Understanding Revisions in Student Writing through Revision Graphs

Antonette Shibani, Simon Knight and Simon Buckingham Shum

University of Technology Sydney, PO Box 123, Broadway, Ultimo, NSW 2007, Australia
antonette.aileenshibani@student.uts.edu.au

Abstract. Text revision is regarded as an important process in improving written products. To study the process of revision activity from authentic classroom contexts, this paper introduces a novel visualization method called *Revision Graph* to aid detailed analysis of the writing process. This opens up the possibility of exploring the stages in students' revision of drafts, which can lead to further automation of revision analysis for researchers, and formative feedback to students on their writing. The Revision Graph could also be applied to study the direct impact of automated feedback on students' revisions and written outputs in stages of their revision, thus evaluating its effectiveness in pedagogic contexts.

Keywords: Learning Analytics, Writing Analytics, Revision Analysis, Writing, Revision process, Visualizations, Revision Graph

1 Introduction[i]

Text revision is considered an important process in writing to support the reworking of writer's thoughts and ideas, playing a major role in the outcome of the writing [1]. The cognitive process theory of writing defines revision as a recursive process that can be called any time during writing [2]. Writers engage in task definition, evaluation, goal-setting and strategy selection to make revisions, thus leading to improvements in a text. To teach students revision skills to improve their writing, it is essential for researchers and educators to understand what contributes to good revision and how it occurs. This can be supported by Writing Analytics, which could be thought of as a sub field of Learning Analytics that involves "the measurement and analysis of written texts for the purpose of *understanding writing processes and products*, in their educational contexts" [3]. Such analytics might be deployed both to provide feedback to students on their revisions, and in research to understand the revision process using textual features.

The focus of this article is on studying the *process* of revision, which can help researchers and educators gain insights into the processes involved in the creation of a written document and the use of feedback in various stages of revision (an extended version can be found at [4]). In earlier work, such processes in revision have been studied using personal testimonies of participants regarding their cognitive process in revising, or by process tracing and participant-observer methods that observe the behaviors involved in revision [1]. Resource intensive manual observation and coding can be

improved with advanced online trace data collection and analysis techniques to develop visualizations that represent the process of drafting and revision. To visualize modification patterns in an online document, Caporossi and Leblay [5] developed a graph theory approach to represent the movement of text through a document using log data of keystrokes and cursor movements from the document editing process. However, there is no evidence that educators would find keystroke-level data insightful for understanding revision patterns, nor that students would find this meaningful feedback to improve their writing. More recent work introduced the use of Sequence Homology Analysis (SHA) to study the evolution of public speech drafts by comparing the changes in characters, and proposing a draft network based on the strength of revisions made [6]. In this paper, we introduce a 'Revision Graph' to visualize the evolution of writing in terms of the actions that led to the final product, and explain its potential for studying writing revisions in various contexts.

2 Research Context

The research context for this paper is a pedagogic intervention that made use of a web-based tool integrated with multiple tasks to help students write better essays for their subject in authentic classroom settings [7]. In the main revision task, students worked on revising a short essay that was provided to them, to produce an improved version (rationale in [8]), in study conditions with and without using automated writing feedback. To study the features of revision, the revised essays were marked by tutors on a scale of 0-3 (0- degraded, 1-no change, 2- minor improvement, 3-major improvements), based on which the essays are characterized as improved or degraded. Drafts from students' revisions were captured every one minute (unobtrusively) for collecting revision data using the AWA-Tutor tool which scaffolds the tasks in the intervention, and students' usage of automated feedback was also recorded [9].

3 A novel approach to revision analysis

We provide a novel analysis of revisions over multiple drafts created through the text-revision exercise using a 'Revision Graph', exemplified by a sample improved essay and a sample degraded essay written by the students in our context. This draft level analysis can aid to uncover the previously unknown processes involved in the editing of the final revised essay. This new manual analysis focuses on the ordering of sentences and revision actions, which could be potentially automated. In this revision graph (Fig. 1), the nodes represent sentences from the drafts and the edges represent changes in the organization of sentences across multiple drafts. The sentences are represented in the sequence of occurrence across the paragraphs. The colors of the nodes indicate the type of revision action made at the sentence level: i) minor revisions are when students predominantly use the given text, but add or substitute few words, ii) major revisions are when students add a substantial number of words and explanations to the given text with the inclusion of their own writing, iii) no changes made and iv) no change in

the current stage, but deleted in the next stage. Red triangles represent that automated feedback was requested during the revision process. Dotted edges are used to represent the repetition of similar concepts across multiple sentences inside a draft. This could be a good indicator of word repetition/ overlap leading to high cohesion in the document.

Fig. 1 (left) shows the revision graph constructed from the sample improved essay's drafts to show the evolution of a high-scoring revised essay. The drafts were selected from certain intervals (every 6 minutes in this analysis) using the time spent on revision. The graph shows the stages in the revision of the given text containing four paragraphs and 15 sentences to the final product containing two paragraphs and 10 sentences. In the first draft stage, the student has deleted some broad introductory sentences from the original essay. The first paragraph of the draft has been shaped up by making minor and major revisions to the given sentences and reordering them, while the other paragraphs remain untouched. In the second stage of drafting, the student has deleted the previous second paragraph and mainly worked on the revision of this paragraph from the other paragraph sentences. Here the text has been reduced to three paragraphs.

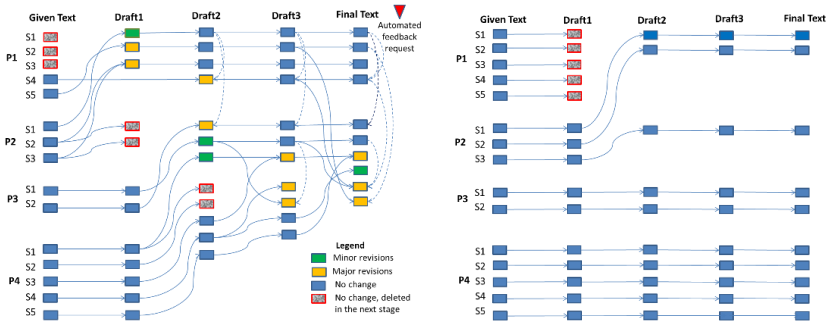

Fig. 1. Revision graph of sample improved essay (left) and sample degraded essay (right)

From the third draft, the first paragraph remains stable. The student has made some extensive changes to the sentences by revising and consolidating them to produce a final text consisting of only two paragraphs. The number of references to the previously written words increases in each stage of the draft as shown by the dotted edges. The final text has many such cross references made to the previous sentences, which has improved the cohesion of the text. This student requested automated feedback (red triangle) after completing the final text and made no more changes after that. This information is made visible by matching the timestamp of feedback request with those of the drafts. It informs that the changes made to the text by the student were *not an effect of the feedback received*. In cases where we do not have such process information to study writing, it is feasible that the revision effect is attributed to feedback, but they are in fact not related. This revision graph is thus serving its purpose of making visible, at an appropriate granularity, the nature of the revisions, and whether the automated feedback component impacted subsequent revisions.

In the revision graph of a sample degraded essay shown in Fig. 1 (right), there are no edits made by the student to the given sentences. The introductory sentences have been removed in the first draft, and sentences have been reorganized in the second draft. No further changes have been made from the second draft to create the final revised essay, leading to a degraded version of the given text. The last three drafts have remained stable, meaning the student has stopped working in the last few minutes of the revision task. The above manually constructed revision graphs could potentially be automated for a large scale analysis of revision process.

4 Conclusion and Future Work

This paper introduced a process centric method to study revision with the construction of a 'Revision Graph' to study the evolution of writing. This novel visualization revealed a pattern of actions that led to the final product like addition, deletion and reorganization of sentences in the generation of the text, showing the importance of understanding textual restructuring and the revision process in writing. It demonstrated the opportunity to study the diverse ways in which good or poor writing may evolve in its revision stages. One could also imagine the visualization being applied to other specific changes we would like to study, like the types of revisions (e.g. content, concepts, rhetorical moves, surface errors, etc.) instead of the revision actions.

An application of this revision graph, as mentioned previously in the revision process analysis of a good revised essay, is to study the effect of automated writing feedback using actual revisions made by students at multiple stages, thus helping to find effective forms of feedback leading to revisions. This way of evaluating the effectiveness of Learning Analytics applications (automated writing feedback in this case) is thus made possible using Learning Analytics itself (tracking the revision process in student drafts for detailed study). This could be the first step towards studying the contexts in which automated feedback can work better, and other contexts in which other forms of feedback like human feedback are well suited. Further cognitive processes can be studied using think aloud techniques to capture the mental models while adopting/ rejecting the feedback. We do not yet know if these techniques can be used to differentiate texts that are not extreme cases of performance; thus, having demonstrated the utility of the revision graph in principle, to test its performance on text corpora at scale requires software implementation. Finally, to extend their usage in educational contexts, further work has to be done to characterize essays based on the discussed features to provide meaningful feedback to educators and students. The feedback might be based on writing patterns that emerge or revision types, e.g. to draw attention to the fact that there have been no substantive changes in graphs after 2 drafts or within a defined time interval, or changes that only involve surface level error corrections. Validation of the Revision Graph in terms of usability and usefulness should also be conducted as to supports its application in future writing research.

References

1. Fitzgerald J: Research on revision in writing. Review of educational research, 57(4): 481-506 (1987).
2. Flower L and Hayes J R: A cognitive process theory of writing. College composition and communication, 32(4): 365-387 (1981).
3. Buckingham Shum S, Knight S, McNamara D, Allen L, Bektik D and Crossley S: Critical perspectives on writing analytics. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, ACM, pp. 481-483, NewYork (2016).
4. Shibani A, Knight S and Shum S B: Understanding Students' Revisions in Writing: From Word Counts to the Revision Graph. Technical Report CIC-TR-2018-01, Connected Intelligence Centre, University of Technology Sydney, Australia, https://utscic.edu.au/research/publications/ (2018).
5. Caporossi G and Leblay C: Online writing data representation: a graph theory approach. In International Symposium on Intelligent Data Analysis, Springer, pp. 80-89 (2011).
6. Wininger M: Measuring the evolution of a revised document. Journal of Writing Research, 6(1) (2014).
7. Shibani A, Knight S, Buckingham Shum S and Ryan P: Design and Implementation of a Pedagogic Intervention Using Writing Analytics. In 25th International Conference on Computers in Education, Asia-Pacific Society for Computers in Education, New Zealand (2017).
8. Rijlaarsdam G, Couzijn M and Van Den Bergh H: The study of revision as a writing process and as a learning-to-write process. Revision Cognitive and Instructional Processes. Springer, pp. 189-207, New York (2004).
9. Shibani A: AWA-Tutor: A Platform to Ground Automated Writing Feedback in Robust Learning Design. In 8th International Conference on Learning Analytics & Knowledge (LAK18), Sydney (2018).

[i] An extended version of this paper can be found at Technical Report [4].