Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 363–368
Marseille, 11–16 May 2020

A Process-oriented Dataset of Revisions during Writing

Rianne Conijn[1,4], Emily Dux Speltz[2], Menno van Zaanen[3], Luuk van Waes[4], and Evgeny Chukharev-Hudilainen[2]
1. Department of Cognitive Science and Artificial Intelligence, Tilburg University, The Netherlands
2. Department of English, Iowa State University, United States
3. South African Centre for Digital Language Resources, North-West University, South Africa
4. Department of Management, University of Antwerp, Belgium
m.a.conijn@uvt.nl, {endux, evgeny}@iastate.edu, menno.vanzaanen@nwu.ac.za, luuk.vanwaes@uantwerpen.be

Abstract

Revision plays a major role in writing and the analysis of writing processes. Revisions can be analyzed using a product-oriented approach (focusing on a finished product, the text that has been produced) or a process-oriented approach (focusing on the process that the writer followed to generate this product). Although several language resources exist for the product-oriented approach to revisions, there are hardly any resources available yet for an in-depth analysis of the process of revisions. Therefore, we provide an extensive dataset on revisions made during writing (accessible via `hdl.handle.net/10411/VBDYGX`). This dataset is based on keystroke data and eye tracking data of 65 students from a variety of backgrounds (undergraduate and graduate English as a first language and English as a second language students) and a variety of tasks (argumentative text and academic abstract). In total, 7,120 revisions were identified in the dataset. For each revision, 18 features have been manually annotated and 31 features have been automatically extracted. As a case study, we show two potential use cases of the dataset. In addition, future uses of the dataset are described.

Keywords: Writing analytics, writing process, keystroke logging, revision

1. Introduction

Revision plays a major role in writing (Flower and Hayes, 1980; Scardamalia and Bereiter, 1983; Fitzgerald, 1987). Revisions are defined as: "changes at any point in the writing process" (Fitzgerald, 1987, p. 484). Hence, revisions do not necessarily have to correct an *error*, but they can be any change in the text produced so far. This broad definition results in a large diversity of types of revisions, ranging from the revision of a typo to a major restructuring of the text. Each type of revision can have a different effect on the written product or writing quality (Fitzgerald, 1987; Barkaoui, 2016). Likewise, different backgrounds of writers and different tasks can result in different types of revisions, and different ways in which these are made. For example, spelling and grammar revisions are made more often by second language (L2) compared to first language (L1) writers (Stevenson et al., 2006). Therefore, it is important to be able to analyze revisions separately and in depth.

In writing research, revisions have been analyzed using a product-oriented and a process-oriented approach (Lindgren and Sullivan, 2006b). With the product-oriented approach, revisions are identified by comparing the differences between two products, such as two drafts, see e.g., Min (2006). The process-oriented approach allows for more in-depth analysis, as revisions can also be analyzed within a single draft, e.g., Zhu et al. (2019). Keystroke logging is often used for the process-oriented approach to analyze revisions. With keystroke logging, every key pressed is recorded, resulting in fine-grained information on when and where keys are inserted and deleted, enabling researchers to identify when and where a revision is made (Leijten and Van Waes, 2013; Lindgren et al., 2019). However, revisions within the keystroke log are usually operationalized rather mechanically, for example, by counting the number of backspaces, e.g., Zhu et al. (2019). Although this operationalization is quick and can be done completely automatically, it does not allow for more in-depth analyses of revisions in writing.

Several language resources are available for the product-oriented approach, such as datasets of Wikipedia revisions (Daxenberger and Gurevych, 2012). For an in-depth process-oriented approach, however, hardly any language resources, and especially no open-source datasets, have been made available yet. An exception is that some keystroke logging programs provide additional analyses on revisions (e.g., Inputlog (Leijten and Van Waes, 2013) or CyWrite (Chukharev-Hudilainen, 2019)), but this is still limited to a few properties of revision.

Therefore, the current article describes an annotated dataset which consists of an extensive set of features of revisions made during the writing process of a single draft. This dataset is based on keystroke data and eye tracking data from writers with various backgrounds (L1 versus L2, undergraduate versus graduate) writing a variety of tasks (academic abstract versus argumentative text), resulting in a diverse set of types of revisions. In total, 49 features related to revisions are extracted, consisting of both manually annotated and automatically extracted (rule-based) features. The features relate to eight of the ten properties of revisions, as described in our revision tagset (Conijn et al., 2020): orientation, evaluation, action, linguistic domain, spatial location, temporal location, duration, and sequencing. Two properties, processing and trigger, were not extracted, as these were not available in the keystroke and eye tracking data. The extracted features can be used to study revisions in more depth. In the following, we provide a description of the creation of the dataset and illustrate the usefulness of the dataset with two use cases: (1) an example experiment and (2) revision visualization based on the data.

2. Description of the Dataset

The current dataset is based on anonymized data obtained through the use of CyWrite, a web-based word processing tool with embedded keystroke logging and eye-tracking capabilities. The CyWrite tool has been used both in research studies (Ranalli et al., 2018; Chukharev-Hudilainen, 2019; Chukharev-Hudilainen et al., 2019) and in various undergraduate and graduate courses taught at a large Midwestern research university in the United States. The CyWrite tool provides a composition interface similar to a low-feature text editor (e.g., Microsoft WordPad), while also collecting timestamped logs of keystrokes and eye fixations during the composition process. All writing sessions conducted in CyWrite are automatically stored in a database (without any personally identifying information). From this database, we semi-randomly selected a subset of 65 writing sessions, from which 20 were completed by English-native graduate students writing a summary of an academic article; 20 were completed by native-speaking undergraduate students writing an essay to an argumentative prompt adapted from the Test of English as a Foreign Language (TOEFL), arguing the power of music to influence and entertain people or whether computer technology is a barrier to developing real friendships; and 25 were completed by non-native speakers of English (most likely undergraduate students based on the original study that contributed to this portion of the dataset) writing to similar prompts.

Feature	Mean	SD	IRR
General			
Revision [Y/N]	91.9%	3.7%	0.96
Position of revision end[a]			0.74
Orientation			
Surface	92.6%	6.2%	0.64
....Typography	50.8%	13.0%	0.71
....Capitalization	1.6%	2.2%	.
....Punctuation	6.4%	3.5%	.
....Spelling	2.6%	3.1%	0.74
....Grammar	9.0%	4.9%	0.69
....Cosmetics/presentation	0.2%	0.6%	0.83
....No change	7.4%	3.8%	.
....Wording/phrasing	21.0%	10.9%	0.75
Semantic (deep)	13.9%	8.6%	0.59
Deep specify[b]			0.22
..Microstructure changes	14.1%	8.6%	.
....Supporting info	6.9%	4.8%	.
....Emphasis	2.0%	2.3%	.
....Understate	0.8%	1.1%	.
....Coherence	1.4%	2.0%	.
....Cohesiveness	0.4%	0.8%	.
....unknown	2.6%	2.9%	.
..Macrostructure changes	0.0%	0.2%	.
....Overall aim	0.0%	0.0%	.
....Subtopic	0.0%	0.2%	.
Evaluation			
Correct start	4.7%	4.2%	0.69
Correct revision	85.2%	9.4%	0.66
Action			
Insertion	40.0%	15.5%	.
Deletion	25.2%	7.6%	.
Substitution	24.4%	10.3%	.
Reordering	3.0%	2.7%	.
Domain			
Domain specify[b]			0.59
Subword	67.7%	11.6%	.
Word	24.1%	9.2%	.
Phrase	4.6%	4.0%	.
Clause	1.3%	1.5%	.
Sentence	2.2%	3.1%	.
Paragraph	0.0%	0.2%	.
Number of backspace/delete keys	2.4	0.5	.
Number of characters deleted	3.5	1.4	.
Number of characters inserted	8.2	8.7	.
Number of words deleted	1.1	0.2	.
Number of words inserted	1.7	1.5	.
Number of sentences deleted	0.02	0.02	.
Number of sentences inserted	0.04	0.09	.
Spatial location			
Word finished [Y/N]	51.0%	12.0%	0.70
Intended word[a]			0.71
Word initial	43.1%	10.9%	0.80
Clause initial	13.7%	6.7%	0.68
Sentence initial	10.2%	6.3%	0.82
Characters from leading edge	69.3	91.4	.
Words from leading edge	11.6	15.7	.
Pre-contextual (= 1 - contextual)	77.9%	17.1%	.
Immediate (= 1 - distant)	86.2%	10.1%	.
Chars from start sentence	69.2	31.7	.
Chars from start writing process	813.7	348.5	.
Chars from start writing product	817.2	344.5	.
Temporal location			
Time from start writing process	8.5	4.0	.
Duration			
Duration (sec)	3.1	3.0	.
Pause before revision (sec)	2.0	1.1	.
Sequencing			
Overrides previous revision	13.8%	7.6%	0.55
Continues on previous revision	14.6%	8.3%	0.27
Repetitive (leading edge)	23.9%	10.1%	.
Repetitive (immediate)	23.8%	10.4%	.
Embedded revision (lead edge)	0.2%	0.5%	.
Embedded revision (imm)	0.2%	0.5%	.
Seq forwards (leading edge)	8.1%	8.3%	.
Seq forwards (immediate)	4.8%	5.2%	.
Seq backwards (leading edge)	1.4%	2.1%	.
Seq backwards (immediate)	1.2%	1.8%	.
Time from prev revision (sec)	6.7	3.4	.
Chars from prev revision	7.0	11.1	.

Table 1: Descriptive statistics of all features and inter-rater reliability (Krippendorff's alpha) of the manually annotated features. Adapted from Conijn et al. (2020).

Notes. [a] Non-numerical variable so no descriptives are provided. [b] Inter-rater reliability is calculated once for the full category, as all labels are mutually exclusive. The . indicates no inter-rater reliability, as feature is automatically extracted. N =7,120.

Within this dataset, we automatically extracted revision events from the writing-process data. A revision event starts when the writer starts deleting character(s), or when the writer repositions the cursor to a different location and starts inserting character(s) there. A revision event ends when a new revision event is started or when the writer continues producing new text once the revision is finished (the continuation of text production was manually annotated as human judgement was required). In total, 7,120 revision events were identified ($M = 110$, $SD = 53$ per student).

Each of these revision events were annotated according to a range of features, detailing the properties of the revision event. The features are related to the orientation, evaluation, action, linguistic domain, spatial location, temporal location, duration, and sequencing of the revision.

Those features that required human judgement were manually annotated, while others were extracted automatically from the process and product data using rule-based algorithms. Table 1 provides an overview of all features in the dataset. Only for the manual features the inter-rater reliability is provided (for the automatic extracted features, inter-rater reliability is not available and indicated with a dot ('.'). Hence, the manual features in the dataset are all features in the dataset which have a value for the inter-rater reliability. The dataset is available from `https://hdl.handle.net/10411/VBDYGX`. Below we shortly summarize the feature extraction procedure; a detailed description can be found in Conijn et al. (2020).

2.1. Manual Annotation

The manual features were annotated by five annotators. Annotation was done after extensive training and using a detailed annotation guide. The annotation guide can be found in the supplementary materials of Conijn et al. (2020). For every revision event, the number of characters inserted and removed was provided, as well as a visual replay of the typing process, including the eye fixation marker. In total, 15 (23%) documents were randomly selected to be annotated twice (by different pairs of annotators). The inter-rater reliability of these documents were estimated using Krippendorff's alpha (Hayes and Krippendorff, 2007; Krippendorff, 2011).

For each revision event, the annotators indicated whether it was indeed a revision or whether it involved merely fluent text production. A revision was indicated when the writer deleted or substituted one or more characters in the text, or when the writer moves the cursor to a different location in the text and then begins producing new text. Fluent text production was indicated when the writer solely produced text at the leading edge. The leading edge was defined as proposed by Lindgren et al. (2019) as the end of the text, but disregarding trailing white spaces or remains of a writing plan (e.g., the word 'conclusion' as a reminder that the conclusion still needs to be written).

If the revision event was flagged as a revision, the annotators marked the point on the timeline where the revision ended and fluent text production resumed. In addition, 18 features were annotated manually. For orientation, the annotators first determined whether it was a surface revision, i.e., conventional copy-editing operations or paraphrasing, or a semantic revision, i.e., one that changed the meaning of the text (Faigley and Witte, 1981). For surface revisions, the annotators also indicated whether it involved typography (slip of the finger), spelling, grammar, cosmetics/presentation, or wording/phrasing (all binary categories). For each semantic revision, the annotators indicated whether it involved changes to provide or delete supporting information (such as examples); emphasize findings or a line of reasoning; understate findings or a line of reasoning; adjust the coherence; adjust the cohesiveness or flow of the text; adjust the overall aim; adjust a subtopic; or unknown (categorical variable). When the specific orientation was unclear, all possible orientations were marked. For typography, spelling, and grammar revisions, the evaluation was annotated, indicating whether the revision start and revision end was correct (Wobbrock and Myers, 2006). In addition, the annotators indicated the domain at which the revision was targeted: subword (i.e., part of a word, such as a morpheme), word, phrase, clause, sentence, or paragraph (Monahan, 1984). For spatial location, the annotators indicated if the word in which the revision started was finished, and if not, what the intended word was (open text entry). Moreover, they indicated whether the characters deleted or inserted were at a word-, clause-, and/or sentence-initial position. Lastly, the annotators indicated the sequencing, or the relation of the revision to the previous revision, whether it overrides the previous revision (e.g., repetitive), or whether it continues on the previous revision (e.g., is caused by) (Kollberg, 1996; Lindgren and Sullivan, 2006a).

2.2. Automatic Extraction

In total, 31 features were automatically extracted using rule-based scripts in JavaScript and R. For orientation, three surface revisions were automatically identified: capitalization, punctuation, and no-change (where some characters were replaced by the same characters). Action was automatically classified into insertion, deletion, substitution, and reordering (Sommers, 1980), using the restricted Damerau-Levenshtein distance (Boytsov, 2011) of the deleted and inserted text. Complementary to the manually annotated domain, we automatically extracted several features related to the size of the revision: number of backspace/delete keys and number of characters, words, and sentences deleted and inserted. For spatial location, we extracted the number of characters or words from the leading edge of the text produced so far (Lindgren and Sullivan, 2006a; Lindgren and Sullivan, 2006b). In addition, we classified a revision as pre-contextual if it was made at the leading edge, or as immediate, if it was made at the cursor position (Thorson, 2000). Moreover, we identified the number of characters from the start of the sentence and the start of the writing process and writing product. For temporal location and duration, we extracted the time from the start of the writing process (Zhang et al., 2016), the duration of the revision event (Xu, 2018), and the pause time (inter-keystroke interval) before the revision event. Lastly, for sequencing we identified whether the revision was at the same location and domain as the previous revision (repetitive), within the domain of the previous revision (embedded), or part of a sequence of revisions in the text (sequence forward or sequence backwards) (Kollberg, 1996; Lindgren and Sullivan, 2006a). All these sequencing variables were calculated for both using the leading edge as well as the point of inscription. In addition, we calculated the number of characters and time from the previous revision.

3. Use Cases

As a case study, showing the potential uses of the dataset, we employed the dataset for two purposes: (1) determining differences between groups of writers and (2) visualizing revision processes.

3.1. Determining Differences

As an experiment, we used the dataset to identify the differences in the types of revision between the three groups in our dataset: undergraduate and graduate L1 writers and undergraduate English L2 writers. The in-depth analysis of revision as opposed to solely counting the number of revisions is important here for two reasons. First, it provides a better understanding of how these groups differ in their revision processes. Second, the simple approach has resulted in conflicting findings, e.g., on the effect of expertise on the total number of revisions (compare e.g., Lindgren and Sullivan (2006a; Barkaoui (2016; Stevenson et al. (2006)). Hence, to determine the effect of expertise on revisions, it is necessary to distinguish between different types of revisions.

Previous work has already shown differences in the orientation, spatial location, and domain of the revision between L1 and L2 writers and between novice and expert writers. L2 students make more spelling and grammar revisions, more revisions at the point of transcription, and more subword revisions, compared to L1 students (Stevenson et al., 2006). Less skilled L2 students make more typographic, spelling, and grammar revisions, make more revisions at the point of transcription, and delete fewer characters, compared to more skilled L2 students (Barkaoui, 2016; Xu, 2018).

In the current study, ANOVAs showed no differences for the orientation of the revision between undergraduate L2, undergraduate L1, and graduate L1 students: no differences were found between the number of typo revisions or the number of language revisions (all $p's > 0.10$). For the spatial location, again no differences were found for the number of immediate and the number of distant revisions between the three groups (all $p's > 0.05$). When considering only immediate revisions without the typo revisions, differences were found ($F(2,62) = 3.55$, $p = 0.04$, $\eta^2 = 0.10$): undergraduate L1 students made fewer immediate revisions (excluding typo revisions; $M = 13.9$, $SD = 6.3$ per 100 words) compared to graduate L1 students ($M = 22.4$, $SD = 15.4$ per 100 words). No differences were found between the L1 and L2 students for these revisions. Lastly, for the domain, again no differences were found between the three groups in terms of number of revisions below word level with and without typos, number of revisions below clause level, and number of characters deleted and inserted (all $p's > 0.10$).

To conclude, the undergraduate L2, undergraduate L1, and graduate L1 students showed limited differences in the orientation, spatial location, and domain of the revision. These findings contradict previous studies, which did find differences in terms of orientation, spatial location, and domain (Stevenson et al., 2006; Xu, 2018; Barkaoui, 2016). The contradicting findings might be explained by the differences in groups. For example, Stevenson et al. (2006) compared students with low and high proficiency in L1, while we considered undergraduate and graduate students, which do not necessarily have to differ in their proficiency. In addition, the sample size might have been too small to show these effects. Nevertheless, this proof of concept shows how the tagset can be used to determine differences in revisions across groups of writers. Future work could further determine how these groups differ in terms of their revisions, potentially by adding or combining features, such as the temporal location (or writing phase).

3.2. Visualizing Revision Processes

Additionally, we used the dataset to visualize the revision processes of the students. Data visualizations are often used in the area of learning analytics, to provide teachers, students, and other educational stakeholders with insight into students' learning processes (Verbert et al., 2013; Verbert et al., 2014). Within this area, these data visualizations of tracked learning activities are also known as dashboards (Verbert et al., 2013; Verbert et al., 2014).

For the current use case, we show how a data visualization may be used to identify differences in the revision processes between two students. This might be used for students to reflect on their own writing process. In addition, this might be used for teachers to identify differences in approaches between students or identify points were students struggle. The visualization is based on three properties: the temporal location, spatial location, and orientation. The visualization of two students is shown in Figure 1. These students show highly distinct processes: student 33 (top) first makes many surface and wording changes, and then for the last five minutes, revises from the beginning of the text to the end of the text with both surface and semantic changes. Student 44 (bottom) misses this final "revision" stage and solely shows one linear process (as shown as a line in the figure) of text production with mostly surface changes and some wording and semantic changes.

To conclude, the dataset can be used to visualize various aspects of the revision process, which may be used by both teachers and students to improve teaching and learning of writing. Future work should identify whether these visualizations indeed have the proposed effects, for example by evaluating the visualizations in the classroom context, cf. Martinez-Maldonado et al. (2015).

4. Discussion

In this article, we provide an annotated dataset detailing features related to revisions made during the writing process of a single draft. The dataset consists of both manually annotated and automatically extracted features, based on data from keystroke logging and eye tracking. The manual features resulted in relatively high inter-rater agreements

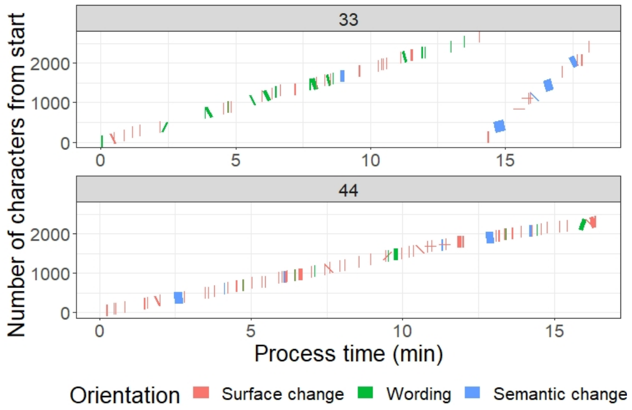

Figure 1: Visualization of the orientation, spatial location, and temporal location of the revisions of two students.

(Krippendorff's alpha ranged from 0.59–0.96), except for the deep revisions and sequencing of revision. This might be caused by the limited number of deep revisions in the dataset. In addition, we provided several automatic features related to sequencing to complement the manual annotations.

4.1. Future Work

The use cases pointed to potential uses of the dataset. We have shown how the dataset could be used for determining differences between various groups of writers. Moreover, the dataset could be used for visualizing the writing processes, which may be used by teachers and learners to reflect on and improve students' revision processes. In addition to the provided use cases and possible extensions, this dataset might also be used for other purposes.

First, given the chronological ordering of the revision events, sequential analyses or pattern mining may be employed to identify patterns in the revision processes over time (e.g., for pauses Zhang et al. (2016)). This moves beyond the relatively simple patterns as identified in the sequencing category which only compared two subsequent revisions.

Second, future work could identify the extent to which the manual features can be classified using machine learning techniques, cf. Zhang and Litman (2015). In addition, the dataset could be used to determine how the classification generalizes across the different subsets in the dataset. For example, whether algorithms trained on L1 data also predict well on L2 data. In this way, the labor-intensive manual annotation of the revision properties might be replaced by automatic classification in the future.

Third, the current visualizations only compared two students. Future work could try to identify whether there are some clusters or groups of students with similar revision processes or writing profiles, cf. Levy and Ransdell (1996), Van Waes and Schellens (2003). In combination with the visualizations (especially related to the temporal and spatial location), this could provide writing researchers with further insight how writing processes might differ (but also how they might be similar).

Finally, to further encourage the availability of language resources allowing for a process-oriented approach to analyzing writing, the authors would like to encourage other researchers to share their writing process data. Moreover, to increase the accessibility of these data and their use, it would be advisable to further develop a standardized XML format for writing process log files. This kind of format would simplify the interchangeability of research data, independent of the logging tool used. A first step towards standardization is formulated in a white paper (Van Horenbeeck et al., 2015). This document describes a generic structure for logging human computer interaction and related XML-tagging in relation to three components: (1) data related to the session, including logger used, time start, and information on the writer, (2) summary statistics of the session, such as writing product, number of words (optional), and (3) logged events, including action (e.g., revision), properties of the action (at a minimum the position and timing of the action), and result of the action (e.g., characters inserted/deleted).

5. Conclusion

To conclude, we provide the first open dataset on revision during the writing process (as opposed to the writing product). This dataset includes an extensive amount of features. We expect that this dataset will enable other researchers to study revision in writing in more depth.

6. Bibliographical References

Barkaoui, K. (2016). What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills. *The Modern Language Journal*, 100(1):320–340.

Boytsov, L. (2011). Indexing methods for approximate dictionary searching: Comparative analysis. *Journal of Experimental Algorithmics (JEA)*, 16:1–91.

Chukharev-Hudilainen, E., Saricaoglu, A., Torrance, M., and Feng, H.-H. (2019). Combined deployable keystroke logging and eyetracking for investigating L2 writing fluency. *Studies in Second Language Acquisition*, 41(3):583–604.

Chukharev-Hudilainen, E. (2019). Empowering automated writing evaluation with keystroke logging. In E. Lindgren et al., editors, *Observing Writing*, pages 125–142. Brill.

Conijn, R., Dux Speltz, E., Zaanen, M. v., Van Waes, L., and Chukharev-Hudilainen, E. (2020). A product and process oriented tagset for revisions in writing. *PsyArXiv*. `psyarxiv.com/h25ak`.

Daxenberger, J. and Gurevych, I. (2012). A corpus-based study of edit categories in featured and non-featured Wikipedia articles. *Proceedings of COLING 2012*, pages 711–726.

Faigley, L. and Witte, S. (1981). Analyzing revision. *College composition and communication*, 32(4):400–414.

Fitzgerald, J. (1987). Research on revision in writing. *Review of educational research*, 57(4):481–506.

Flower, L. and Hayes, J. R. (1980). The cognition of discovery: Defining a rhetorical problem. *College composition and communication*, 31(1):21–32.

Hayes, A. F. and Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1):77–89.

Kollberg, P. (1996). S-notation as a tool for analysing the episodic structure of revisions. In *European writing conferences*.

Krippendorff, K. (2011). Computing Krippendorff's alpha-reliability. *Departmental Papers (ASC)*, 43:1–10.

Leijten, M. and Van Waes, L. (2013). Keystroke logging in writing research: Using inputlog to analyze and visualize writing processes. *Written Communication*, 30(3):358–392.

Levy, C. M. and Ransdell, S. (1996). Writing signatures. In *The science of writing: Theories, methods, individual differences and applications*. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, US.

Lindgren, E. and Sullivan, K. P., (2006a). *Analysing online revision*, pages 157–188. Elsevier.

Lindgren, E. and Sullivan, K. P., (2006b). *Writing and the analysis of revision: An overview*, pages 31–40. Elsevier.

Lindgren, E., Westum, A., Outakoski, H., and Sullivan, K. P. H. (2019). Revising at the leading edge: Shaping ideas or clearing up noise. In *Observing Writing*, pages 346–365. Brill, Leiden, The Netherlands.

Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., and Clayphan, A. (2015). LATUX: An iterative workflow for designing, validating and deploying learning analytics visualisations. *Journal of Learning Analytics*, 2(3):9–39.

Min, H.-T. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of second language writing*, 15(2):118–141.

Monahan, B. D. (1984). Revision strategies of basic and competent writers as they write for different audiences. *Research in the Teaching of English*, pages 288–304.

Ranalli, J. R., Feng, H.-H., and Chukharev-Hudilainen, E. (2018). The affordances of process-tracing technologies for supporting L2 writing instruction. *Language Learning & Technology*, 23(2):1–11.

Scardamalia, M. and Bereiter, C. (1983). The development of evaluative, diagnostic and remedial capabilities in children's composing. In M. Martlew, editor, *The psychology of written language. Developmental and educational perspectives*, pages 67–95. New York: McMillan.

Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College composition and communication*, 31(4):378–388.

Stevenson, M., Schoonen, R., and de Glopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing*, 15(3):201–233.

Thorson, H. (2000). Using the computer to compare foreign and native language writing processes: A statistical and case study approach. *The Modern Language Journal*, 84(2):155–170.

Van Horenbeeck, E., Pauwaert, T., Van Waes, L., and Leijten, M. (2015). A generic XML-structure for logging human computer interaction [white paper]. Retrieved from http://www.inputlog.net/wp-content/uploads/Generic_XML_structure_version-1_3.pdf.

Van Waes, L. and Schellens, P. J. (2003). Writing profiles: The effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of pragmatics*, 35(6):829–853.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., and Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10):1500–1509.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Assche, F., Parra, G., and Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, 18(6):1499–1514.

Wobbrock, J. O. and Myers, B. A. (2006). Analyzing the input stream for character-level errors in unconstrained text entry evaluations. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(4):458–489.

Xu, C. (2018). Understanding online revisions in L2 writing: A computer keystroke-log perspective. *System*, 78:104–114.

Zhang, F. and Litman, D. (2015). Annotation and classification of argumentative writing revisions. In *Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 133–143.

Zhang, M., Hao, J., Li, C., and Deane, P. (2016). Classification of writing patterns using keystroke logs. In L. Andries van der Ark, et al., editors, *Quantitative Psychology Research: The 80th Annual Meeting of the Psychometric Society, Beijing, 2015*, pages 299–314. Springer International Publishing.

Zhu, M., Zhang, M., and Deane, P. (2019). Analysis of keystroke sequences in writing logs. *ETS Research Report Series*.