Corpus Methods and Textual Visualization To Enhance Learning in Core Writing Courses

David Kaufer
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
+1 412-268-1074
kaufer@andrew.cmu.edu

Suguru Ishizaki
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
+1 412-268-4013
suguru@cmu.edu

ABSTRACT

Writing tasks require countless composing decisions that are typically beyond the conscious grasp of writers. Much of the skill of being "text-aware" inheres in understanding that texts produced from classroom assignments are not just composed of words and sentences but of highly structured and often highly predictive composing decisions. However, the decision-making underlying writing is an extremely abstract idea that is hard to make tangible for students. Although a significant number of pedagogical approaches have been investigated in the past three decades, the means to help students acquire more tangible understanding and control of their composing decisions has not been addressed.

We propose to address this gap by developing a corpus-based learning tool to help students notice and reflect on composition decisions in their writing and to become resultantly more self-aware and reflective writers. This approach builds on an existing corpus-based text analysis tool called DocuScope, which for over a decade was successfully used for these purposes in a graduate pilot course. The goal of this project is to extend this approach to support the core writing courses at our university.

Keywords

Textual Awareness, Textual Visualization, Corpus-Based Instruction

1. INTRODUCTION

Writing tasks require countless composing decisions that are typically beyond the conscious grasp of writers. Much of the skill of being "text-aware" inheres in understanding that texts produced from classroom assignments are not just composed of words and sentences but of highly structured and often highly predictive composing decisions. A fundamental goal of Carnegie Mellon's core writing courses is to help students develop this textual awareness so that they are able to make appropriate compositional decisions for different text types. Unfortunately, the decision-making underlying writing is an extremely abstract notion and hard to make tangible for students. While various pedagogical approaches have been investigated over the past 30+ years, making tangible the decision-making underlying writing has eluded these approaches.

The goal of our project is to develop a suite of corpus-based learning tools that will help students notice hidden structures and composing decisions in writing, and become more self-aware and reflective writers.

2. OUR APPROACH

Our approach builds on a graduate-level writing course developed and taught by Kaufer over a decade, in collaboration with Ishizaki. In the course, students used DocuScope [1]—a dictionary-based tool for rhetorical text analysis with a suite of tools for interactive visualization—that allowed students to visualize differences in the rhetorical strategies underlying their drafts and across the different genres they were assigned to write.

DocuScope transformed the writing classroom into a design studio–like environment for writing, where—unlike a typical writing course—students could compare their writing at a glance as if they were comparing posters on a wall (Figure 1). DocuScope, then, would allow students to select specific writing to view how certain rhetorical strategies are implemented in terms of composing decisions (Figure 2).

We informally observed that the visualizations helped enhance students' awareness of (a) their composing decisions and (b) the relationship of their decision-making to their writing context and the genre of text they were seeking to produce. Although we have no definitive understanding of how this works, we suspect that allowing students to see their composing decisions visualized after the fact creates grounded evidence for claiming ownership of those decisions and using those decisions to explain their situated goals of composing with sharpened clarity.

In our current project, our goal is to extend the use of DocuScope to a much larger scale by embedding it in a freshman-writing course and a popular professional writing course. Each student will receive feedback based on the text-analysis that compares and situate his or her writing against the historical student data. Students of any cohort on any assignment will be able to compare their writing against a historical cohort writing on the same assignment.

More specifically, we are developing a tool for automatically generating visual reports that highlight salient structures and composition decisions in the students' own writing in relation to the historical data as well as writing by other students in class. We hypothesize that enhancing students' awareness of their low-level composition choices can enhance their overall metacognitive awareness as writers.

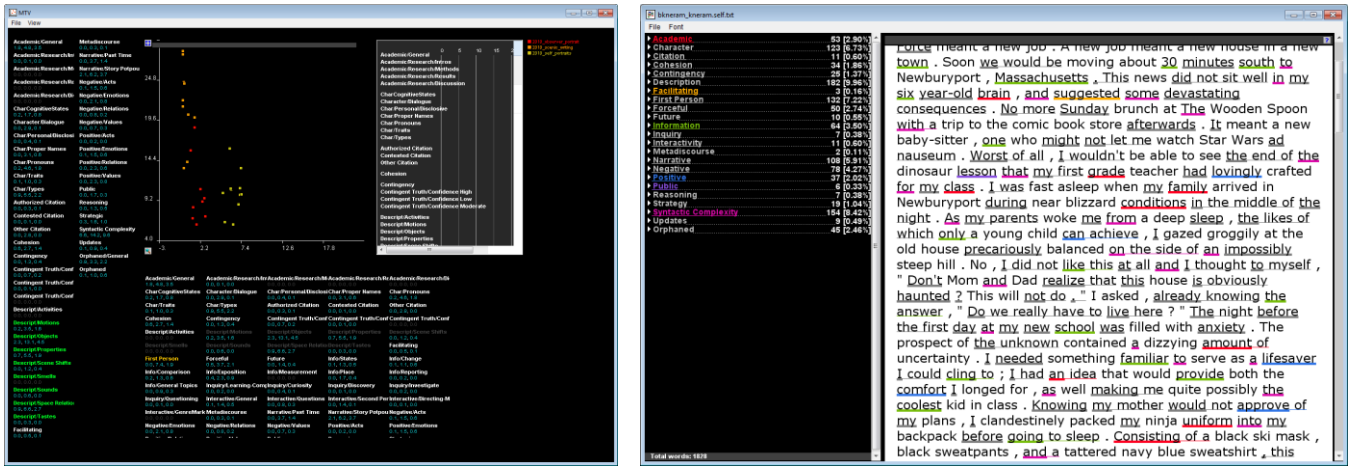

Figure 1. LEFT: Multi-Text Visualization (MTV)—This screenshot shows three genres of a writing course. Yellow dots indicate a single discrete student writer's text on the self-portrait assignment. Red dots indicate a single discrete student writer's text on the observer-portrait assignment. Orange dots indicate a single discrete student writer's text on the scenic writing assignment. The X-axis represents the amount of "first person" in each text. The Y-axis represents the amount of "description" (writing for the eyes and ears) in each text. Notice that the self-portraits are separated from the other genres on first person. Notice that the scenic texts are separated from the other genres on description.

RIGHT: Single-Text Visualization (STV)—In this screenshot, we see how a student writer or teacher can drill down from MTV and see how DocuScope categories tag individual words and word strings. A number of categories are highlighted. Notice how the word "suggested" is tied to the facilitating category through color-coding. To suggest something is to help another facilitate action.

3. CHALLENGES

While the course taught by Kaufer was successful [2, 3], the text analysis tool was not fully automated. Running DocuScope therefore required a manual process that had to be handled by the instructor (Kaufer). This original context worked as well as it did because (1) the instructor was extremely familiar with the tool and (2) he was able to assist students in interpreting the analysis.

In order to scale the use of this environment for core writing courses with many sections with different instructors, we must make it highly user-friendly and capable of presenting results clearly to non-writing experts—i.e., students. Accordingly, we are currently addressing the following specific research questions.

- What are optimal ways to integrate automated reporting into undergraduate writing instructions? We are exploring how these reports can be integrated meaningfully for students in our core writing classes. We are also examining the extent to which these reports can positively impact student understanding of structures and composition decisions in their own writing.
- What are the optimal statistical methods for uncovering the most salient composing choices from data generated from DocuScope? In order to fully automate the analysis and report generation, we are exploring statistical methods for uncovering salient features in a student's writing.
- What are optimal ways to visualize the results of statistical analysis? We are exploring optimal ways students' composing decisions can be visualized.

4. DEMO

In this demonstration, we will provide an overview of the technology we have developed so far, including the tool to mine the corpus, the visualizations (i.e., reports) we are experimenting to provide feedback to students.

We are currently working with a team of statistics professors and students to help us answer some of these questions. By the time of the workshop, we should have more concrete results about helpful visual feedback to students. We will also discuss our pedagogical philosophy for the way students can productively use this feedback, as well as some of the challenges of getting this ambitious project off the ground.

5. ACKNOWLEDGMENTS

Our thanks to Danielle Wetzel, Necia Werner, Xizhen Cai, Ann Lee, Joel Greenhouse, Arianna Garofalo, Chushan Chen and Binghui Ouyang for vital help on this project.

6. REFERENCES

[1] Ishizaki, S., & Kaufer, D. (2011). Computer-aided rhetorical analysis. In P. McCarthy & C. Boonthum (Eds.), Applied Natural Language Processing and Content Analysis: Advances in Identification, Investigation, and Resolution. Hershey, PA: IGI Global.

[2] Kaufer, D., Geisler, C., Vlachos, P., & Ishizaki, S. (2006). Mining textual knowledge for writing education and research. In L. v. Waes, M. Leijten, & C. Neuwirth (Eds.), Writing and Digital Media (pp. 115-130). Oxford, UK: Elsevier Science.

[3] David Kaufer, Suguru Ishizaki, Jeff Collins, and Pantelis Vlachos, (2004) "Teaching Language Awareness in Rhetorical Choice Using IText and Visualization in Classroom Genre Assignments." Journal for Business and Technical Communication, 18:3 361-40