Teresa M. Dobson, PhD
Associate Professor
Department of Language and Literacy Education
University of British Columbia
2125 Main Mall
Vancouver, BC V6T 1Z4
Office: (604) 822-8365
Fax: (604) 822-9515
Email: teresa [dot] dobson [at] ubc [dot] ca

Stan Ruecker, PhD
Associate Professor
Institute of Design
Illinois Institute of Technology
350 North LaSalle Street, 4th floor
Chicago, Illinois 60654
sruecker@ualberta.ca

Monica Brown
Doctoral Candidate
Department of English
University of British Columbia
397 – 1873 East Mall
Vancouver, BC   V6T 1Z1
mm2brown@interchange.ubc.ca

Dustin Grue
Doctoral Student
Department of English
University of British Columbia
397 – 1873 East Mall
Vancouver, BC   V6T 1Z1
dustin.grue@gmail.com

Neither Bicycles Nor Sheep: Crowdsourcing Semantic Encoding for Elements of Plot
Teresa Dobson, Stan Ruecker, Monica Brown, and Dustin Grue

In this paper, we discuss our recent efforts to use crowdsourced workers as a means of obtaining semantically-encoded XML of Hemingway's short story "Hills Like White Elephants." Our approach was to have people recruited through Amazon's Mechanical Turk carry out the encoding for micropayments, after which we were able to compare the results of their attempts with a previously-encoded copy of the same text completed by graduate students from the Department of English at

University of British Columbia in Canada who were research assistants in this project.

The larger context for this experiment is the Plotvis project (Figure 1), where we set out, using a rather conventional digital humanities approach, to develop a piece of interactive software that would address a recognizable gap in teaching and research (e.g. Dobson et al. 2011).
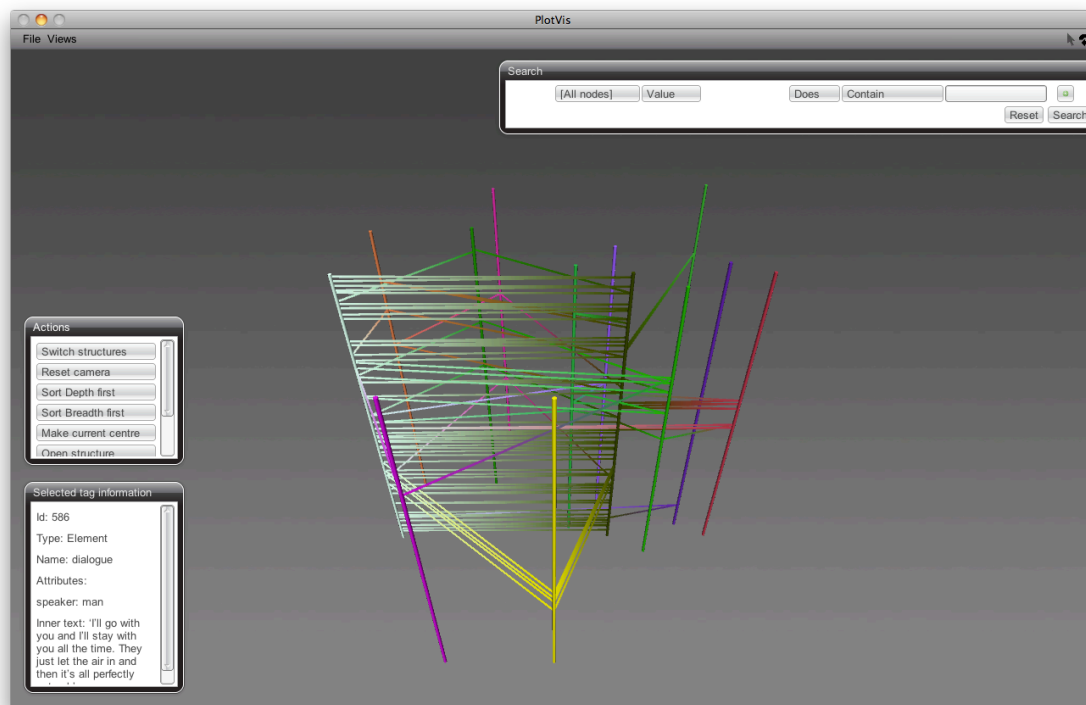

Figure 1. This screenshot from the Plotvis system shows interactions between characters and objects in Hemingway's "Hills Like White Elephants". The design is by Piotr Michura; the system was built in Unity by Omar Rodriguez.

We were interested in providing readers with additional models of plot that could be used to supplement conventional approaches to graphing plot structure such as Freytag's Pyramid (Figure 2), which is in widespread use in grade school English literature classrooms across North America and beyond. Devised by Freytag based on his analysis of Shakespearean and classical tragedies, Freytag's pyramid is inappropriately applied to other kinds of literature and is also limited in representing the complexity of narrative by its reliance on the Cartesian graph.
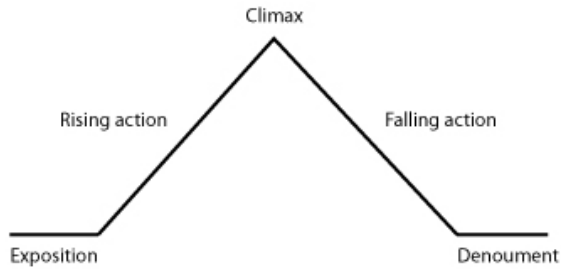
Figure 2. Freytag's pyramid (Frytag, 1863/1983).

In support of the Plotvis visualizations, we developed an XML schema that encodes for elements related to the narrative, namely action, dialogue, narration, characters, and objects (Brown et al. 2011). The process was relatively complex and time-consuming, but eventually we had assembled a set of three short stories that were encoded in the schema and could be viewed in the software. In addition, we ran a pilot study, where students in grades 11 and 12 were given short stories and asked to begin encoding them for the purposes of identifying what they considered to be the relevant parts of the plot (Grue et al. 2011).

These activities brought us to the point of having a testbed of encoded fiction and a prototype tool that relied on that encoding for its core functionality. The question then arose: since encoding of this kind is relatively slow and requires a relatively high degree of human intelligence, would it be possible to successfully crowdsource it, using readily available conventional crowdsourcing mechanisms like the Mechanical Turk (M-Turk), which rely on micropayments rather than volunteer enthusiasts? We were familiar with a number of previous projects that had used M-Turk for artistic and digital humanities goals, such as the Bicycle Built for 2000 (Koblin and Massey 2009), where notes sung by individual people in 71 countries were assembled into a song, the Sheep Market (Koblin 2007), where ten thousand people each made a line drawing of a sheep, and "A Modest Proposal" (Organisciak 2011), where in an ironic commentary on human exploitation, M-Turk workers rewrote Swift's famous short story in contemporary register for Twitter (follow him at @swiftsays).

In our case, we set up a Mechanical Turk project for workers to encode Hemingway's short story "Hills Like White Elephants", which had been previously encoded by graduate research assistants working with our team. We broke the story into 137 text snippets consisting of individual sentences or short paragraphs and tasked each worker with applying one tag as many times as necessary to each snippet. So one worker encoded, for instance, all of the objects in the sentence "The hills across the valley of the Ebro were long and white", while another encoded all the characters in "The American and the girl with him sat at a table in the shade, outside the building." It was possible for someone to decide that the sentence required no tagging. We paid two cents per sentence or short paragraph, for a total

of $13.70 per round. For the purposes of being able to take an average result, we duplicated the exercise three times for a grand total of $41.10.

- $0.02 per snippet
- 137 sentences
- 5 tags

The project is still in progress and will be complete in February. We will report our results and conclusions in this proposed conference paper.

References
Brown, Monica, Dobson, T.M., Grue, D. & Ferreira, L. (2011). Towards a Schema Design for Narrative Encoding. Paper presented at the Digital Humanities 2011 conference. Stanford, June 19-21.
Dobson, T.M., Michura, P., Ruecker, S., Brown, M., & Rodriguez, O. (2011). Interactive visualizations of plot in fiction. *Visible Language*, 45, 166-185.
Dobson, T.M., Ruecker, S. Michura, P. & Rodriguez, O. PlotVis. [Computer Software]. Retrieved from http://blogs.ubc.ca/plotviz/
Freytag, G. (1863/1983). *Die technik des dramas*. Leipzig.
Grue, D., Dobson, T.M. and Brown, M. (2011). From Envisaging to Visualization: Young People's Narrative Reading Practices. Paper presented at the Digital Humanities 2011 conference. Stanford, June 19-21.
Koblin, A. (2007). "The Sheep Market." Retrieved from http://www.thesheepmarket.com/
Koblin, A. & Massey, D. (2009). "Bicycle Built for 2000." Retrieved from http://www.bicyclebuiltfortwothousand.com/
Organisciak, Peter. (2011). "A Modest Proposal". Retrieved from http://www.porganized.com/