CATMA5.0
Tutorial

CATMA
FOR UNDOGMATIC
TEXTUAL MARKUP
AND ANALYSIS

In this tutorial, you will learn how to use CATMA 5.0 – in 5 easy steps:

1. **How to get started** 🐱 **create your own account or use the guest login!**

2. *Manage Resources* 🐱 **get to know your central desktop!**

3. *Manage Tags &* 🐱 **annotate your texts!**
   Annotate

4. *Analyze* 🐱 **browse your texts with queries and more!**

5. *Visualize* 🐱 **get an image of your text(s)!**

Additionally, this tutorial provides you with sample tasks that will help you get used to CATMA's functionalities. Answers and results can be found at the end of the tutorial – but don't look them up too quickly, give it a try first!

CATMA

FOR UNDOGMATIC
TEXTUAL MARKUP
AND ANALYSIS

1. How to get started

You have two options:

      a) Use the guest login

      b) Create your own account

a) Use the guest login

If you don't have a particular project you want to work on with CATMA, you can choose to take a shortcut and use a guest login.

The advantage: Several preparatory steps have already been carried out automatically – CATMA will open with Edgar Allan Poe's *The Tell-Tale Heart*. A Tag Library as well as an Annotation Collection have been created for you as well. What is all that? You'll find out throughout the next sections. The sample tasks in this tutorial require this preloaded material.

The disadvantage of this option is: Your work in a guest account is not going to be saved. You need a personal CATMA account for that.

Click on the CATMA icon as shown below to login as a guest user, answer the question that pops up, and you are ready to go. In this tutorial, you can proceed with section 2.

b) Create your own account

To create your personal CATMA account, go to www.catma.de and click this button:

Click on "Sign in" in the top right corner and choose "Create a CATMA account". If you already have a Google account, you can simply "Log in with your Google account" here. That'll save you the next step.
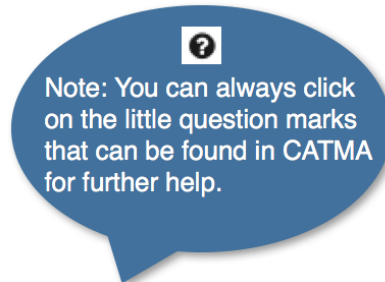
If you don't feel like a robot today, you can confirm that fact in the registration window that opens, and proceed by entering your email address. After clicking "SEND", you'll get an email from CATMA, asking you to confirm your account setup by following the link.

2. *Manage Resources*

In CATMA, there are five modules, which you can use by clicking on the respective tab. You can open several tabs in each module, like you do in your web browser.

| Manage Resources | Manage Tags | Annotate | Analyze | Visualize |

CATMA's Modules

Note: You can always click on the little question marks that can be found in CATMA for further help.

After the login, you will automatically be taken to "*Manage Resources*": your central desktop in CATMA.

Here, you can upload texts and organize them in corpora (on the left side), as well as access the single documents in these corpora (in the middle). You can assign uploaded documents to additional corpora with drag and drop. Under the single documents, the respective Annotations will be displayed in a drop-down menu.

Manage Resources in CATMA

CATMA
FOR UNDOGMATIC
TEXTUAL MARKUP
AND ANALYSIS

Here, you can upload texts and organize them in corpora (on the left side), as well as access the single documents in these corpora (in the middle). You can assign uploaded documents to additional corpora with drag and drop. Under the single documents, the respective Annotations will be displayed in a drop-down menu.

The right section of "*Manage Resources*" displays more detailed information about what's selected, for example corpora or documents. In the lower section, you can organize your Tag Libraries, once you have created them.

What are *Annotations?*

When you annotate a text in CATMA using tags, this markup will be saved in an individual collection called *Annotations*. You can have different types of annotations for one text, or you might annotate a text in a team where everybody creates their own version of the markup. *Annotations* is where all annotations pertaining to one text are stored.

You will learn more about how to create and access Annotations in the next section of this tutorial.

CATMA

FOR UNDOGMATIC
TEXTUAL MARKUP
AND ANALYSIS

What is a *Tag Library*?

Let's start the other way around: *Tags* are the descriptive labels which you can assign to any piece of text loaded into CATMA. Tags are grouped in *Tagsets* – for example, for *The Tell-Tale Heart*, we pre-loaded two Tagsets: *character* and *narrator*. Every Tagset is part of a *Tag Library,* which can contain as many tags and Tagsets as you need.
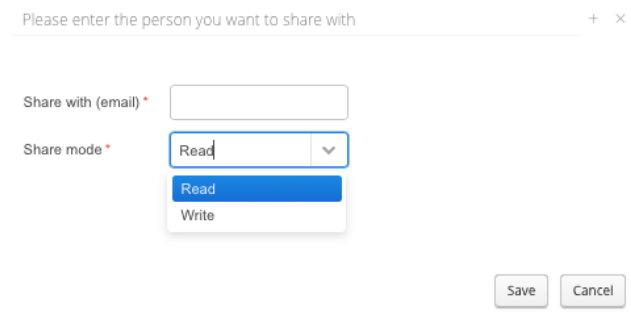

Sample Tag Library

Teamwork and CATMA's "share" function

CATMA supports teamwork. If you collaboratively work on a corpus, a document, a Tag Library, or even Annotations, you will want to make use of the "share" function. Start from *Manage Resources*, highlight the corpus, document, Annotations under a document, or Tag Library which you want to share, and choose "More actions" under the respective column. Here, you can select to share either a corpus, a document, Annotations, or a Tag Library.

The following window will pop up.

Here, you have to enter the email address of the person you want to share things with. Make sure you use the address the other one used to register on CATMA. You also have to choose whether you want the other person to be able to only read what you share with her, or whether you want to give the permission to write in the corpus, document, Annotations, or Tag Library as well. Note that whenever someone makes changes to a write-shared instance, these changes will affect your own version as well. If you want to avoid this, but still be able to work on the same texts or Tagsets, you will have to use a workaround: Export the instance as a file to your computer and ask your partner to re-upload it to CATMA. You'll hence be able to work with two versions of the same corpus, document, Annotations, or Tag Library, without affecting each other.

Note, too, that everything that you add to an already shared corpus will be shared automatically with the other team members. If a corpus is shared in the "Write" mode, everything else (i.e. documents and Annotations) is writable as well. If a corpus is shared in the "Read" mode, the other team members can see the documents and your Annotations, but cannot write in them. The other team members can also create Annotations on read-shared corpora, but these Annotations will be private and you won't be able to see them.

2. *Manage Tags* & *Annotate*

After choosing a single document in *Manage Resources*, you can open it (e.g. with a double click) and *Annotate* will display the text on the left. Active Tagsets and Annotations are shown on the right.

CATMA
FOR UNDOGMATIC
TEXTUAL MARKUP
AND ANALYSIS

the *document*
you're
working with

the set of tags
available to
annotate the
text

the individual tag
that you chose to
annotate a specific
piece of text – your
markup is stored in
Annotations

Annotate in **CATMA**

CATMA 5.0 Tutorial – page 14

CATMA
FOR UNDOGMATIC
TEXTUAL MARKUP
AND ANALYSIS

Tag Creation and *Manage Tags*

(You can skip this step if you are using the guest login.)

In CATMA, tags are managed by grouping them into Tagsets, which in turn are arranged in Tag Libraries. This hierarchical structure guarantees that you can extend and modify your tags at any time – in other words, you don't have to worry about some missing tags when you start; you can always add them later. Additionally, the structure enables you to share selected Tagsets with other CATMA users and/or to import theirs.

Here we show you how to create Tag Libraries, Tagsets, and tags from within *Annotate* – 'on the fly'. You can as well execute these steps in *Manage Tags* itself. *Manage Tags* allows for a better overview of already created Tagsets.

In the *Annotate* window, click on "Open Tagset." First create your own Tag Library (e.g., "My Tag Library") – this is a virtual folder where you collect all your (reusable) Tagsets.

- Within that Tag Library, now define a first Tagset. Tagsets allow you to group a set of related annotation labels (for example, "Characters," "Geographic Entities," "Themes").

- Back in *Annotate*, click on your Tagset and create the actual tags that you require (e.g.: "Character name," "Behavior," "Land," "Waters," "Nature," "Darkness," etc.) by choosing "Create Tag" on the right. Make sure the colors of the tags can be clearly distinguished from each other.

Annotate your text

To annotate your text with the defined tags and save your annotations in an Annotation collection, execute the following steps:

- In *Annotate*, you go to the "Active Annotations" tab and click "Open Annotations." In the pop-up box, click on "Create Annotation Collection." This is where all your annotations relating to this particular source text will be stored.

🐶 Back in *Annotate*, go to the "Active Tagsets" tab and click "Open Tagset." Select your Tag Library and the desired Tagset. Then click "Load Tagset into currently active document" – you can now use all tags within the Tagset to annotate your text.

🐶 In *Annotate*, close-read your source text and pay attention to the phenomena for which you have created your tags. If you identify a phenomenon, select the corresponding text string and click on the color next to the tag in the "Active Tagsets" tab (you may need to unfold the respective Tagset by clicking on the arrow symbol next to it). You've now officially set your first annotation, congrats!

Sample Task for *The Tell-Tale Heart*

1. With the "narrator" Tagset of the Tag Library we already created for you in the guest login (i.e. "attitude" and "style" tags), annotate the relevant parts in the text. If you are working in a team, you can split the text or the Tagset among you to save time.

CATMA

FOR UNDOGMATIC
TEXTUAL MARKUP
AND ANALYSIS

Let CATMA annotate your texts automatically

CATMA can automatically annotate a bunch of phenomena in German texts: Parts of Speech, Tenses, and Temporal Signals. The categories that can be annotated automatically are to be increased in the future. You can let CATMA annotate whole corpora by choosing "More actions" in the corpora column of *Manage Resources*, and then "Generate annotations."

However, be careful to use this functionality in workshop contexts – if many users choose this option at once, we might run out of resources. We are working on improving CATMA's performance continually.
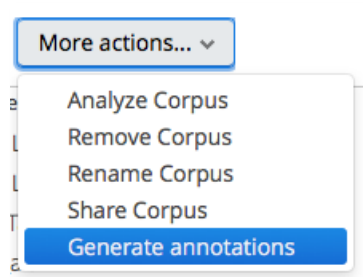
3. *Analyze*

Whenever we engage with texts in order to explore and discuss their meaning, two operations go hand in hand: analysis and interpretation. Their mutual dependency results in what is known as the *hermeneutic circle*.
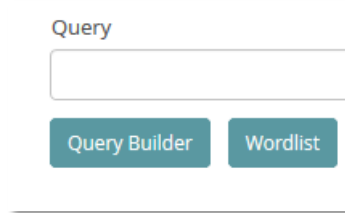
CATMA supports this approach by combining text analysis and text annotation in a cyclical, iterative workflow. Once you have uploaded your digital text to CATMA, you will typically combine your work in *Annotate*, as described above, with using the merits of the *Analyze* module. Whereas you pursue a close reading in *Annotate*, *Analyze* supports a more distant reading approach. Here you can browse your texts without reading them yourself.

You reach *Analyze* through *Annotate* by clicking on the "Analyze Document" button below the displayed source text. Whenever you access *Analyze* through an active document, a new tab is created for each new document. Likewise, different analyses of a single document open new tabs (numbered serially), so you don't lose your previous results. You can also open a new tab by clicking the "+" button in the upper right corner.

In *Analyze,* you see two buttons in the top left section:

Create a Wordlist

CATMA automatically creates a Wordlist of the respective text if you click on the "Word-list" button. The Wordlist can be sorted in ascending or descending frequency by click-ing on the "Frequency" field once or twice.

Sample Task for *The Tell-Tale Heart*

2. What is the most common content word (= a word with "more" semantic meaning than function words such as articles, pronouns etc.)?

Kwic visualization (Kwic = keyword in context)

Select a word in the Wordlist that you are particularly interested in (e.g. "heart" in our sample). Tick the "Visible in Kwic" box for this word and look at the Kwic display on the right-hand side of the screen. If necessary, change the scope of the displayed context using the slider at the bottom (setting to 1–30 tokens possible).

If you double-click on one of the keywords in the Kwic visualization, CATMA will jump to the selected keyword's full-text location back in *Annotate*.

Query Builder

You are free to enter your query directly into the Query box, but a large part of all possible queries can be generated with the Query Builder, which is much more convenient, especially if you aren't familiar with the query language. (By the way: After executing a query with the Query Builder, you can see the respective query phrase in the query mask and thus learn it casually.)

You use the Query Builder by clicking on the corresponding button. In the opening window, you decide what you want to search for:

Choose an instance and click "Next." If you want to search for the word "heart," for example, use the "by word or phrase" query option. Note that queries are case sensitive.

Click "Show in preview" to test whether your query generates a result. If it does, click "Finish" at the bottom right. Your query is now being executed.

You can also use the Query Builder to create complex queries (i.e. to refine, combine, and exclude search results). Just tick "continue to build a complex query" in the query mask to do so.

Sample Tasks for *The Tell-Tale Heart*

3. How many words occur more than five times? Use the query function "by frequency" for this.

4. How many words occur between five and fifteen times? Which physical senses seem to be predominant in the story?

5. How many words have a 70% similarity to "heart?" Use the "by grade of similarity" query option. Open a new *Analyze* tab (by clicking the "+ New Query" button on the upper right) and increase the similarity to 80%. Do it again with 85%.

6. How often does the word "mad" appear near the word "me" (within a span of ten words). Use the "collocation" function.

Analyze your annotations

Analyze also offers you to analyze your own annotations that you set in *Annotate* before. Use the Query Builder to search for a specific tag ("by Tag"), or directly type the query: tag = "tag name". The analysis of your annotated text(s) depends on your research interest and your annotations.

After searching specific annotations, you can examine the "Result by Phrase" (as you did in the other searches) or as well the "Result by Tag" (tabs in the wordlist).

Sample Task for *The Tell-Tale Heart*

7. Search for the different "narrator" annotations you set in task 1. Can you tell something interesting about their frequency or co-occurrences? What can you infer from your annotations concerning the reliability of the narrator?

Annotate your text(s) in *Analyze*

Since the hermeneutic workflow in CATMA is circular, *Analyze* can help you annotate your text, too. Repetitive manual annotation of the same word across an entire text, for example, can be avoided by running a query first and then annotating all search results at once. Execute the following steps for this:

- Run a query as described above.

- Activate the Kwic display and highlight the relevant Kwic lines.

- Click on "Annotate selected Results" in the bottom-right corner: A box displaying your Tagset and its tags will open.
- Select and draw the relevant tag onto the Kwic lines (drag & drop).

All the selected search results have now been annotated with the same tag.

Sample Tasks for *The Tell-Tale Heart*

8. Select all occurrences of "I," "me," "my," "My," "myself," and "mine" to be visible in Kwic and let *Analyze* annotate them with the "hero(ine)" tag. How many annotations does it set?

9. Do the equivalent thing with the "opponent" tag. Which words or phrases qualify for being annotated as "opponent"?

Analyzing text corpora

If you are working with more than one source text, you should access *Analyze* directly from *Manage Resources* by analyzing a selected corpus:
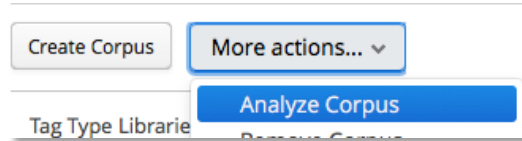
Analyze opens a new tab, named after your corpus. On the right side, you can as well see a list of the documents in the corpus. Now all actions that you execute in *Analyze* will relate to the whole corpus of texts. Be aware that creating a wordlist or executing queries on more than a single text takes more time, depending on the size of the respective corpus.

General considerations before you start tagging and analyzing your texts

CATMA's digital infrastructure forces you to clarify your research problems. You really need to make sure what you want to find out about the source text(s) you are investigating, and thus reflect on your own work, your research questions, and your methods.

Take some time to identify the questions and problems that you would like to investigate upfront. For example, you might be interested in analyzing the presence of the characters in the text, their behavior, their character traits or the like. If you are more interested in geographical features, you could analyze types of geographic entities – countries, cities, waters, islands, but also special places, etc. If you are interested in topics, identify relevant topics, and so on.

If you look at your list of questions, they usually fall into three different categories:

a. Questions that can be answered directly by investigating the actual words and phrases that constitute the text. For example: Which character in *The Tell-Tale Heart* is mentioned *verbatim* more often – "I" or "the man?" A wordlist analysis with *Analyze* already answers this one: "I" is mentioned more often than "the man."

b. Questions that require you to pre-interpret textual elements in order to find out what they mean. For example: Characters in *The Tell-Tale Heart* are not identified by names, but by pronouns like "he" or function words like "officers." But does every "he" point to the old man or maybe to one of the officers? You can make those instances of "he" that refer to the old man explicit by annotating them with CATMA tags.

c. Questions that require you to combine both strategies – direct textual analysis and pre-analytical text interpretation. So, if you really want to find out every mentioning of the narrator in the story, you can use *Analyze* to do a combined search for literal occurrences of "I" plus a search for all occurrences of "me" etc. that you annotated as an indirect reference to the narrator.

Type (b) and (c) questions are the ones where tags come into play – that is, interpretive annotation labels that you attach to a word in order to make its implicit meaning or function explicit to CATMA.

4. *Visualize*

Next to annotating and analyzing your text(s), you will want to get an image of them – literally. You either want to get a quick overview of phenomena in your text or corpus or you want to visually display results of your work in *Annotate* and *Analyze*. This is where *Visualize* comes into play. Currently, CATMA offers two different forms of text visualization: a DoubleTree and a Distribution Graph. There will be more visualization functions in the future.

DoubleTree visualization

The DoubleTree is a very compact and more interactive way to visualize your Keywords in context (Kwic). Choose a word you are interested in from the wordlist, highlight it, and click on the button below it. You can expand the DoubleTree to both sides.
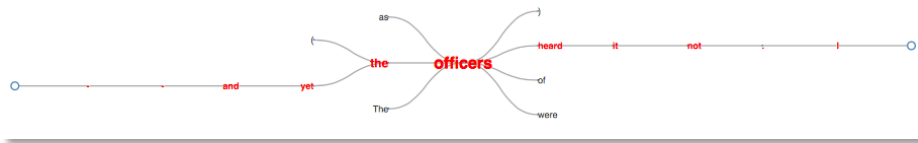
**DoubleTree of "officers"**

Sample Tasks for *The Tell-Tale Heart*

10. Start from the wordlist in *Analyze*. Look for the word "man" in the list, highlight it and click on the <u>DoubleTree</u> button. Click on the words in the DoubleTree visualization and try to figure out how the DoubleTree works. Can you tell something about the character?

11. Try the same with "I."

Distribution Graph

The Distribution Graph allows you to visualize and compare the occurrence of single or multiple phenomena and their distribution and potential development in your text – or even to compare these phenomena throughout several texts in a corpus (in this case multiple diagrams will be displayed next to each other in *Visualize*). You can either combine several words or tags in one graph or show one graph for each word/tag in a diagram. To combine words or tags into one graph, select a group of words or tags that will be interesting to you in the word or tag list of *Analyze*. This is done by selecting the appropriate lines (by holding the command key). If you now click the distribution graph button ( ), the words will be displayed in a single distribution graph. You can also click on the graphs to get back to the respective text passages in *Annotate*.
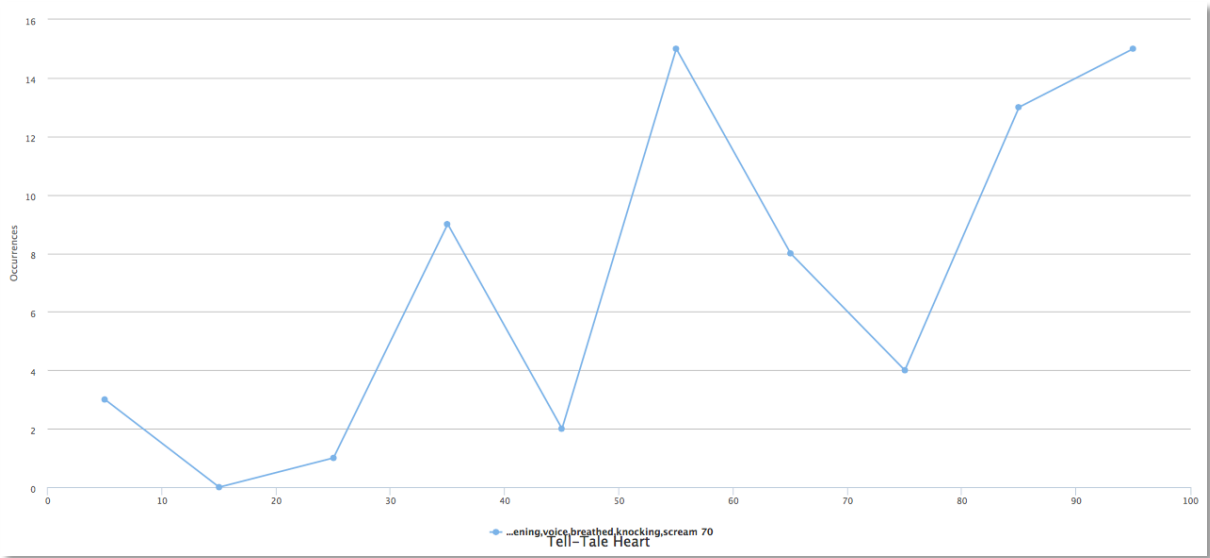
Distribution Graph for sounds and hearing in *The Tell-Tale Heart*

Sample Tasks for *The Tell-Tale Heart*

12. Build a complex query and search for the "hero(ine)" tag as well as the "opponent" tag in *Analyze*. (This will only work if you executed tasks 8 and 9, or set the tags manually.) Highlight the "hero(ine)" tag in the Tag list and click on the distribution button below. What can you tell from the diagram?

13. Go back to *Analyze* and select the tag "opponent." If you now click the distribution graph button again, it will be displayed in the same diagram. Is there something interesting that can be seen from the representation of the two tags?

ANSWERS for the sample tasks

Question 2:

The word "was" appears 33 times (or, in case you don't consider pronouns function words, "I" appears 124 times).

Question 3:

There are 79 words that occur more than five times.

Question 4:

There are 73 words that occur between five and 15 times. Especially sound but also vision are the predominant physical senses.

Question 5:

Eleven words have a 70% similarity to "heart," eight words with 80%. With 85%, only "hear," "HEART," "tear," "heartly," "earth," and "heart" are listed.

Question 6:

The words "mad" and "me" co-occur three times within a span of ten words.
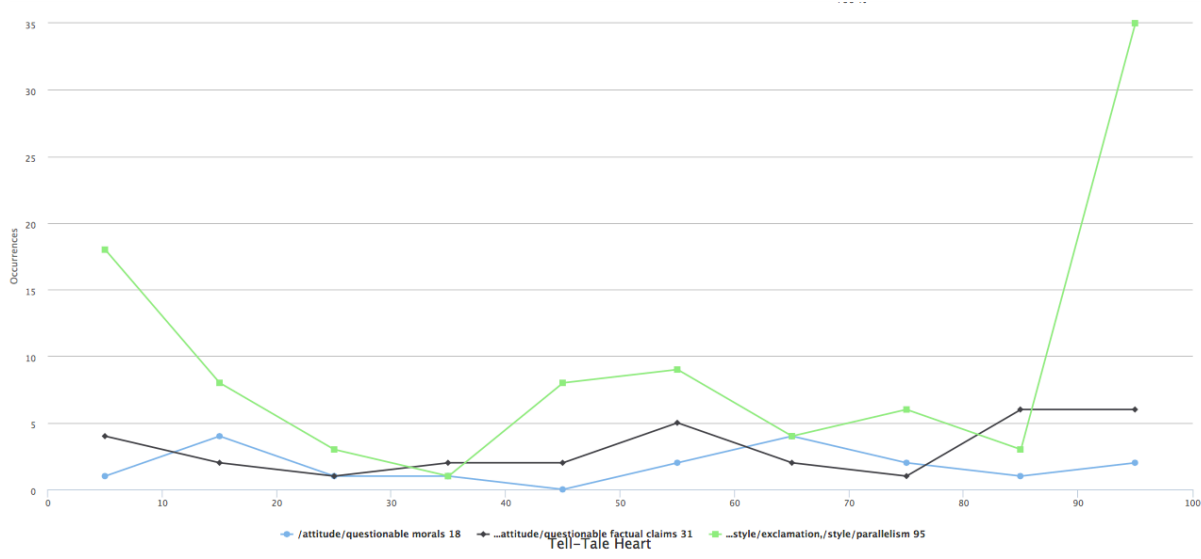

Question 7:

The first distribution chart shows the frequencies of the narrator's attitude (which is "questionable morals" and "questionable factual claims") throughout *The Tell-Tale Heart*. The most striking thing might be the increasing amount of questionable factual claims at the end of the story, while the narrator's morals seem more questionable at the beginning and the beginning of the last third of the story.
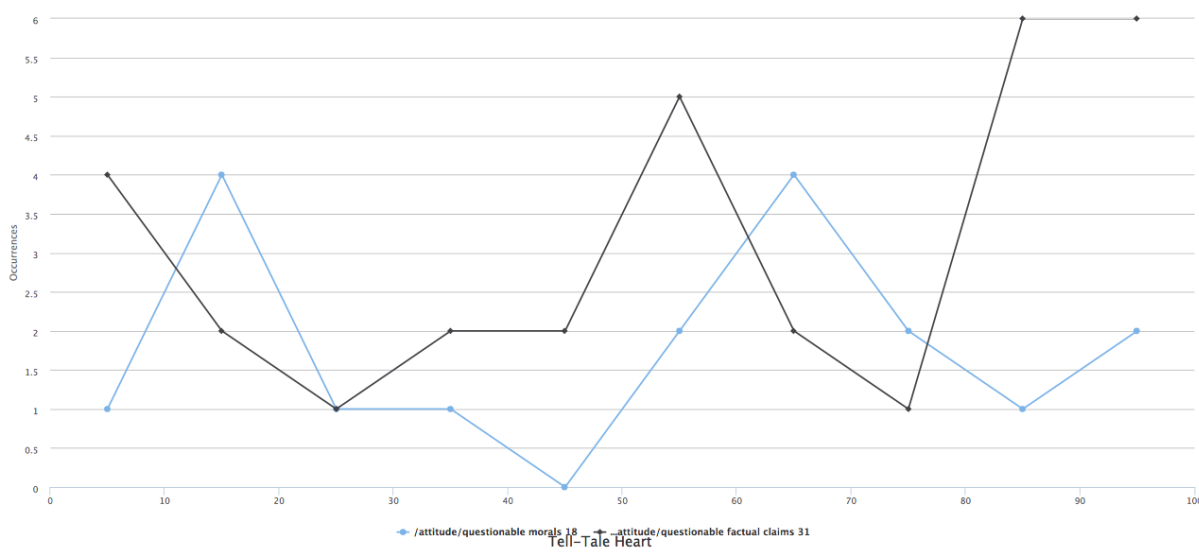
Chart 1: narrator's attitude in *The Tell-Tale Heart*

Chart no. 2 displays co-occurences of these attitude tags with annotations of the narrator's style (which is addresses to the reader, exclamations, parallelisms, and repetitions). One could observe that stylistic devices and a questionable attitude tend to peak at the same points of the story. An augmented emotional involvement of the narrator could be inferred by this, and also a trace of his unreliability on the level of form. However, you should note that especially the attitude tags tend to be set interpretively rather than in a way that generates intersubjectively reproducible results.

Chart 2: narrator's attitude and style in *The Tell-Tale Heart*

Question 8:

There are 178 hero(ine) annotations set.

Question 9:

Words/phrases like "old man," "eye," "heart," "corpse," "he," "him,", "his," "body," "police," "officers," "they," "them," "their," etc. qualify to be annotated with the "opponent" tag.

Question 10:

The DoubleTree allows you to explore a word in its various contexts. For example, the man is obviously old, scared and he dies. This may lead to conclusions about the character of the man as well as his feelings to and/or his behavior towards the narrator.

Question 11:

In the case of "I," a clear picture cannot be seen at first glance – there are significantly more and clearly more variable results.

Question 12:

The hero is predominant throughout the story. Since the hero is the homodiegetic narrator of the story, too, this doesn't come as a surprise. Interesting might be the low point in the middle of the text. By clicking on this point, we find out that this is exactly the moment shortly before the killing, a moment where the narrating I feels overpowered by the "evil eye" of the old man.

Question 13:

The distribution graph makes clear where there are two important moments of decision in the story: when the narrator decides to kill the old man and when he meets the police officers. In these moments, his "opponents" overpower or nearly overpower his presence in the text. Further hypotheses of interpretation could be built on this.
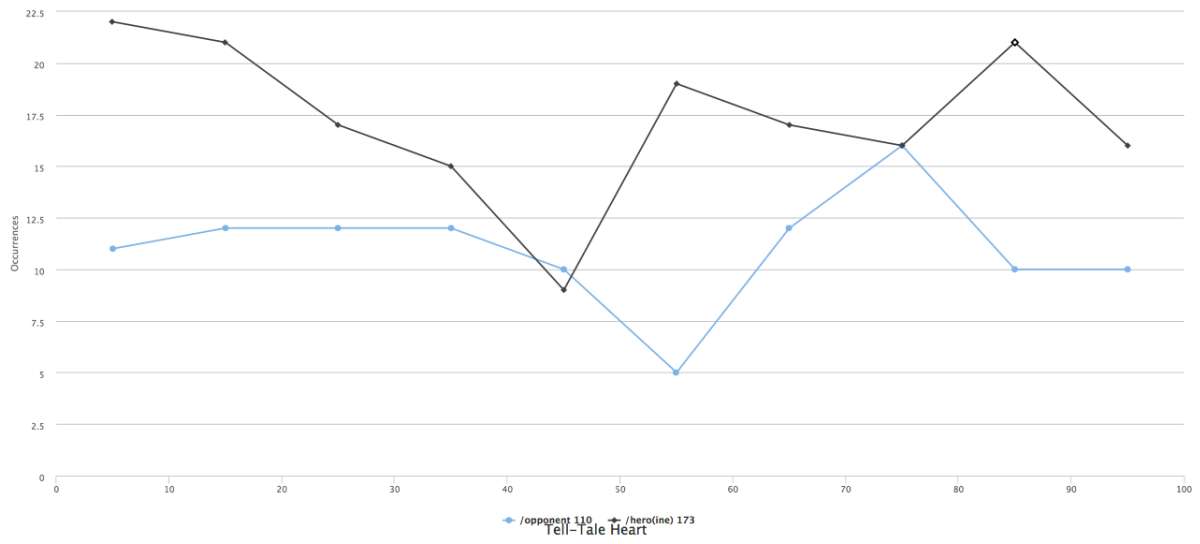
Distribution Graph for "hero(ine)" and "opponent" tags in *The Tell-Tale Heart*

CATMA 5.0 Tutorial – page 42