DIAGRAM Center’s Tools for Improving Image Descriptions
presented by Steve Landau

Touch Graphics’ is working under subcontract to DIAGRAM Center on new ways to improve the quality and consistency of image descriptions in digital talking books (DTB). This year’s work will focus on developing templates for streamlining creation of well-crafted descriptions of textbook images as part of an overall textbook transcription workflow. Touch Graphics will demonstrate an online tool that presents human transcribers with a series of questions about individual images; responses to these questions will be used to build descriptive paragraphs to replace visual images in the final DTB. The system will mix human- and system-generated sentences and phrases, and then allow human transcribers to revise and elaborate on descriptions based on their own interpretation of the graphic and their awareness of the context of the page where it appears.

We believe that image descriptions produced using these methods will be superior to those by human transcribers using traditional, less automated means. We know from observation that readers appreciate descriptions that are terse and consistent, both qualities that human transcribers often find hard to achieve when working without a suitable framework. Readers appreciate concision and predictability when working to rapidly develop accurate mental models of images, and we hypothesize that our online tools will lead to superior results in reader comprehension, while improving throughput in large-scale, geographically-distributed transcription operations such as Bookshare’s POET cloud-based service. This year’s work will include a field test of the online system. Those interested in participating should contact Ting Siu at ts@touchgraphics.com

The example to the right shows how a typical description paragraph is generated for a common type of graphic. Human and system-generated fragments are combined to produce a descriptive paragraph, ready for review and post-editing by the human transcriber for added nuance and specificity. In the example to the right, we use color-coding to indicate the meaning of each fragment:

- black = system-generated text.
- blue = human-generated responses to generic questions.
- red = prompts appear where additional information should be added by the human transcriber
- green = comments or embellishments supplied by a human transcriber in post-editing.

**Project collaborators**
- Josh Miele, Ph.D. Smith-Kettlewell Eye Research Institute
- Lucia Hasty Rocky Mountain Braille Associates
- Steve Landau Touch Graphics, Inc.
- Ting Siu University of California, Berkeley
- Val Morash University of California, Berkeley