Design of a corpus and an interactive annotation tool for graphic literature

Jochen Laubrock
laubrock@uni-potsdam.de
University of Potsdam, Germany

Sven Hohenstein
hohenstein@uni-potsdam.de
University of Potsdam, Germany

Eike Martin Richter
eike.richter@uni-potsdam.de
University of Potsdam, Germany

The project "Hybrid Narrativity" combines work by language and literature studies, cognitive psychology, and computer science with the overarching goal to arrive at an empirically founded narratology of graphic literature, including comics and graphic novels. Comics and graphic novels provide a unique cultural form that has developed its own vocabulary, allowing for a fascinating interplay of text and visual art. After a period of neglect, they have recently been theoretically analyzed in detail by scholars in the arts, humanities and linguistics (McCloud, 1993; Groensteen, 2007; Cohn, 2013). Our aim is to provide an empirical testbed for these theories. The foundation of this endeavor is a large collection of graphic novels, which are annotated using a variety of methods. These include high-level descriptions of the work, mid-level descriptions of pages and panels, including the actors / characters, text, objects, and panel transitions, and low-level descriptions of visual elements in terms of descriptors developed in computer vision such as color histograms, GIST, SIFT, and SURF features. We are currently evaluating the addition of mid-level features from deep networks trained on photographs of real-world scenes, with quite promising first results.

These descriptions in terms of material properties are complemented by eye-tracking data, providing an empirical measure of the reader-level attention distribution and the time course of attention shifts. Thus, the digitized representation of literary and artistic works includes information on the side of "recipients", that is, readers, viewers, spectators and appreciators with their psychological and physiological responses. In a first step, eye-tracking data on a small sample of pages from selected works were collected from a large number of participants in order to evaluate general principles of attentional selection in graphic literature. Results show that reading of graphic novels is primarily governed by reading of text, and that inspection of graphical elements is apparently governed by top-down selection of story-relevant elements. In perspective, eye-tracking data will be collected for each of the works in the corpus, using a sample of pages and a smaller number of readers.

A graphical annotation tool is in development and has first been released to the public at DH 2016. This tool is based on an XML dialect that allows for the annotation of language as well as graphical elements. Future versions will include OCR support for comics fonts, and provide customizable annotation schemes, allowing other researchers to implement their own research ideas. We will also briefly present ideas on the potential to incorporate gaze-based interaction in the user interface of the tool, e.g., for the intuitive selection of objects, which will become important with the projected availability of low-cost eye trackers in the near future.

We have developed the Graphic Novel Markup Language (GBML) as an extension of John Walsh's Comic Book Markup Language (CBML; Walsh, 2012) to facilitate the description of graphical elements. These descriptions are imperative for defining regions-of-interest based mapping of eye movement data to the stimulus material. Material has been annotated using our editor, and a custom R package is under development and in use for statistical analysis of eye movements. Visual features are currently extracted using OpenCV (Bradski, 2000, 2016) and VLFEAT (Vedaldi & Fulkerson, 2008) libraries from Python and Matlab, since R does not yet provide sufficiently extensive packages for this purpose (for a promising approach see imager, Barthelmé, 2016). Deep features are based on Deep Gaze II (Kümmerer, Wallis, & Bethge, 2016), which is in turn based on the VGG-19 network (Simonyan & Zisserman, 2014). A description of artworks in terms of visual features has shown promising results in other domains (Elgammal & Saleh, 2015).

A number of analyses using the corpus data show the potential for comparative studies as well as detailed study of individual works. Many of the testable hypotheses can be derived from the

theoretical work cited above. For example, McCloud (1993) speculated that the cognitive effort of the recipient depends on the kind of panel transition, or that the empty space between panels is used to signal the passage of time. We provide empirical support for both of these hypotheses. Other examples include visual trends that can be identified across time or between regions, linguistic and visual analyses that can be used to compare text and visual complexity between different genres, and network analyses of interactions between characters that allow for an easy quantification and visualization of roles within a work, and for a comparison between works. They can also be used, e.g., to compare the complexity of a novel and its adaptation to the graphic novel format.

An in-depth analysis of a single work, the graphic novel adaptation of Paul Auster's City of Glass by Paul Karasik and David Mazzucchelli, shows that readers of graphic literature benefit from a specific expertise in decoding the different channels of information conveyed by image and text. Comics experts spend significantly more time on the image part of the panels, and this is correlated with a significantly deeper understanding of the narrative. New data suggests that this pattern replicates across samples, labs, and languages.

Taken together, we present the design of a corpus of graphic literature that is annotated using a variety of levels, including readers' eye movements. Ideas for how to make use of these data for interactive future versions are developed, and analyses of the collected data in terms of description as well as reception of the works of art are presented.

Bibliography

Barthelmé, S. (2016). imgager: Image Processing Library Based on 'CImg'. R package version 0.31. [Computer Software: https://CRAN.R-project.org/package=imager]

Bradski, G. (2000). The OpenCV library. Dr. Dobb's Journal of Software Tools, 25(11):120, 122–125.

Cohn, N. (2013). The Visual Language of Comics. Introduction to the Structure and Cognition of Sequential Images. London: Bloomsbury.

Elgammal, A. & Saleh, B. (2015). Quantifying Creativity in Art Networks. 6th International Conference on Computational Creativity (ICCC'15).

Groensteen, T. (2007). The System of Comics. Translated by B. Beaty and N. Nguyen. Jackson, MI: University of Mississippi Press.

Kümmerer, M., Wallis, T.S.A., & Bethge, M (n.d.).: DeepGaze II: Reading fixations from deep features trained on object recognition. arXiv:1610.01563

McCloud, S. (1993). Understanding Comics: The Invisible Art. New York: Harper Collins.

Simonyan, K. & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. In: CoRR abs/1409.1556. url: http://arxiv.org/abs/1409.1556.

Vedaldi, A. & Fulkerson, B. (2008). VLFeat: An open and portable library of computer vision algorithms. [Computer Software: http://www.vlfeat.org/]

Walsh, J. (2012). Comic Book Markup Language: An Introduction and Rationale. Digital Humanities Quarterly, 6(1).