similarity to the word “man”. Type in that the word is similar to “man” and adjust the grade of similarity scale to 70.

Please note: If you choose 100 as grade of similarity, only the exact word you entered will be displayed in the results.

To display the preview of your current search results, please click *Show in preview*. The syntax of your current Query and a preview of your current Query results are now displayed in the bottom section of the Query Builder.

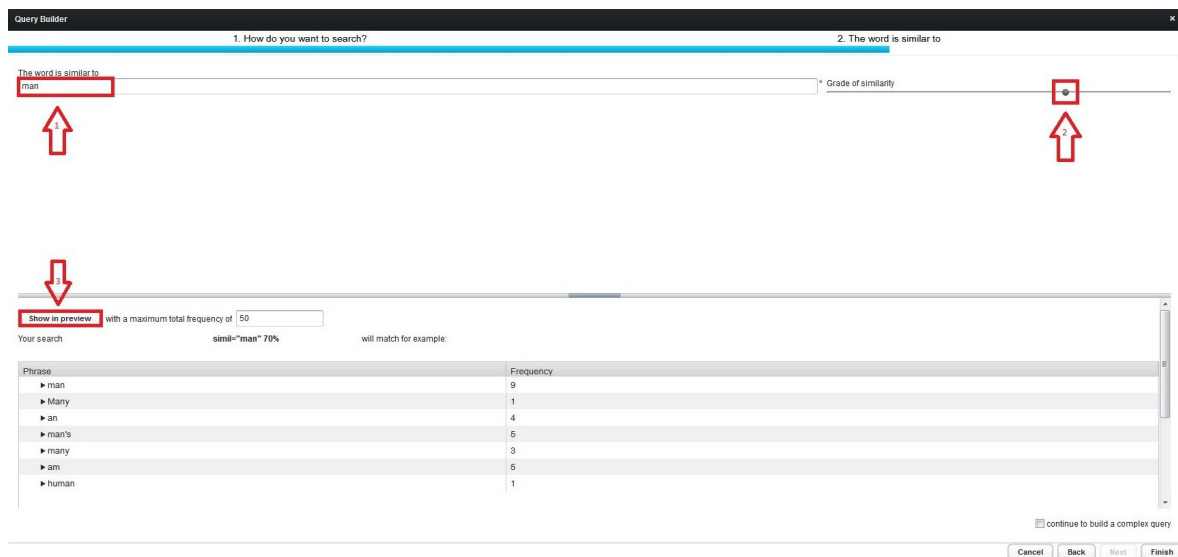


fig. 15

If you would like to change the maximum number of results displayed in the preview, please enter the desired maximum in the bottom section of the Query Builder and click *Show in preview*. If your Query results include different word types that meet your Query,

different words will be displayed in different rows, together with their respective frequency. At the bottom of the preview section you find the total count of word types that meet your Query on the left-hand side, and the total frequency of occurrences in the text on the right-hand side. When finished, please click *Finish* in order to display the Query results in the Analyzer. For information on how to extend your Query, please see section 3.2.1.2 Complex Queries.

### 3.2.1.1.3 Searching by Tag

Another option is to search by Tag. To do so, please select *by Tag* and click *Next*. (Note: It is necessary that you have already started tagging your text before you can run Queries by Tag.) You now find a list of the Tagsets containing the Tags you already used for tagging your text. Please click the little arrow in front of a Tagset to display the corresponding Tags.

For example:

You want to search for all the parts of the text that you tagged with a Tag named “Exclamation” which is part of the Tagset “Signs of Emotion”. Just click the little arrow in front of the Tagset “Signs of Emotion” and select the Tag “Exclamation”.

As soon as you select one of the Tags, the syntax of your current Query and a preview of your current Query results are displayed in the bottom section of the Query Builder.

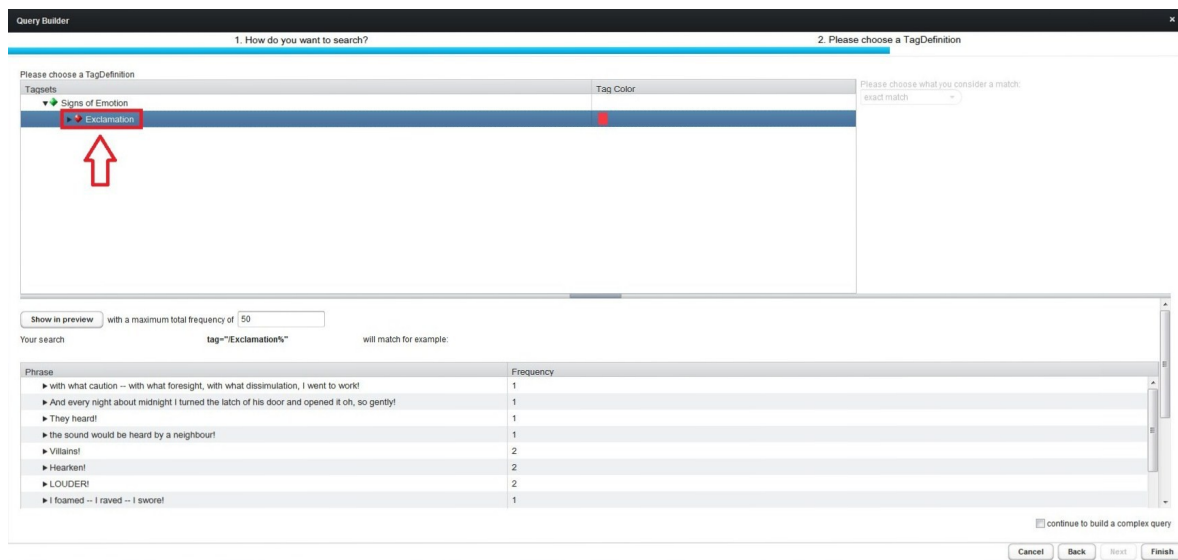


fig. 16

If you would like to change the maximum number of results displayed in the preview, please enter the desired maximum in the bottom section of the Query Builder and click *Show in preview*. If your Query results include different word types that meet your Query,

different words will be displayed in different rows, together with their respective frequency. At the bottom of the preview section you find the total count of word types that meet your Query on the left-hand side, and the total frequency of occurrences in the text on the right-hand side. When finished, click *Finish*. For information on how to extend your Query, please see of section 3.2.1.2 Complex Queries.

When you search by Tag, you can also decide that the search results are to be displayed by markup instead of by phrase. To do this, please select the tab *Result by markup* above the listing of the results. If your Query results include different Tags, every Tag will be displayed in a different row.

### 3.2.1.1.4 Searching by Collocation

Another option is to search by collocation. This means that you can look for occurrences of a certain word with the precondition that it must appear near a specific other word. To do so, please select *by collocation* and click *Next*. Now type into the first field the word of which you would like to search all occurrences that appear next to the word that you type into the second field. In the third field you may enter the span of words within which the two words shall be found.

For example:

You may want to search for all occurrences of the word “louder” that appear within a 20 word span together with the word “heard”. Just type in that you would like to search for all occurrences of “louder” that appear near “heard” and choose 20 as span.

If you click *Show in preview*, the syntax of your current Query and a preview of your current Query results are displayed in the bottom section of the Query Builder.

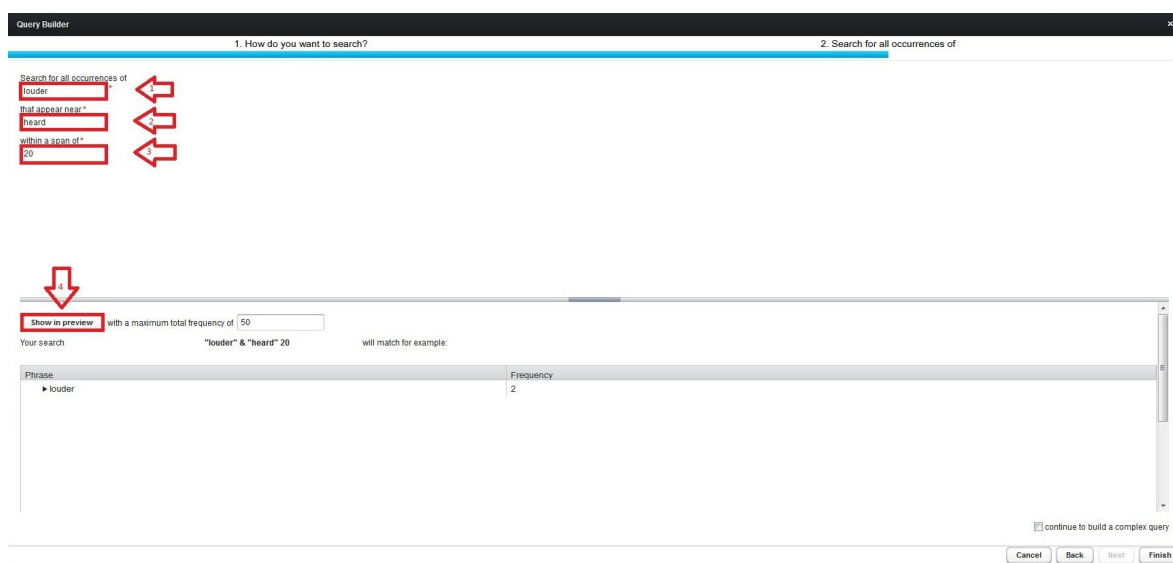


fig. 17

If you would like to change the maximum number of results displayed in the preview, please enter the desired maximum in the bottom section of the Query Builder and click *Show in preview*. If your Query results include different word types that meet your query, different words will be displayed in different rows, together with their respective frequency. At the bottom of the preview section you find the total count of word types that meet your Query on the left-hand side, and the total frequency of occurrences in the text on the right-hand side. When finished, click *Finish*. For information on how to extend your Query, please see section 3.2.1.2 Complex Queries.

### 3.2.1.1.5 Searching by Frequency

Another option is to search by frequency. To do so, please select *by frequency* and click *Next*. You can now 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 lie between the first and the second value you have entered in the two input fields.

For example:

You may want to search for every word that occurs between ten and 20 times in the text. Just select “between” from the drop-down list, type “10” into the first box and “20” into the second one.

If you click *Show in preview*, the syntax of your current Query and a preview of your current Query results are displayed in the bottom section of the Query Builder.

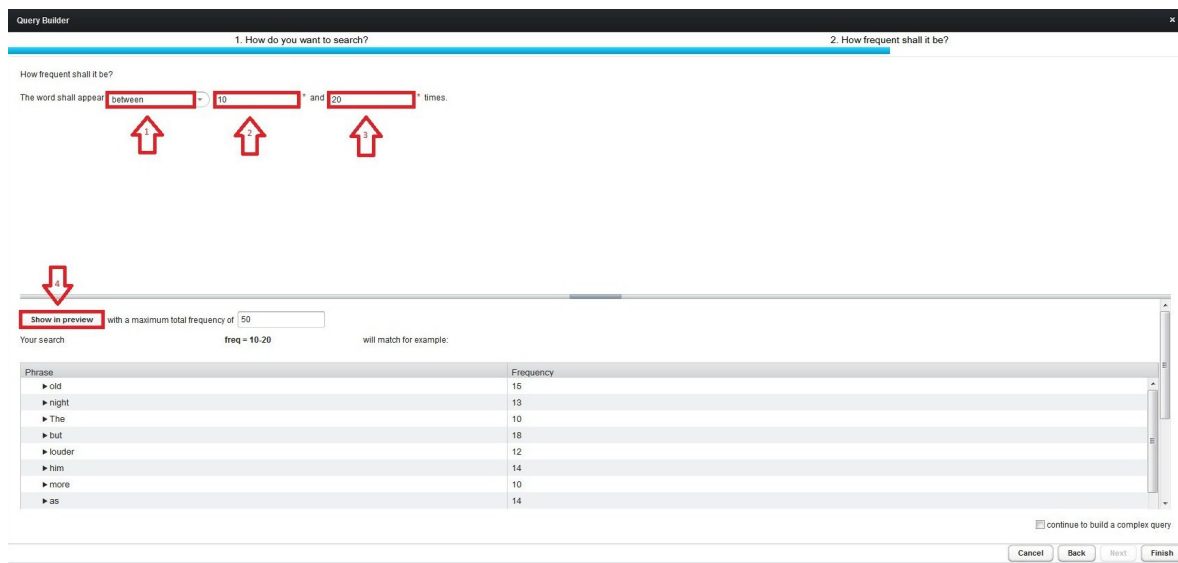


fig. 18

If you would like to change the maximum number of results displayed in the preview,



please enter the desired maximum in the bottom section of the Query Builder and click *Show in preview*. If your Query results include different word types that meet your Query, different words will be displayed in different rows, together with their respective frequency. At the bottom of the preview section you find the total count of word types that meet your Query on the left-hand side, and the total frequency of occurrences in the text on the right-hand side. When finished, click *Finish*. For information on how to extend your Query, please see section 3.2.1.2 Complex Queries.

### 3.2.1.2 Complex Queries

If you want to make your Query more complex, do not click *Finish* yet, but tick off the box *continue to build a complex query* in the bottom right corner of the Query Builder and click *Next*.

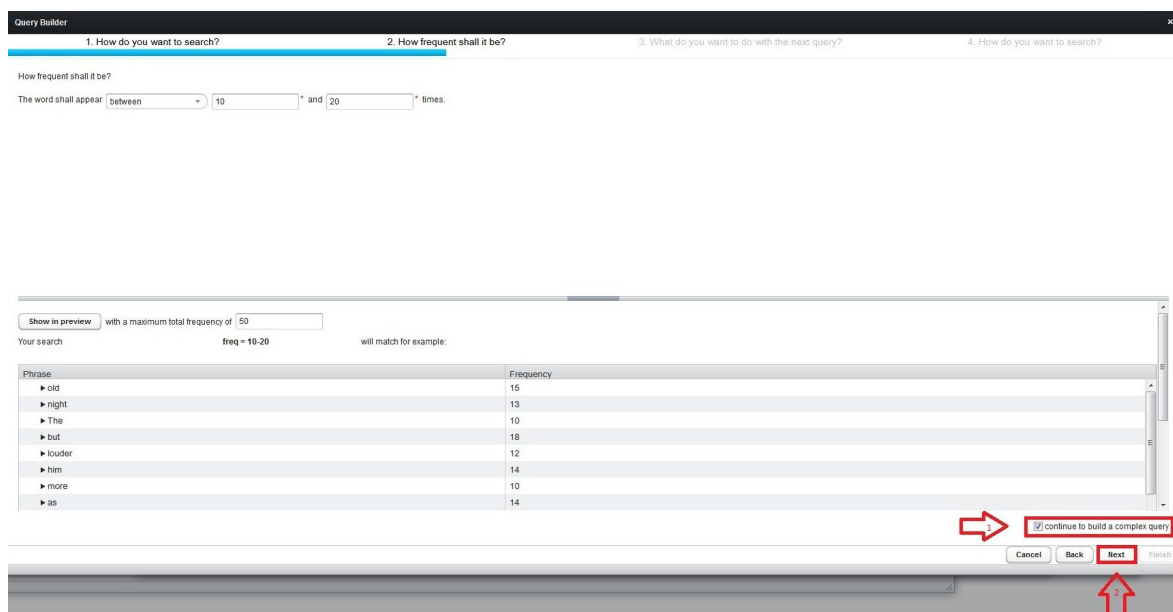


fig. 19

You can now choose the manner in which you would like to extend your Query—by either adding more results, or by excluding hits from the previous results, or by refining the previous results. Please make your choice and click *Next*.

For any kind of expansion you choose, you now have to decide, just like for the first step of the Query building, whether your refinement is specified by word or phrase, by grade of similarity, by Tag, by collocation and by frequency. (For information on these options, please see the previous sections.)

### 3.2.1.2.1 Adding More Results

If you choose to add more results, you are basically combining two Queries.

For example:

You may want to search for every occurrence of the word “eye” and for every word that occurs more than 50 times in the text. Just select *by word or phrase* in the beginning of the Query building and click *Next*. Type in that the word is exactly “eye”, tick off *continue to build complex query* and click *Next*. Now select *add more results* and click *Next*. Select *by frequency* and click *Next*. Choose *greaterThan* from the drop-down list, type “50” into the next field, click *Show in preview* and click *Finish*.

### 3.2.1.2.2 Excluding Hits from Results

If you choose to exclude hits from previous results, you can subtract certain hits from a Query.

For example:

You may want to search for all the words that are 50 percent similar to the word “man”, but only if they do not appear less than 10 times in the text. Just select *by grade of similarity* in the beginning of the Query building and click *Next*. Type in that the word is similar to “man”, adjust the grade of similarity scale to 50, select *continue to build complex query* and click *Next*. Select *exclude hits from previous results* and click *Next*. Select *by frequency* and click *Next*. Choose *lessThan* from the drop down list, type “10” into the next field, click *Show in preview* and click *Finish*.

### 3.2.1.2.3 Refining Results

If you choose to refine the previous results, you are asking CATMA to list the results of your first Query only in case they meet a further requirement.

For example:

You may want to search for all the occurrences of the word “louder” that are also tagged with the Tag “Exclamation”. To do so, please select *by word or phrase* in the beginning of the Query Building and click *Next*. Type in that the word is exactly “louder”, select *continue to build a complex query* and click *Next*. Select *refine previous results* and click *Next*. Select *by Tag* and click *Next*. Click the little arrow in front of the Tagset that contains the Tag “Exclamation”, select the Tag “Exclamation” and click *Finish*.

Please note:

For refinements with Tags, you have the possibility to choose what you consider a match before you finish the Query building.

Please consider the previous example:

If the word “louder” and the Tag “Exclamation” are supposed to be an exact match, please select *exact* from the drop-down list. If the word “louder” is supposed to be entirely contained in the part of the text that is tagged as “Exclamation”, select *boundary*. (Please note: When you choose *boundary*, you allow for “louder” to be part of a larger Tag named “Exclamation”.) If the word “louder” and the Tag “Exclamation” are supposed to have overlapping parts, select *overlap*. (Please note: When you choose overlap, you also allow for only parts of the word “louder” to be part of the Tag “Exclamation”.)

### 3.2.1.2.4 Bracket Particularities of Complex Queries

When you compile complex Queries, the Query Builder places brackets for you in order to regulate in which sequence the information must be processed. Please note that the Query syntax functions like a mathematical formula: The information in brackets is processed first. As it may be the case, the Query Builder does not place brackets in the way you would like it to, because it would change the result of your Query. In this case you have to change the brackets manually. If you, for example, start the Query Builder with the *by word or phrase* option, enter an “s” in the field *The first word starts with*, refine your Query by using the Tag “Exclamation”, select *exclude hits from previous results* and exclude all words with more than five occurrences and add more results by using the *by word or phrase* option again, but this time you enter a “t” in the field *The first word ends with*, then you have compiled this Query:

```
((reg="\b\Qs\E\S*") where (tag="Exclamation")) - (freq > 5) ,  
(reg="\b\S*\Q\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="Exclamation")**), 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\*\Q\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="Exclamation")**), but you do not only want to exclude all words that 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="Exclamation") - (freq > 5 ,  
reg="\b\S*\Q\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="Exclamation"))—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\*\Q\E(?:\W)”)” in brackets.

### 3.2.2 Query Syntax

Although most Queries can be generated by using the Query Builder, you may sometimes want to enter your Query directly. To do so, please enter your Query in the field in the top left corner of the Analyzer and click *Execute Query*.

One tip for learning CATMA's Query Language is to always examine the Queries the Query Builder generates when you use it. By doing this, you will get used to the special expressions and the syntax that are necessary for achieving certain results. However, in the following sections you find a detailed guide on how to compile Queries for different purposes.

#### 3.2.2.1 Simple Queries

##### 3.2.2.1.1 Queries with Words or Phrases

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

For example:

```
"eye"
```

This Query would select all occurrences of the word “eye” from the text that you have loaded. If you use the Wordlist and select “eye” from the list, the results would be identical.

Building Queries with phrases is just as easy. Just enter a phrase and put the whole phrase in inverted commas.

For example:

### "I think it was his eye"

This Query will list all occurrences of the phrase "This is my bicycle". Please note that only exact matches are returned.

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:

"\"Villains!\" I shrieked"

### 3.2.2.1.2 Queries with Regular Expressions

Regular Expressions allow you to search for words as well as for parts of words. Whenever you want to create a Query with a Regular Expression, please start your Query with **reg=**

and put the following Query in inverted commas.

Here is a list of possible operators for Queries with Regular Expressions and their respective functions:

Operator	Description
.	period/dot: represents any single character.  For example: If you search for <b>reg="b..t"</b> , you will search for any sequence of four characters, where the letter "b" is followed by two arbitrary characters and the letter "t". Please note that the dot may also represent whitespace characters. So if you run the example Query, you might get "beat" as well as "by t" (as part of "by the observations") among the results.
*	asterisk: represents zero, one or more consecutive occurrences of the preceding character.  For example: If you type in <b>reg="dis*"</b> you will search for any sequence of the letters "di", followed by the letter "s" zero or more times. So if you run the Query, you may get "di" (as part of the word "did"), "dis" (as part of the word "disease" and "diss" (as part of the word "dissemble") among the results.
?	question mark: represents zero or one consecutive occurrences of the preceding character.

	<p>For example: If you type in <b>reg="dis?"</b>, you may get “di” and “dis” among the results.</p>
+	<p>question mark: represents zero or one consecutive occurrences of the preceding character.</p> <p>For example: If you type in <b>reg="dis+"</b>, you may get “dis” and “diss” among the results.</p>
<b>\b</b>	<p>Represents a word boundary marker.</p> <p>For example: If you type in <b>reg="\bb..t\b"</b>, you may get “beat” among the results, but not “by t”.</p>
[ ]	<p>square brackets: represent a set of characters, i. e. a character class.</p> <p>For example, <b>reg="b[aeiou]"</b> will list every occurrence of the letter “b” where “b” is followed by a vowel, like “be” or “ba” (as part of the word “back”).</p>
^	<p>circumflex accent: is used to negate any character class.</p> <p>For example, <b>reg="b[^aeiou]"</b> will list every occurrence of the letter “b” where “b” is followed by a non-vowel, like “by” or “br” (as part of the word “brain”).</p>
-	<p>hyphen/minus: is used to specify a range in a character class.</p> <p>For example, <b>reg="b[a-m]"</b> will list every occurrence of the letter “b” where “b” is followed by any letter that lies between “a” and “m” in the alphabet, including “a” and “m”, like “be” or “bl” (as part of the word “impossible”).</p>

Some of the operators have a different function when combined with other operators. The description of these functions would exceed the scope of this manual. Also, Queries with Regular Expressions may work rather slowly when applied to large documents. A useful alternative for some of the Queries with Regular Expressions pose Wildcard Queries (see

section 3.2.2.1.3 Queries with Wildcard).

Please note: If you actually want to search for a character that can also be used as an operator in a Query with Regular Expressions, please escape the operator by putting a backslash in front of it.

For example:

If you want to search for any five letter word that ends with “cute” and finishes a sentence, you have to type in **reg=".cute\."**.

### 3.2.2.1.3 Queries with Wildcard

Some of the functions of Queries with Regular Expressions can be achieved by using Wildcard Queries. They allow you to search for words by entering only parts of the word.

Whenever you want to create a Wildcard Query, please start your Query with

**wild=**

and put the following Query in inverted commas.

Here is a list of the operators for Wildcard Queries and their respective functions:

Operator	Function
%	percent: represents an unknown part of a word.  For example: If you type in <b>wild="%ill"</b> , you will receive any word that ends with “ill”, like “will” and “still”.  Please note: This operator may also be used in Queries with Tags. For further information, please see section 3.2.2.1.5 Queries with Tags.
_	underscore: represents any single character within a word.  For example: If you type in <b>wild="_ill"</b> , you will receive any four-letter word that ends with “ill”.

Please note: If you actually want to search for a character that can also be used as an operator in a Wildcard Query, please escape the operator by putting a backslash in front of it.

For example: If you want to search for any word that starts with “aim\_”, please type in **wild="aim\\_%"**.

### 3.2.2.1.4 Queries with Similarity

Sometimes it is useful to look for words which are similar to some other word. If you would like to do this, please put

**simil=**

in front of your Query and put the following Query in inverted commas.

The Analyzer uses the Ratcliff/Obershelp Pattern Recognition Algorithm to determine similarity. (A discussion of the algorithm is beyond the scope of this user manual, but if you are interested, please have a look here: <http://www.drdoobs.com/database/pattern-matching-the-gestalt-approach/184407970?pgno=5> )

Here is an example of how similarity Queries work:

**simil="man" 70%**

This Query selects any word which is to 70 per cent similar to the word “man”. Note that the percentage symbol is optional.

### 3.2.2.1.5 Queries with Tags

If you have tagged the text you are working on, you will be able to use this meta-information in your Queries. To do so, just put

**tag=**

in front of your Query and put the following Query in inverted commas.

Here is an example:

**tag="Exclamation"**

This Query would select any word or phrase that is tagged with the Tag “Exclamation”.

If you would like to search for a specific Subtag, you can type in the whole path that leads to the desired Subtag.

For example:

**tag="Exclamation/Single Word"**

This Query would list every instance of the Subtag “Single Word” that belongs to the Tag “Expression”.

It is also possible to use the percent operator from the Wildcard Queries when searching for Tags.

For example:

**tag="%Single Word"**

This Query would list every instance of a Tag named “Single Word” in the text, regardless



to which Tag the Subtag “Single Word” belongs, for there may be more than one Subtag with this name.

### 3.2.2.1.6 Queries with Collocation

Collocate Queries allow you to search for words that appear near other words. You may, for example, want to find out where the word “louder” appears near word “heard”. This can easily be done with the ampersand operator.

Operator	Description
&	ampersand: is used to define collocate Queries. For example: " <b>louder</b> " & " <b>heard</b> " <b>20</b> will list all occurrences of “louder”, where “heard” appears within a predetermined number of words to either side of “louder”. In this case, “20” is the span size for the collocation. If you omit the span size, the default value is five words to either side.

Please note that collocate Queries are evaluated from right to left, meaning that the leftmost word will always be the one that is selected. If you compile a collocation Query with more than two components, though, you need to place brackets to define which part of the Query is supposed to be processed first.

### 3.2.2.1.7 Queries with Frequency

You can also build Queries that will list you all the words that occur in your text with a certain frequency. Whenever you want to create a Query with frequency, please start your Query with

**freq=**

Here are a few examples of the possibilities for Queries with frequency:

Operator	Description
=	equal sign: helps to search for the exact frequency. For example: <b>freq = 5</b> will list all the words that occur with a frequency of five in the text.
<	smaller than: helps to search for word that occur less often than the chosen frequency in the text.

	<p>For example: <b>freq &lt; 100</b> will list all the words that occur less than 100 times in the text.</p> <p>Please note: This operator may also be combined with the equal sign.</p> <p>For example: <b>freq &lt;= 100</b> will list all the words that occur 100 or less times in the text.</p>
>	<p>larger than: helps to search for words that occur more often than the chosen frequency in the text.</p> <p>For example: <b>freq &gt; 50</b> will list all the words that occur more than 50 times in the text.</p> <p>Please note: This operator may also be combined with the equal sign.</p> <p>For example: <b>freq &gt;= 100</b> will list all the words that occur 50 or more times in the text.</p>
-	<p>hyphen/minus: helps to search for a range of frequencies.</p> <p>For example: <b>freq = 5-10</b> will list all the words that occur between five and ten times in the text.</p>

### 3.2.2.2 Complex Queries

Complex Queries allow you to combine everything you have learned about Queries so far in order to create even more powerful Queries. In the following sections, you will find information on different ways of building complex Queries. For a summary chart of all the operators that can be used in complex Queries, please see section 3.2.2.2.5 Summary Chart of Operators for Complex Queries.

Please note: For any complex Query that consists in more than two components, it is necessary that you place brackets to define which part of the Query is supposed to be processed first.

#### 3.2.2.2.1 Combining Queries

You can combine the results of multiple basic Queries. This can simply be achieved by using the comma

,

operator.

One example would be:

**"louder" , "heard"**

This Query would list all the occurrences of “tree” and all the occurrences of “house”.

A more complex example would be:

**wild="%ed" , freq > 20**

This Query would list any word ending with the letters “ed” as well as any word which occurs more than twenty times.

### **3.2.2.2.2 Queries with Exclusions**

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 are not interested in some of these results. This can be achieved by using the hyphen/minus

-

as an operator.

You might, for example, be interested in any occurrence of the word “louder” that is not tagged with the tag “Exclamation”. The Query you are looking for is the following:

**"louder" - tag="Exclamation"**

This Query will list every occurrence of the word “louder” that is not tagged with the Tag “Exclamation”.

### **3.2.2.2.3 Queries with Adjacency**

Queries with adjacency allow you to search for certain results that will only be listed if they are directly followed by the results of a different Query. These Queries are built by using the semicolon

;

as an operator.

For example:

**"an" ; wild="u%"**

This Query will list every occurrence of “an” that is directly followed by any word starting with “u”.

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

### 3.2.2.2.4 Queries with Refinements

Technically, Queries with exclusions and Queries with adjacency are subclasses of Queries with refinements. But there are many more possibilities of refining a Query, e. g. refining it by similarity, by Tag and by frequency. To build Queries with refinements, you have to use the

**where**  
operator.

Here is an example for a refinement by similarity:

**wild="a%" where simil="man" 50%**

This Query will select all words starting with the letter “a”, which are 50% similar to “man”.

Here is an example for a refinement by frequency:

**wild="\te%" where freq = 100**

This Query allows you to search for all the words that start with “te” that occur exactly 100 times in the text.

Here is an example for a refinement with a Tag:

**"louder" where tag="Exclamation"**

This Query would select all instances of “louder” that are tagged with the Tag named “Exclamation”.

Please note that, for refinement with Tags, you may also specify what you consider a hit. There are three possibilities: exact, boundary and overlap.

Here is an example for an exact match:

**"louder" where tag="Exclamation" exact**

This Query will list all occurrences of “louder” where the exact word “louder” is tagged with the Tag “Exclamation”. It will not list occurrences of “louder” where “louder” is part of a larger Tag named “Exclamation” or where only parts of the word “louder“ are tagged as “Exclamation”.

Here is an example for a boundary match:

**"louder" where tag="Exclamation" boundary**

This Query will list only those occurrences of “louder” that are entirely included by a Tag named “Exclamation”. This is the case when “louder” and the Tag “Exclamation” are exact matches, but also when the whole word “louder” is part of a larger tag named “Exclamation”.

Here is an example for a boundary match:

**"louder" where tag="Exclamation" overlap**

This Query will search for every part of the text where the word “louder” and a part of the text that is tagged as “Exclamation” have an overlap. So the word and the Tag do neither have to be an exact match, nor does the whole word “louder” have to be entirely included in the Tag “Exclamation”.

It is also possible to work with complex refinements by either using the comma

,

as a logical “and” operator or by using the pipe

|

as a logical “or” operator.

Here is an example for the “and” operator:

**"louder" where tag="Exclamation", tag="Capital Letters"**

This Query would select all instances of “louder” that are tagged both as “Exclamation” and as “Capital Letters”.

Here is an example for the “or” operator:

**"louder" where tag="Exclamation" | tag="Anxious"**

This Query would select all instances of “louder” that are tagged as “Exclamation” or as “Anxious”.

In the case of multiple refinements, the comma operator takes precedence over the pipe operator. Use brackets to preserve the meaning of your refinement in cases where the operator precedence would change the intended meaning.

### 3.2.2.2.5 Summary Chart of Operators for Complex Queries

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.

-	hyphen/minus: is used to define exclusions. (Please note: When used in a Query with frequency, the hyphen is used to define a range of frequencies.)
;	semicolon: is used to define adjacency.
<b>where</b>	is used to specify refinements.

### 3.3 Key Word in Context

From every listing you generate in the Analyzer, like Wordlists or Query results, you can proceed by selecting phrases from the list that you would like to be shown in their contexts. To do so, please tick off the boxes behind the words you would like to be listed in their contexts in the *Kwic* (short for Key word in context) column of your list. You can also select or deselect all words from the list at once, using the buttons in the bottom left corner of the Analyzer window. The words you choose for Kwic will be listed in the bottom right section of the Analyzer window, displaying your key word together with the ten words surrounding the key word in the text.

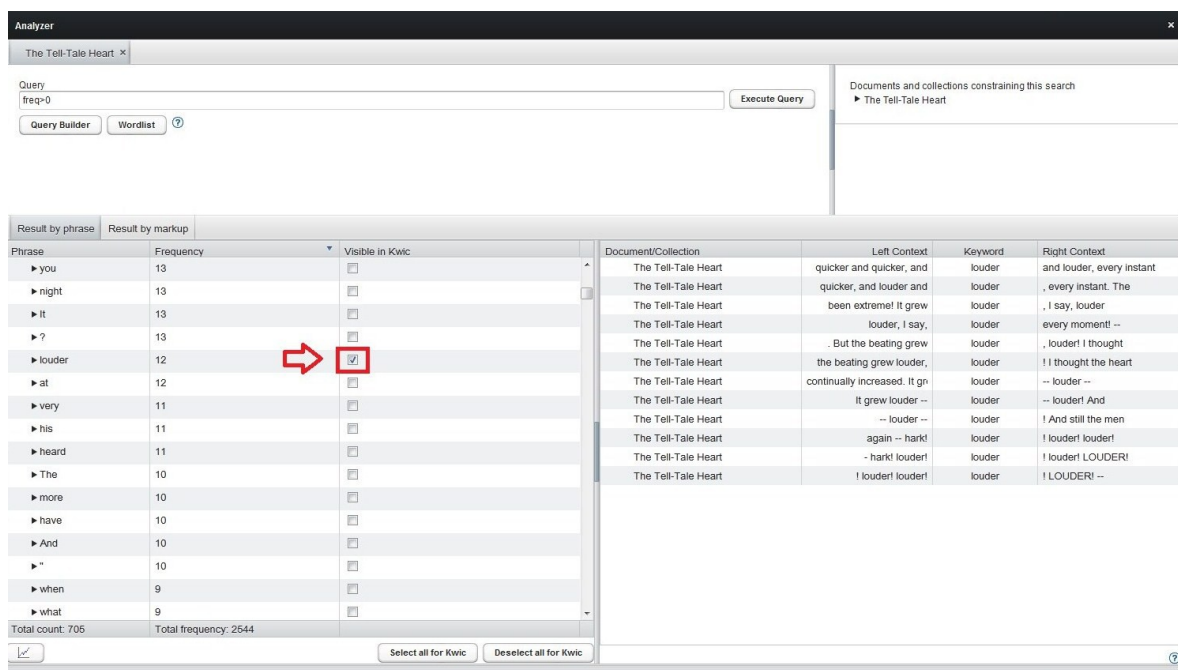


fig. 20

If you would like to view more context than the one displayed in Kwic, double click a key word in context in the bottom right section of the Analyzer window in order to open the Tagger and to jump to the corresponding position in the text.

The Kwic view also offers the possibility to tag words from the Wordlist or Query results. To do so, please select occurrences of a word from the Kwic list via mouseclick, then open

the Tag Manager (if you have not used the Tag Manager yet during your current working session, you need to open it via the *Open Tag Library* button in the Repository Manager) and drag and drop the desired Tag from the Tag Manager to the occurrence of the word, as listed in Kwic, that you want to tag. Confirm the operation in the upcoming pop-up window.

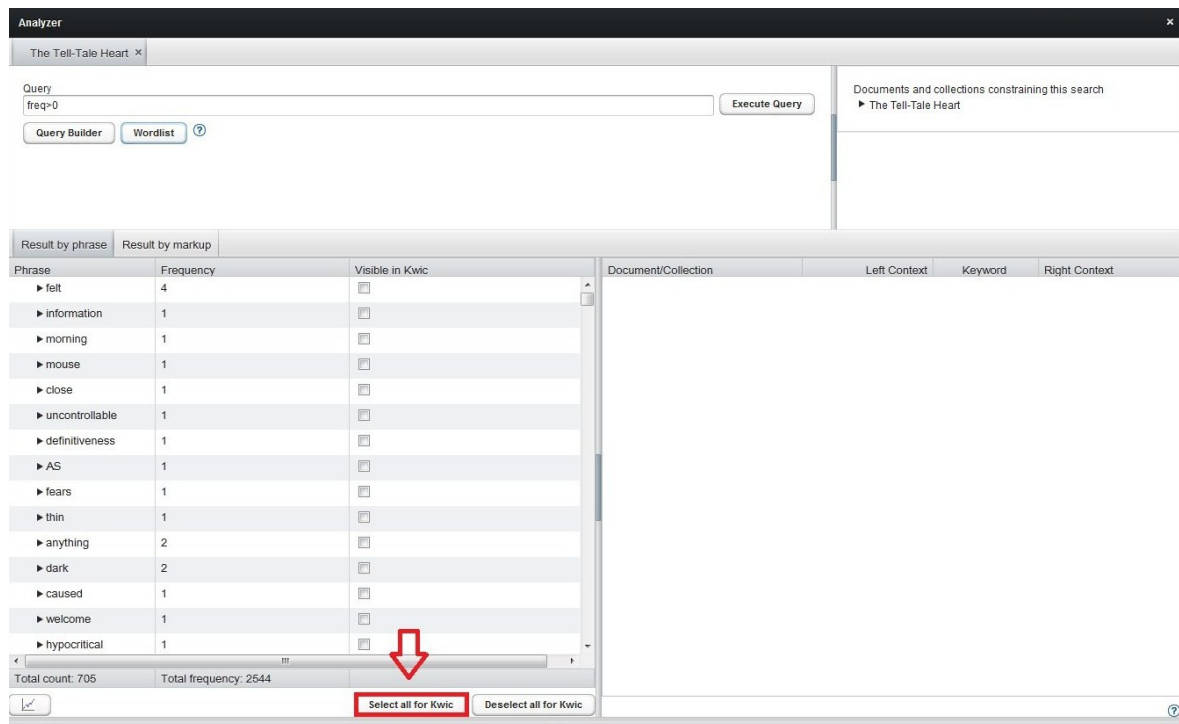


fig. 21

To select more than one occurrence, please hold the Control key of your keyboard after having selected one occurrence and choose further occurrences. To choose all occurrences from Kwic for tagging, please select the first occurrence from the Kwic list, then hold Shift, scroll down to the last occurrence and select it. Now all occurrences in Kwic should be selected. Now drag and drop the desired Tag from the Tag Manager to Kwic and confirm the operation.

You can also untag occurrences via Kwic, but only if you directly search for Tags by using a Query. For information on how to run a Query by Tag, please see section 3.2.1.1.3 Searching by Tag or section 3.2.2.1.5 Queries with Tags. If you have run a Query by Tag, please select *Result by Markup* in the bottom left section of the Analyzer and select all for Kwic. Select the occurrences you wish to untag from the Kwic section and click the *Untag selected Kwics* button.

### 3.4 Analyzing Corpora

Every analyzing operation that can be applied to a single document can also be applied to a whole corpus. Just select the corpus in the Repository Manager, click *More actions...* below the corpora section and select *Analyze Corpus*. When a Wordlist or Query results are listed, you can click the little arrow in front of an entry to display the distribution of the occurrences among the different documents of the analyzed corpus. The analyzed documents are displayed in the top right corner of the Analyzer window.

## 4. Visualizing Results

The Visualizer module offers the possibility of displaying the distribution of words from the Wordlist or Query results in your text. To open the Visualizer, please open the Analyzer and create a Wordlist or run a Query.

### 4.1 Visualization with a Simple Graph

You can now select an entry from the bottom left section of the Analyzer. Now click the little chart symbol in the bottom left corner of the Analyzer window in order to open the Visualizer.

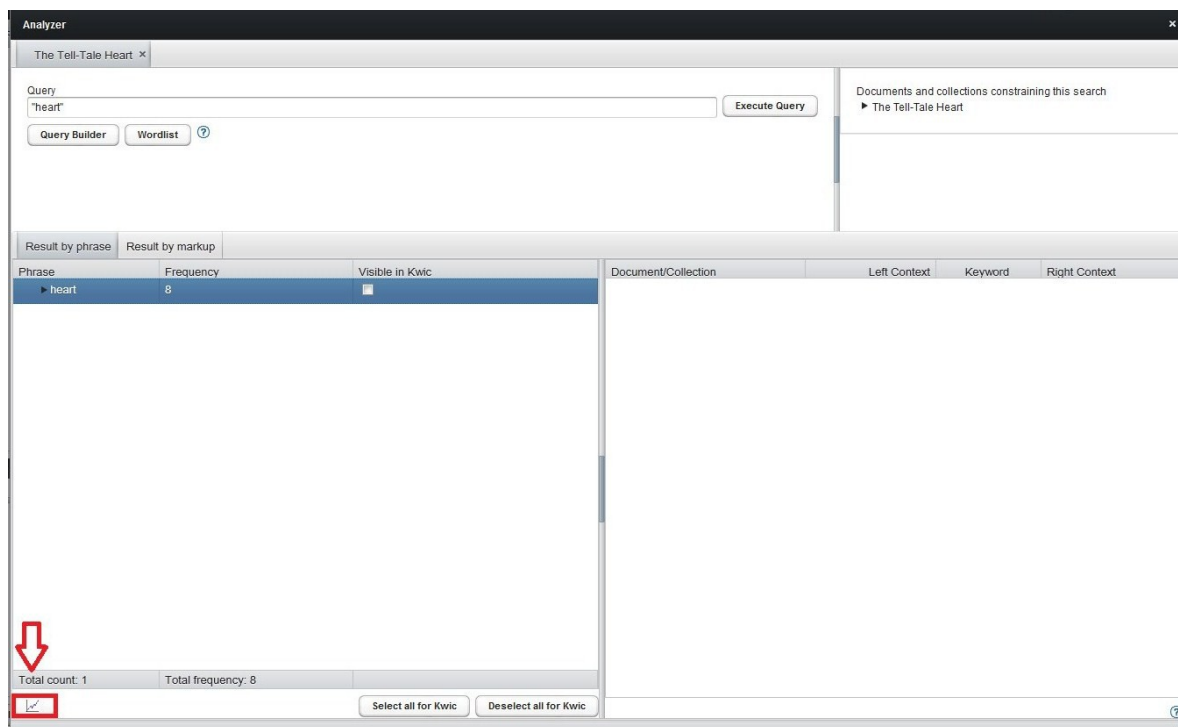


fig. 22

The Visualizer now shows a graph displaying the distribution of the selected results in your



text. The chart shows the number of type occurrence (y-axis) per text section (x-axis). Please note that, for this purpose, the text is divided in ten equal parts—one text section accordingly consists of ten percent of the text. If you navigate the cursor over one of the dots of the graph, the exact number of type occurrences and the text section will be displayed. If you double-click on a dot, you will jump directly to the according text part in the Tagger window.

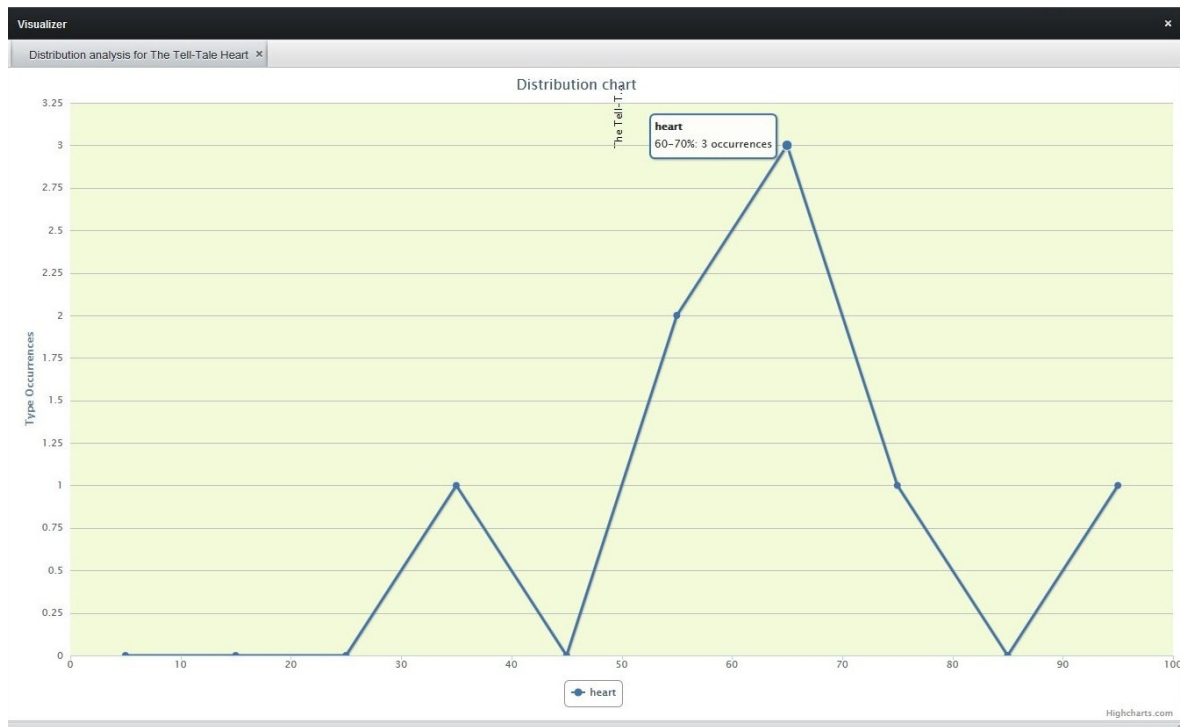


fig. 23

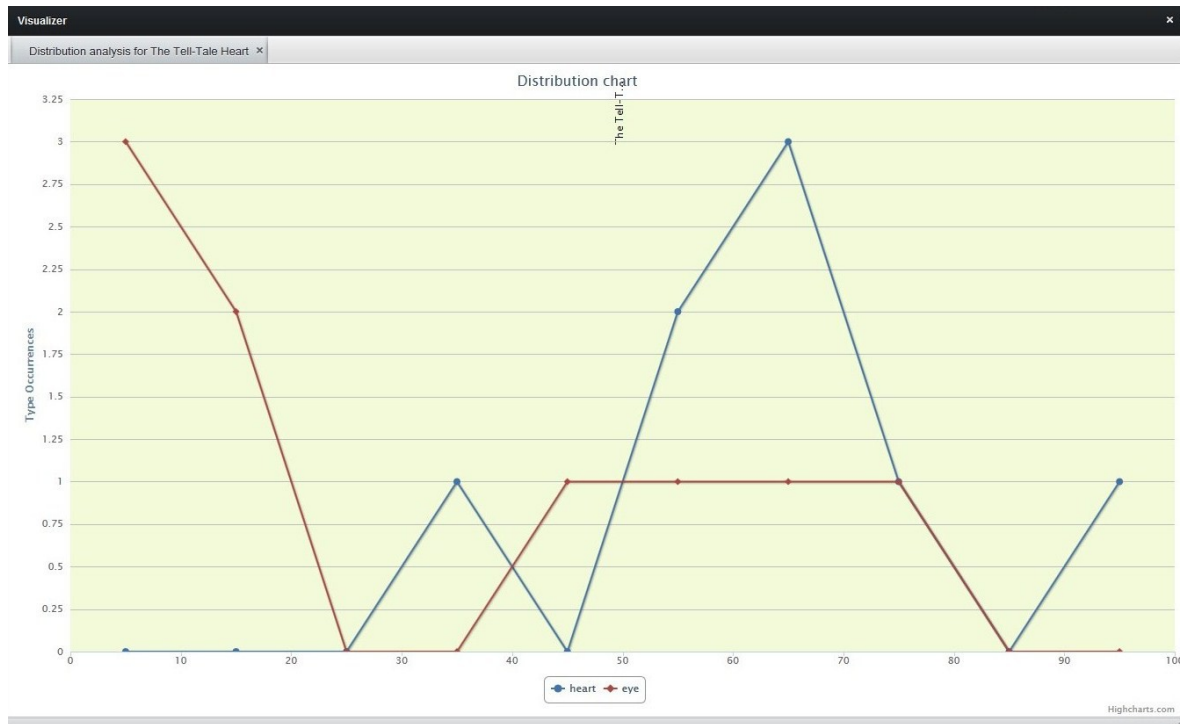
Please note that, technically speaking, the distribution dots should not be connected to form a graph. However, this representation ensures a clear view when more than one graph is shown in the same chart.

## 4.2 Combining Results into One Graph

You may also combine different words or Query results in one graph. To select more than one entry from the bottom left section of the Analyzer window, please hold the Control key of your keyboard after having selected one entry and choose further ones. To choose all results, please select the first occurrence from the list, then hold Shift, scroll down to the last entry and select it. Now all results should be selected. Then click the little chart symbol in the bottom left corner of the Analyzer. The selected entries will now be displayed in one graph in the Visualizer. A list of the entries can be found on the button beneath the chart.

### 4.3 Visualization with Multiple Graphs

It is also possible to show different graphs in one chart. To do so, please generate a distribution graph as described above. Then go back to the Analyzer and generate another graph. Both graphs will now be displayed in the same chart. The graphs can be distinguished by their different colors and dot symbols.



*fig. 24*

Below the chart you find a listing of the graphs and the results they are representing. By clicking the symbol of a graph, you may hide it. This may be useful when very many graphs are shown at once.