CATMA – Voyeur collaboration Hamilton 01.-15. March final notes
Authors:
Marco Petris
Lena Schüch

Participants:
Prof. Stéfan Sinclair (McMaster University, Hamilton, Canada)
Andrew MacDonald (McMaster University, Hamilton, Canada)
Lena Schüch (Hamburg University, Hamburg, Germany)
Marco Petris (Hamburg University, Hamburg, Germany)
Date: 31 March 2011

Content

Goals

CATMA
- export of CATMA markup data as input for Voyeur
- conceptualize and implement an appropriate data format to enable Voyeur visualization of data tagged with the space tagset.

CATMA side design and implementation has been carried out by Marco Petris and Lena Schüch.

Voyeur
- import of CATMA data for visualization
- use standard visualization of voyeurtools.org to visualize CATMA data like tags and their instances.
- conceptualize and implement a visualization for data tagged with the space tagset.

Voyeur side design and implementation has been carried out by Stéfan Sinclair and Andrew MacDonald.

What does this mean to and how does this help someone studying literary texts?

Author: Lena Schüch

In a test version, it is now possible to import texts that have been tagged in CATMA into Voyeur. In this way, one can use the visualization tools of Voyeur to display the distribution of tags and collocations of tags (the latter has not been possible in either CATMA nor Voyeur before). At the moment, it is only possible to import texts that do not have tags that overlap, but in the future it should be possible to also import documents with an overlapping tag structure. Especially the possibility to display the collocations of tags has been a long-awaited feature that helps to find re-occurrent patterns in the text, e.g. dependencies between certain elements that indicate the underlying mechanisms of a phenomenon like the literary representation of space.

Furthermore, one can use tags to create a sunburst visualization, i.e. a visualization that arranges the occurrences of tags in a concentric structure (a center and several rings around it). Since the user can decide before generation according to which criterion he wants to arrange the tags by assigning numbered properties to the respective tags, the visualization is quite flexible and usable for various purposes. It can be used for the illustration of different semantic, structural, perspective, temporal, narrative etc. layers of a text that evolve around a main aspect or core. For example, for the representation of narrative layers, one could tag the different levels of a narrative and could determine that aspects of the embedded story should appear in the first ring, aspects of the different frame narratives could be arranged in the rings around the embedded story (similar to a matryoshka).

In the example, sunburst is used for the illustration of (literary) space: The character is the center, and around it, there are different spatial dimensions that are arranged according to the closeness to the character and the perceived factuality of these spatial objects. In this way, one can, among other things, determine which spatial objects are very close to or far away from the character, and which have moved. Since one can compare how many occurrences appear in the different rings, one can determine if the narrator or character rather focuses on spatial objects that are close or far away (either in a spatial, temporal or hypothetical/counterfactual sense). These kinds of information help to find out about, describe and interpret the representation of space in literature.

An example: space visualization

Author: Lena Schüch

Tagset „Space"

	<place>	A place (room, landscape, house, city, country, etc.) mentioned in the text
	<direction>	A direction (preposition, local adverb, adjective) mentioned in the text
	<movement>	A movement of the acting character (like go, walk, stop)
	<deix-space-fig>	Deictic expressions that refer to the acting character as deictic center (s. also perspective tagset)
	<deix-space-nar>	Deictic expressions that refer to the narrator as deictic center (s. also perspective tagset)
	<direct-frame>	A space in which the acting character is located
	<secondary-frame>	A space which is close to the acting character (in assessable distance)
	<inac-frame>	Inaccessible frame: Spatial objects that are in the same spatial-temporal continuum like the acting character but that cannot be entered by him/her
	<spatio-tempo-dis-frame>	Spatio-temporal-distant-frame: Spatial objects that cannot be entered for reasons of spatio-temporal-distance
	<general-space>	General spatial expressions (like "in the world")
	<counterfactual-frame>	Spatial objects that do not refer to frames that exist in the narrated world (dreams, imaginations, etc.)
	<hypothetical-frame>	Spatial objects that are marked as hypothetical
	<space-nar>	Spatial objects that are perceived by the narrator, not by the character (can be used for all frame tags, "neutralizes" the character-bound status of the frame tags)
	<space-ref>	Spatial objects that refer to an extra-textual reality

Possible distribution display of frame tags

Similar to Ronen's overview[1] of spatial frames, there could be a distribution display that arranges the spatial frame tags in a spatial manner: The acting character would be the center, around him there would be the different layers that are marked by the spatial frame tags: The first ring around him would be filled with the occurrences of the direct-frame tag, the second with those of the secondary frame tag and so on. The ring is of the same color as the respective tag.. With mouse-over on the occurrences, the context of the tag and/or its number of occurrences could be shown.

> (Advanced: Furthermore, it could be possible to connect those occurrences that appear in different rings in the order they appear in the text. For example, the spatial object "the road" is a direct frame at the beginning, then becomes a secondary frame in the course of the story, and at the end is a spatio-temporal-distant frame in the story – this movement could be indicated via a connecting line between these occurrences in the display (so that the change in status of the occurrence over time can be illustrated. One should have the possibility which occurrences ought to be connected since the same spatial objects will not probably always be named in the same way).
>
> It would be also very good, if one could choose the occurrences of one frame tag and the occurrences in the other rings could be displayed according to their closeness in the text are close together in the text could also be displayed close together in the ring structure in a "ring-overlapping" structure (i.e., occurrences of the direct frame are displayed close to possible 'neighbors' of the secondary frame and so on).

Since this structure could not only be used for the illustration of spatial frames, the tool ought to be constructed in the way that the user can chose which tag ought to represent which ring in the structure before generation (so the user has to decide which is the major criterion for him/her to order the frames; the space tags could be ordered by factuality or by distance from the character, for example). In this way, the display tool could also be used for other concentric thinking structures such as different text levels ("source" text and its hypertexts), levels of narration (main story and frame stories), the different modes of thought representation (increasing degree of mediation from the center).

2. The <space-ref>-Tag could be used to show those places in Google Maps. Before a map is generated, there should be an interstage on which the user may decide for each tag whether it is a country, city (and in which country if there are several entries), street (and in which city if there are several entries) or famous place, etc.

---

1 Ronen, Ruth. *Space in Fiction*. In: Poetics Today 7, 1986, p. 421–438.

Technical Achievements

Author: Marco Petris

CATMA export capabilities

Source + tags all in one document

CATMA is now able to export a simple XML document that contains the source text labeled with tags. The CATMA tags are applied to the source text as simple XML elements named after their tag counterpart. Only selected (active) tags are considered during export process. Text which is not marked with any tag is framed by a special XML-element named „untagged".

This is a provisional format which does not support overlapping. This simplification makes it easier for Voyeur to import the data. To overcome this limitation Voyeur would need the ability to process standoff markup.

The data is send via HTTP POST to Voyeur. The resulting corpus ID is then merged with the voyeurtools.org URL. CATMA sends this URL to the system's default browser.

List of tags

CATMA is now able to export a list of tags in order of the occurrences in the user markup document. The tags are separated by a blank and the list is plain text UTF-8.

The data is send via HTTP POST to Voyeur. The resulting corpus ID is then merged with the voyeurtools.org URL. CATMA sends this URL to the system's default browser.

Sunburst

CATMA is now able to export specially crafted data for Sunburst visualization. The user can enrich a set of tags with a property named „position". This property has an integer value which indicates the ring of the Sunburst this particular tag occupies. The lowest number indicates the ring that is at the closest position to the center. The higher the number the further away from the center the ring of the tag would be. The user can also define the name of the label of the center of the Sunburst.

The data itself is send as a tab separated value table:

column number	description	part of the Sunburst
1	tag name	
2	type name	label of a wedge of the ring
3	fraction of the type frequency	together with the total tag frequency which can be obtain by the sum of all third column values that have the same first column value this value will define the size of the wedge in correspondence to other wedges the ring.
4	tag color	ring color

The data is send via HTTP POST. Unlike the other two export processes mentioned previously, the

send of the Sunburst data is not executed by CATMA directly. The data is wrapped by a HTML document which executes a POST onload. This HTML document is created as a temporary file, which is send to the system's default browser. The resulting Sunburst visualization is loaded within the same browser window and tab.

Voyeur visualizations

Source + tags all in one document

The import of this document is realized by creating a corpus. Right now voyeurtools.org reflects only the Source Document data of the document but the tag information is available from the corpus and can be exploited at a later stage.

List of tags

The import of the list of tags is realized by creating a corpus. The tags are treated as normal text and allow the usual analysis provided by voyeurtools.org

Sunburst

The data gets converted directly into the Sunburst visualization.