Evaluating a Text-to-Scene Generation System as an Aid to Literacy

1Bob Coyne, 1Cecilia Schudel, 2,3Michael Bitz, and 1Julia Hirschberg

1Columbia University, New York, NY, USA
2Ramapo College of New Jersey, Mahwah, NJ, USA
3Center for Educational Pathways, USA
coyne@cs.columbia.edu, cms2223@columbia.edu, bitz@edpath.org, julia@cs.columbia.edu

Abstract

We discuss classroom experiments using WordsEye, a system for automatically generating 3D scenes from English textual descriptions. Input is syntactically and semantically processed to identify a set of graphical objects and constraints which are then rendered as a 3D scene. We describe experiments with the system in a summer literacy enrichment program conducted at the Harlem Educational Activities Fund with 6th grade students, in which students using the system had significantly greater improvement in their literary character and story descriptions in pre- and post- test essays compared with a control. Students reported that using the system helped them imagine the events in the stories they were reading better. We also observed that social interaction engendered by this process was a strong motivator.

Index Terms: text-to-scene generation

1 Introduction

Rendering complex 3D images normally requires special training and expertise – especially images that convey realistic human emotion. WordsEye [1] is a software system that converts text input into a 3D scene without requiring any kind of technical knowledge from the user. We tested our current version of WordsEye in a Harlem-based summer enrichment program for 6th grade students, who used the system in an English literature class. Before the class began, students provided a writing sample. Over the course of five weeks, students were asked to re-create scenes from Aesop’s Fables and George Orwell’s Animal Farm in WordsEye. They then provided a new writing sample at the end of the summer program. We found that students who used WordsEye showed significantly higher improvement on written essays than students who took the same course without WordsEye.

2 Previous Literature

The connections between visual perception and language acquisition have been a major focus of researchers in the fields of cognition, psychology, and neuroscience. Barbara Landau [2], for example, presented subjects with a block on a box and told them “the block is acorp the box.” Subjects interpreted acorp to mean on. In contrast, when shown a stick on a box, subjects interpreted acorp as across. These results, and those of many other similar experiments, highlight well-documented theories of cognitive development by Piaget [3]: people form mental schema, or models, to assimilate the knowledge they already have and accommodate new information presented to them. Gardner [4] identified the importance of visual-spatial “intelligence,” furthering the body of research linking visual perception to other important human activities, including language acquisition. Despite this research and base of knowledge, schools in the United States have placed a much larger emphasis on language acquisition than visual perception. Perceptual psychologist and art theorist Rudolf Arnheim [5] demonstrated that the dismissal of perception, and subsequently of the visual, has resulted in widespread visual illiteracy.

Over the past two decades, many educational researchers have focused on the connections between visual literacy and conventional literacy – that is, reading and writing. For example, there is extensive scholarship in this area in relation to a very common artifact: children’s picture books. Picture books wed written text and visual imagery in a natural and cohesive manner. Lawrence Sipe [6], a researcher of children’s literature, thoroughly examined the interplay between words and pictures in a picture book. Sipe described a text-picture synergy that produces an effect greater than that which text or pictures would create on their own. He cited the work of numerous scholars who attempted to depict the relationships between text and imagery in a picture book, ranging from the idea that pictures extend the text [7] to the concept that pictures and text limit each other [8]. Sipe concluded, “As readers/viewers, we are always interpreting the words in terms of the pictures and the pictures in terms of the words...The best and most fruitful readings of picture books are never straightforwardly linear, but rather involve a lot of reading, turning to previous pages, reviewing, slowing down, and reinterpreting” (p. 27).

Sipe’s connections between visual imagery and written text in children’s books primarily relate to the discipline of reading. Research in the realm of comic books in education further this connection to the discipline of writing. Frey and Fisher [9] presented a series of chapters dedicated to this subject, highlighting the possibilities for writing development through visual imagery. Bitz [10] demonstrated the power of student-generated comic books for writing development through a national program called the Comic Book Project. Similar outcomes for both reading and writing have been demonstrated through other visual media, including videogames [11]; film [12]; and website development [13].

3 The WordsEye Text-to-Scene System

WordsEye is a web-based application that generates 3D scenes from text typed by a user into a browser. Scenes are generated as follows: Input sentences are parsed into a phrase structure representation using a hand-constructed grammar with feature unification to enforce subject-verb agreement and other syntactic and semantic constraints. The parse tree is then automatically converted to a dependency structure where individual words are represented by nodes, and arcs between nodes represent the syntactic relation between the words. This conversion is facilitated by the grammar that designates a head component for each production rule. Dependency structures
OpenGL (http://www.opengl.org) and displayed in the user’s
constraints to the selected 3D objects. The scene is rendered in
automatically generated by applying these graphical
relations representing the position, orientation, size, color,
texture, and poses of objects in the scene. The graphical
objects themselves are automatically selected from a library of
2,200 3D objects and 10,000 images (used as textures). The
library objects are tagged with spatially-relevant regions used
to resolve spatial relations (for example, the area of a chair
where someone would sit). The 3D scene itself can now be
automatically generated by applying these graphical
constraints to the selected 3D objects. The scene is rendered in
OpenGL (http://www.opengl.org) and displayed in the user’s
browser as a jpeg image. Users can manually adjust the
viewpoint, causing the server to quickly redraw the scene from
a new angle and update the image. A user is free to modify
text as often as they like to add more objects, fix mistakes, or
specify different spatial relations. Once the user is satisfied
with a scene, the system will optionally create a higher-quality
final rendering with reflections and shadows by ray-tracing
their scene in Radiance (http://radsite.lbl.gov/radiance). Users
can post their final renderings to an online gallery where they
can give it a title and post comments on each other’s pictures.

The face library was built using the FaceGen 3D graphics
package (http://www.facegen.com), which allows for high-
level control of facial features based on statistical modeling.
This software provides overall facial controls like Age and
Gender, as well as highly specific Shape, Color, and
Asymmetry controls, and various facial Morphs that apply
emotion to the face. With this software, we were able to
incorporate FaceGen capabilities into the WordsEye system,
allowing users to choose faces of well-known people and add
particular emotions to these characters. The system currently
supports the 6 “basic” emotions: sadness, happiness, fear,
ger, anger, surprise, and disgust. Emotion words from WordNet
and a thesaurus were gathered and divided into three degrees
(high, medium, low), depending on their Activation scores
from Whissell’s Dictionary of Affect in Language [15]. These
words were then associated with particular FaceGen
parameters and added to WordsEye as keywords.

One major challenge faced by our system is handling the
ambiguity and vagueness of language. 3D scenes require all
objects and spatial relations to be explicitly defined. As a
result, only the most literal language can be directly translated
to graphics. For example, if the user were to type The boy fed
the cat, the system would not know the location, the size and
color of the cat, what the boy was wearing, what sort of food
he fed it, and so on. We are currently working on addressing
some of these issues by developing a resource of lexical and
real-world knowledge useful in providing default locations and
poses to help resolve underspecified or vague text. In our
educational testing, these limitations forced the students to
adapt their descriptions to what the system can handle. Several
students reported that they enjoyed the problem solving
aspects of translating their ideas into concrete and explicit
language. One related challenge for the system is providing
appropriate feedback in cases when the input cannot be
processed, whether because of spelling or grammatical
mistakes or because of limitations in the system itself. One
example is that missing objects and misspelled words are
handled by simply inserting extruded 3D text of the word into
the scene, giving the student immediate visual feedback.

4 Classroom Testing

We performed preliminary testing of the system in schools in
Virginia. After seeing WordsEye at the Innovate 2007
Exposition (hosted by the Virginia Department of Education),
K-12 public school teachers from Albemarle County asked to
use it in their classes as a tool for ESL remediation, special
education, vocabulary enhancement, writing at all levels,
technology integration, and art. Feedback from teachers and
students was quite positive. In one school with a 10% ESL
population, a teacher used it with 5th-6th graders to reinforce
specificity of detail in descriptive writing, noting that students
are “very eager to use the program and came up with some
great pictures.” Another teacher tested it with 6th-8th grade
students “in a special language class because of their limited
reading and writing ability,” most reading and writing on a
2ND-3RD level. The students found the software fun to use,
an important element in motivating learning. As one teacher
reported, “One kid who never likes anything we do had a great
time yesterday...was laughing out loud.”

To test the hypothesis that WordsEye could serve as an
effective alternative literacy tool, we designed a controlled
experiment for middle school children enrolled in a summer
enrichment program run by the Harlem Educational Activities
Fund (HEAF). In the trial, twenty seven emerging 6th grade
students in a HEAF literature course were given a writing pre-
test. 41% of the students were female and 59% were male;
89% were African American, 4% were Native American, and
7% were other ethnicities. Half of the students (control) were
randomly chosen to participate in a conventional literature
course (i.e., group discussions and thematic analysis); the
other half (treatment) were introduced to WordsEye and
shown how to use it to construct pictures from text. The

Figure 1: Three stages of processing for A red cat is on the green chair.
control curriculum was determined and implemented by HEAF without interference from the researchers. It consisted of a variety of activities ranging from book discussion groups to the development of an original puppet show. There was some technology integrated into a number of those activities, consisting of word processing, Internet searches, and visual design software (Adobe Photoshop). The control and treatment groups spent the same amount of time on their activities. Over the next 5 weeks, the WordsEye group used WordsEye for 90 minutes a week to create scenes from the literature they read, including Aesop’s fables and Animal Farm.

We developed a curriculum (See Appendix) that helped instructors integrate WordsEye into the learning goals of the summer academy. The WordsEye group was also introduced to WordsEye’s face manipulation capabilities, which allowed them to include their own and other well-known people’s faces in scenes, and to modify the expressions of these faces by specifying particular emotions. For the rest of the course, the WordsEye group participated in the same type of class discussions and writing exercises as the control group. At the end of the course, all of the students wrote essays based on the literature they had read; pre- and post-course essays were scored by independent, trained raters. The criteria were: a) Organization and Structure b) Written Expression c) Quality and Depth of Reflection d) Use of Vocabulary e) Mechanics: Grammar, Punctuation, Spelling. Each category was judged on a scale of 1 (poor) to 5 (excellent). The average pre-test treatment score was 15.8; the average pre-test control score was 18.0. We determined that this was as close to a baseline measurement as possible, given the parameters of the summer program. The raters met to review the evaluation rubric and then independently judge two writing samples. The scores were discussed among the group, and inter-rater reliability was determined sufficient at 92%. The WordsEye group showed significantly greater improvement than for control group (Table 1). Note that, as this study was a straightforward comparison between the growth scores of two independent and randomly assigned groups with a small sample size, the researchers used a two-sample t-Test to determine statistical significance: Difference = mu (1) - mu (2); Estimate for difference: 4.81; 95% CI for difference: (0.08, 9.54); t-Test of difference = 0 (vs not =): t-Value = 2.16 p-Value = 0.047 DF = 16. As one of the students said “When you read a book, you don’t get any pictures. WordsEye helps you create your own pictures, so you can picture in your mind what happens in the story.”

5 Discussion and Future Research

We have demonstrated that Text-to-Scene generation can provide a useful approach to literacy skills training through a formal field experiment. In future work we will enhance features of WordsEye, including integrating face manipulation into the program; adding a capability to infer appropriate emotions from input text more robustly; and extending the system to identify spelling and grammatical errors. We are also developing a set of self-contained instructional modules to allow/test distance learning in future pilots with remote school systems. We also want to explore educational settings where English is a second language for most students.

6 Appendix: WordsEye Pilot Curriculum

Session 1: Introduction of the Platform

a. Introduction:
• Give students the same sentence starter: The dog is on the
• Change image selection of dog.
• Change color of dog.
• Change size of dog with descriptors (large, huge, tiny) and numbers (10 feet tall).
• Change color of the sky.
• Change texture of the ground.
• Add something on top of the dog.
• Add something below the dog.
• Render the final image.
• View scene in My Portfolio and the Gallery

b. Scene re-creation:
• Give students the following scene and see who can recreate it most accurately:

(c) Literary exploration:
• Have students re-create a scene from one of Aesop’s fables, including characters, backgrounds, and anything else important to understanding the scene in depth.

Session 2: Working with Fables

a. Introduction:
• Have students open their scene for: The dog is on the...
• Change the point-of-view with the camera angles.
• Zoom in or out with the camera angles.
• Change the texture of the ground.
• Import an image and add it to the scene.
• Render the final scene.
• Use 2D effects to add another visual element to the scene.
• View scene in My Portfolio and the Gallery

b. Warm-up:
• Provide students with the following text from Aesop’s fables. Put students in teams, one for each sentence. Ask them to recreate their assigned sentence in WordsEye so that the entire fable is recreated at the end of the activity: “A monkey perched upon a lofty tree saw some Fishermen casting their nets into a river, and narrowly watched their proceedings. 2) The Fishermen after a while gave up fishing, and going home to dinner left their nets upon the bank. 3) The Monkey, who is the most imitative of animals, descended from the treetop and endeavored to do as they had done. 4)”

<table>
<thead>
<tr>
<th>WordsEye Group</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td>18.05</td>
<td>20.59</td>
<td>2.54</td>
</tr>
</tbody>
</table>

Table 1: Evaluation of student essays
Session 3: Animal Farm Study (Part 1)

a. Warm-up:
- Give students the following Animal Farm text to recreate as a scene in WordsEye, reminding them that they won’t be able to include all of the details and vocabulary.
- At one end of the big barn, on a sort of raised platform, Major was already ensconced on his bed of straw, under a lantern which hung from a beam. He was twelve years old and had lately grown rather stout, but he was still a majestic-looking pig, with a wise and benevolent appearance in spite of the fact that his tushes had never been cut.

b. Literary Exploration:
- Explain that WordsEye will be a tool to storyboard their Animal Farm skits at the final presentation of the Summer Academy, where the storyboards will be enlarged to poster-size and presented. Start discussion on how they can use WordsEye to plan their skit scenes.
- Have students work in groups to create their skit scenes. They can divide the skit into beginning, middle, and end. Each scene should include the background & foreground for what the audience will see during the performance.
- Have each group present their WordsEye creations to the class. Tell students that the completed scenes will be made into posters and displayed at the final presentation.

Session 4: Animal Farm Study (Part 2)

a. Warm-up:
- Ask students to think of their favorite Animal Farm character. Now create a WordsEye scene by placing that character in the middle of New York City. What does the character see? Does the street scene look different because of this character’s presence there?

b. Literary Exploration:
- Have students continue designing scenes for their Animal Farm skits. Scenes should be finalized by the end of the session. Tell students that the completed scenes will be made into posters and displayed at the final presentation.
- Ask students to volunteer to present their scenes. See if other students can discern the moral of the fable from the scenes and presentation.

Session 5: Final Presentation

- Have each group present their WordsEye creations to the class, giving a synopsis of their stories and showing panels of their work. Have class discuss each scene: characters, imagery, text, connection to Animal Farm.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No.IIS-0904361. The authors thank Danielle Moss Lee, Nicole Smith, and Tanya Wiggins (at HEAF), and Adele Chase, Kenneth Harvey, and Daniel Bauer (at Columbia) for their help.

References