

Book of Abstracts

THE EIGHTH INTERNATIONAL CONFERENCE ON WRITING ANALYTICS

Academic Writing in Digital Contexts: Analytics, Tools, Mediality

5-6 September 2019, Winterthur, Switzerland.

<https://writinganalytics.zhaw.ch/>

Conference Chairs and Organizing Committee: Christian Rapp, Otto Kruse &
Noah Bubenhofer
Zurich University of Applied Sciences

Papers (alphabetical order)

Author(s)	Title	Abstract
Chris M. Anson, Ian G. Anson	Text Recycling in STEM Disciplines: Results from a Text Analytic Study	Text recycling (TR), sometimes called “self-plagiarism”, is the reuse of text verbatim from one’s own existing documents in a newly-created text. In a (US) federally-funded grant, we have been studying TR patterns using several methodologies. In one strand of

		<p>this research, we have created a tool in R to analyze large corpora of published articles to determine the extent of TR occurrence. In this presentation, we briefly describe the goals of the project and the analytic system we have developed. We then share the results of an analysis of 400 published papers associated with eight disciplines. Results show the frequency of TR across disciplines in articles generated from the same grants. These results demonstrate that substantial variation exists in text recycling practices across disciplines and individual authors. We conclude by speculating about the causes of these patterns, especially as they have evolved over time.</p>
Kalliopi Benetos	Developing meaningful analytics: beyond time-stamps and counts	<p>The development and testing of C SAW, an authoring tool to support the development of argumentative writing skills while fostering critical thinking and self-regulation, it was necessary to identify how the software was being used and the effects of different uses on outcome variables. Data was collected through the recording of scripts activated during use of different devices and used to visualize writing processes and stages for each participant. Transitional probability matrices were crucial to identifying which aids and devices invoked which writing activities and which writing activities led to the solicitation of which aids. This data permitted the visualization of participants writing processes and the identification of problematic or successful use patterns. The study also revealed future development needs in data analytics and visualizations of writing processes in digital environments that can meaningfully represent uses and writing processes for researchers and instructors.</p>
Noah Bubenhofer, Selena Calleri, Klaus Rothenhäusler	Leveraging corpus linguistic methods to provide writing aids	<p>The ThesisWriter tool (https://thesiswriter.zhaw.ch/) provides a component that offers linguistic help to users. We describe the compilation of the corpus and the procedures used to automatically identify typical language use in scientific writing, that build the backbone of this module.</p>
Mădălina Chitez, Roxana Rogobete	The impact of e-feedback on student academic writing: a corpus-informed assessment	<p>The practice of electronic feedback has become popular with both educators and students in the last decade. However, in the Romanian HE context, little is known about the impact of digital technologies on academic writing improvement processes. The present case study examines how a group of 16 Philology students integrate computer-mediated feedback delivered by expert writers into their literary analysis texts. The e-feedback was part of a multi-modal feedback strategy: (a) three stages of writing (summary, first two pages, final draft) followed by written comments via e-mail and Microsoft Office tools; (b) each feedback-revision session was followed by a group discussion and integrated the iThenticate platform. Corpora batches of revised text versions have been compiled and linguistic features correlated with revision categories. We draw conclusions on the effectiveness of the e-feedback integration into student texts, and propose a scenario for didactic applications of e-feedback for academic writing processes.</p>
Mark Cieliebak	A Brief Survey of Natural Language Generation: Technologies and Applications	<p>Natural Language Generation (NLG) is a sub-field of Artificial Intelligence where machines generate new texts. Simple rule-based approaches were already proposed in the 1960's, and have been refined ever since. Nowadays they have reached main stream media, where news articles on sports, weather and business news are generated on a daily basis. On the other hand, recent developments in machine learning enable computers to create new, genuine texts from scratch. A typical approach is to</p>

		<p>output a sequence of characters, one by one, which surprisingly forms reasonable sentences and texts in many cases. This can be applied, for instance, to generate new product descriptions from structured data (i.e. product features), to generate replies of a chatbot to arbitrary user utterances, or to generate questions about scientific or literary texts for online-learning or (pre-)exams. In this talk, we survey the current state-of-the-art in Natural Language Generation. The talk is suitable to a broad audience, and no technical expertise is required.</p>
<p>Rianne Conijn, Menno van Zaanen, Mariëlle Leijten, Luuk van Waes</p>	<p>How to typo? Building a process-based model of typographic error revisions</p>	<p>The analysis of writing is complex, with planning, translating, and reviewing processes interacting in a non-linear fashion. Intuitively negligible activities such as the revisions of typing errors can have a large influence on the writing process, and hence also on the analysis of writing processes. On the one hand, these types of revisions are low-level, and hence less-important, so we would like to filter them. On the other hand, typing errors, and especially the revision of typing errors, can (unwillingly) break the (linear) flow in writing. Therefore, it is important to identify these revisions, to be able to determine their effect on disfluency and activation of other subprocesses. Previous work on typing errors commonly focuses on the writing product rather than the writing process. In this way, revised typing errors are omitted. In addition, no distinction is being made between typographic errors (slips of the fingers) and other types of errors. Therefore, we aim to build a process-based model of typographic errors and their revisions. Three analysis were conducted. First, typographic errors and their revisions were characterized by using data from 2,103 copy tasks. For this, temporal and character bigram properties were extracted from keystroke data. Second, a process-based model on typographic error revisions was trained on the copy task dataset, to try to automatically identify these errors. Lastly, this model was evaluated in a more natural setting: on keystroke data obtained from a regular (source-based) writing task. The characterization showed that typographic errors are made and revised in a variety of ways. However, we do see some generic patterns in character bigram properties and timing of the keystrokes which might be used to model typographic error revisions using process data only. Results on these process-based models indeed showed that it is possible to identify typographic errors using keystroke information only, especially in a copy task. Yet, the models on the source-based writing task still lead to a high number of false positives. Using these models, a more nuanced analysis of fluency and revision in writing can be performed.</p>
<p>Elena Cotos, Aaron Bertram, Angelica Jasper, Lynn Lundy-Evans, Kristin Terrill</p>	<p>Thesis and Dissertation Writing Quality Needs Analysis</p>	<p>This study involved analyzing data from a corpus of published graduate student theses and dissertations in a mid-western U.S. university. Concerns about writing quality in suggest that students would benefit from additional support for these genres. To describe these needs, a text analysis study was conducted that involved collecting a corpus of published theses and dissertations from the 2015 - 2017 cohorts. Results showed that most of the texts in the corpus contained either local, global, or both types of writing quality concerns, and these concerns occur with different frequency across discipline and student level. Conclusions from this needs analysis will inform the development of pedagogical tools for graduate student writers as they edit their theses and dissertations and prepare them for publication. Of particular interest are implications of this study for the development of an</p>

		automated writing evaluation tool that could be implemented in a broader context.
Lieve De Wachter, Margot D'Hertefelt, Serge Verlinde, Geert Peeters, An Laffut, Kristin Blanpain	AWA KU Leuven. Online support during the writing process	The online tool Academic Writing assistant (AWA) developed for Dutch, English and Afrikaans at KU Leuven (Belgium) fits into the increased attention for process-oriented writing support that has led to the development of a new generation of writing assistants. AWA is used by students in consecutive revising phases of their academic texts. It detects possible mistakes, mainly based on string-pattern matching and some NLP techniques, highlighting words and phrases that may need improvement, it offers possible alternatives and it provides suggestions to enrich the text, this way serving the functions of detection and prediction. The immediate feedback given allows users to improve and enrich their text during the writing process. As a web-based application, AWA distinguishes itself from writing assistants available as an app, a software package or an add-in for word processors. AWA belongs to the category of Automated Writing Evaluation tools, with a number of additional features of Intelligent Tutoring Systems, such as more individualised feedback. Because of its easy online accessibility, AWA is currently used both as a stand-alone and self-directed tool and along with in-class instruction.
Curtis Gautschi	Predicting CEFR levels of student essays in placement tests using an Automated Essay Scoring tool in R: a corpus-based approach	This paper describes the design and first results of an Automated Essay Scoring tool and CEFR-level prediction algorithm created and experimentally implemented in an online English placement test in a higher-education setting. The algorithm was developed employing a prediction-accuracy pseudo-black box approach using a small training corpus of texts with known CEFR levels (N=50). Written and run entirely in R, the tool can be integrated with other advanced text analyses possible in R. Efficiently calculating the CEFR levels of large numbers of student essays, the tool is ideal for quick assessment of writing skills for placement purposes. While gold-standard (human) validation evidence is still required, external validation checks demonstrate the accuracy of the tool. Prediction patterns were found to be closer to officially graded CEFR levels of a selection of texts than from other online AES systems. Further research perspectives and dissemination will also be discussed.
Vassiliki Kourbani	Establishing a 'culture of collaboration'; From Writing Center to Virtual Education Hub	The Writing Center (WC) at the Hellenic American University is a breakthrough in what conventional Writing Centers entail, in terms of both target audience and nature. Since its services go beyond the traditional face-to-face approach, it aspires to showcase how collaborative writing is realized in terms of the cooperation among the particular (WC) stakeholders and the learning outcomes involved. The services of the Writing Center go beyond the traditional face-to-face (f2f) 'paper-pencil' approach, through the use of e-equipment and e-learning software (the Blackboard e-learning platform) that encourage individualized online tutoring synchronously and asynchronously, retaining nonetheless a degree of reciprocity and interactivity. The study investigates the connection between WC tutor synchronous feedback and revision by non-native speakers (NNS), who speak English as a foreign language (EFL), at an American university in Greece. Drawing on socio-cultural theories of language and literacy, the study examines tutor-tutee critical moments by linking various features of the interaction during synchronous communication with the revision process to discover ways tutors

		and tutees co-construct their roles within the multimodal WC tutorial.
Susan Lang	Constructing a Functional Taxonomy of Writing Analytics	As the field of Writing Analytics continues to expand and attract increased attention, mapping and understanding the relationship of the various methodologies and stakeholders becomes even more of a critical task. This presentation will describe one attempt to create a taxonomy. In this presentation, presenters will explain the process and results of constructing a model of functional, rather than conceptual, knowledge of the field. We will discuss the following topics: 1) What are the current, critical writing analytics labor areas? 2) What are the research foci of recent published projects? 3) What are the most common research targets and artifacts of research published over the last decade? The answers to the above will help us conduct a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis of the field of writing analytics and discuss with the audience how this analysis will potentially influence the field's future in both North America and Europe.
Anni Jürine, Djuddah A.J. Leijen	Bottom-up vs top-down: Refining the methodology for Compiling an Academic Phrasebank	This study contributes to the field of (teaching) academic writing by suggesting a methodology for compiling academic phrasebanks. A phrasebank is a collection of typical phrases in academic texts and can be used as a pedagogical tool to support students' writing. Using such resources can help students find appropriate rhetorical and linguistic means when writing academic texts (Morley, n.d.). As such, phrasebanks help students to shape their writing, so that it meets the expectations of their disciplinary audiences. The aim of this paper is to validate the methodology for compiling an Academic Phrasebank for Writing in Estonian (Author 1, 2019), which has been compiled using a corpus-based bottom-up approach that relies on frequent N-grams (see Cortes, 2013). This method is efficient as it allows identifying the most typical expressions and their functions across large bodies of text. However, the method has been criticized for being too coarse-grained to identify all relevant functions (Moreno and Swales, 2018). In this study, we compare the functions and expressions in the Introduction section obtained with the bottom-up approach to that of top-down approach. The preliminary results indicate that while there are differences in the expressions obtained with the bottom-up and top-down approach, there is considerable overlap in the functions identified with the two approaches.
Marije Lesterhuis, Nina Vandermeulen, Gert Rijlaarsdam, Elke Van Steendam, Sven De Maeyer	How students write synthesis texts: Clustering process measures	Source-based writing, such as synthesis writing, is highly difficult for students. This study aims at identifying writer profiles based on keystroke logging data as insight into such profiles could help to focus instruction on the current approaches students use. More specifically, we pose the questions to what extent 1) process profiles can be distinguished, and 2) these profiles are linked with the quality of the produced texts. Therefore, we look into the process measures of 2276 texts written by 658 students (age 15 to 18) in the Netherlands. We executed a latent class analysis, focusing on the number of keystrokes, the number of pauses, the duration of pauses, the transitions between sources and from sources to the own document in the first phase of the writing process. The analyses result in different profiles related to text quality. We propose how to use these insights for feedback and instruction.
Johanna Phelps	An Overview of Recent Updates to	This presentation offers an overview of international research ethics policies and a synopsis of some of the major regulatory

	International Policies Impacting Research in Writing Analytics	updates that impact research in writing analytics. Specifically, attendees are offered focused information about the Common Rule updates, implemented in 2019, (the regulations that mandate Institutional Review Board review of human subjects research in the United States) and the General Data Protection Regulations (GDPR) in the European Union, implemented in 2018. If researchers' data collection, storage, and/or analysis occurs in environments governed by these policies, they must ensure they adhere to required standards (e.g. a US-based researcher collecting data from EU schools, or an EU researcher engaging with US-based students). The impact of these updates on the international, multidisciplinary, and multi-method environment of writing analytics requires attention, given the vast impact of these updates on data collection, storage, and analysis.
Daniel K. Schneider, Kalliopi Benetos, Valérie Follonier	Tools and strategies for scaffolding and evaluating student contributions in a MediaWiki	Wikis have interesting affordances for numerous pedagogic strategies, such as writing-to-learn and peer-reviewing, and supporting contribution-oriented pedagogies. MediaWiki technology is fairly popular in higher education. However, few tools leverage its full potential of this technology for education. A particular shortcoming is the scarcity of functional tools that summarize participants' contributions, another is the lack of support for structured writing activities. The present study reports the scaffolding of writing and multimedia productions in a self-study class, using "Semantic Media Wiki" and "Page Forms" technology. Learner productions and progress are presented in the form of tables, graphics or dashboards that can engage participants in reflection with respect to their different goals, roles, tasks, productions, and therefore help coordinating learner activities. We will present and discuss observations and data from a five-year cohort. We will examine advantages and disadvantages of semantic MediaWiki technology and discuss pedagogical effectiveness of our various technology-enhanced learning scenarios.
Sladoljev-Agejev, T., Snajder, J. & Kolic-Vehovec, S.	Computational Analyses of the Effects of a Structure Strategy on College-Level Summaries: Cohesion and Rhetorical Structure	Summary writing is a useful strategy for developing reading and writing. This particularly applies to college-level assignments in a foreign language (FL) which may be demanding both conceptually and language-wise. Text management may then be facilitated by structure strategies leading to more coherent summaries with clearer rhetorical structure, more cohesion, and better text organisation. We present the analyses of about 100 summaries written by junior business students in a quasi-experimental research study which investigated the effects of a structure strategy on the quality of summaries (e.g. coherence, cohesion) written in English as a foreign language (EFL). Positive effects were found using Coh-Metrix for measuring cohesion (Graesser et al., 2004; McNamara and Graesser, 2012) and a tool based on Penn Discourse Treebank (Prasad et al., 2004) for the rhetorical structure.
Cornelia Tschichold	Combining automated feedback with tutor feedback in an academic writing class	The use of grammar checkers is not very wide-spread in academic writing classes despite the fact that direct explicit corrective feedback is known to be helpful. The study presented here shows results from an experiment that aims to explore and integrate the use of ETS's e-rater into regular university pre-sessional writing classes. 70 students on a 12-week course participated; those in the experimental group were taught to use e-rater to help them revise and proof-read their own drafts before submission of two essays that were part of their assessed portfolio. They were later asked to

		fill in a questionnaire on the perceived usefulness of the tool. All students also received the standard amount of tutor feedback on the content, structure and language of their essays. Results showed improvement in the targeted areas of article and agreement errors for both groups, but the experimental group improved more.
Nina Vandermeulen, Elke Van Steendam, Gert Rijlaarsdam, Mariëlle Leijten, Luuk Van Waes, Marije Lesterhuis	The design and effect of two types of process-oriented feedback on synthesis writing	We conducted an intervention study with 67 Dutch students (grade 10) to explore the effects of two types of process-oriented feedback. Participants received a customised process report generated with keystroke logging tool Inputlog. They were encouraged to reflect on their writing. In the position setting feedback condition, students compared their writing process to that of students with a similar text quality score. In the feed-forward feedback condition, they compared their writing process to better scoring students. These exemplary writing processes were selected from a national baseline study with more than 700 Dutch students. We will present the development and implementation of the feedback. Moreover, we will present results on its effectiveness: in the feed forward condition, the intervention was effective: in one week the students made a progress comparable to one year of regular schooling.
Anna Wärnsby, Asko Kauppinen	Instructor Feedback as Data for Curriculum Design in an ESL Academic Writing Course	Moving writing courses into digital environments allows systematic access to data pertaining to student production and instructor assessment. This makes possible – and encourages – new questions and techniques of inquiry of writing patterns, testing of exploratory questions, and teasing out what kind of information such data can be made to yield. This study uses corpus methodology to analyse summative instructor feedback. The data was collected 2014-2016 from an undergraduate course at Malmö University, Sweden, and consists of instructor comments on 328 papers. In total, the corpus comprises approximately 40,000 words. Given the continuity of course settings, instructions and instructors, we treat the student and instructor body as homogenous and representative of ESL contexts for teaching and learning academic writing at tertiary level. By looking at features of affective language in instructor feedback, we explore criticism and praise in summative assessment. We draw from research on affective and cognitive discourse features (cf. Mirador [2014] on describing the genre of teacher comments through linguistic expressions). Similarly to Wärnsby et al. (2018), we investigate boosters (really, indeed), hedges (maybe, perhaps), cognitive verbs (think, believe), adjectives (good, clear), expressions of suggestion (suggest, you better), personal pronouns (I, we, you), and adversative transitions (however, on the other hand). We find patterns of distribution of different features of affective language, and that some, but not all, affective language features in instructor feedback correlate to grades. Furthermore, analysis of affective language reveals that summative feedback frequently incorporates elements of formative feedback, intentionally or unintentionally. This may be neither time-efficient nor beneficial for student learning, and may reflect uncertainty about the role and function of summative feedback. These findings have an impact on curriculum design, in particular for planning instructor interventions and assessment practices.
Roger Yallop, Djuddah Leijen	Coding cover letters used as a substitute for	Many writing instructors use online peer feedback systems (e.g. My Reviewers) to gather research data. These systems allow the instructors to provide reviewers with suitable writing assessment

	writing assessment rubrics in online peer feedback systems	grids to base their written feedback on. As opposed to author self-devised assessment criteria (cover letters), these rubrics restrict the authors' flexibility of devising their own assessment criteria, providing textual background information and using social presence to develop a sense of community within their writing groups. This presentation investigates how twenty matched PhD students use cover letters in lieu of these rubrics over a three-month period. One group received direct cover letter instruction; the other group did not. Their cover letters were analysed for genre and social presence using thematic analysis. The results show the typical contents and use of social presence between the two groups, and the impact of instruction. These findings will help us develop algorithms to code cover letters.
--	--	---

Posters and Demos (alphabetical order)

Author(s)	Title	Abstract
Gerd Bräuer, Ruth Arimond	Promoting reflective practice in the training of teachers using e-portfolios	Between 2016 and 2018, four institutions in international teacher education collaborated in a project with the aim to develop a theoretical and practical framework for writing electronic portfolios. Based on local need analyses and research on digital reflective practice, a reflective writing platform was created and, in the second phase of the project, further developed through learning analytics of task design and student performance. The poster presentation will provide information about design, content and results of the learning analytics procedure.
Otto Kruse, Christian Rapp	A writing platform for thesis writers (Demo)	The revised and extended, second version of <i>Thesis Writer</i> has been released and we will shortly explain its rational and developmental background. Along with the new support tools we will demonstrate the new working space for writers and some new features to record and evaluate user data.
Theresa Kruse, Ulrich Heid	A specialized e-dictionary as a tool for improving academic writing in the field of mathematical graph theory	We plan to create a domain-specific ontology and an electronic dictionary for the mathematical field of graph theory. It is based on two corpora, one in German and one in English, composed of textbooks and scientific papers. We extract terms from the texts and aim at creating the ontology automatically. The purpose of the dictionary is to help mathematics students to improve their academic writing regarding terminology, because they have to deal with two challenges: First, they are writing their theses in German but most of the sources are in English. Secondly, the students have to learn the meaning of the terms in the context of graph theory. To examine the effect of the dictionary and ontology use, student theses will be analyzed with regard to terminology usage by methods of corpus linguistics, from a quantitative and a qualitative perspective.
Marije Lesterhuis, Maarten Goossens, Roos Van Gasse, Sven De Maeyer	D-PAC: Assessing writing in a comparative way (Demo)	To ensure the development of writing skills, assessments are key. Evaluating these texts is problematic, as analytic scoring procedures, which are traditionally used for the rating of writing skills, face problems in reliability and validity. These are caused by, among others, assessors differing in how they interpret and use rating criteria. Therefore, D-PAC introduces a digital tool that uses pairwise comparisons as an alternative scoring method. This method is grounded in "the law of comparative judgment" which states that people are far more reliable in comparing two texts and indicating which text is better, than in making absolute judgments about the quality of a single text. Moreover, by including multiple assessors, who all have their own perspectives of what a quality performance comprises, the validity of this

		method is warranted. Based on the pairwise comparisons of all assessors, the tool provides a rank-order of students' performances. Results from various try-outs indicate that D-PAC is a credible tool.
Kyle Oddis	Write Like a Human: A Model for Using Analytics to Teach "Human" Writing Processes and Practices (Demo)	What does it mean to write as a human? How can machines illuminate our human writing practices and processes so that we can better prepare our students to write in fields that increasingly rely on machines to produce and evaluate content? This session will show how to integrate machine-produced literature and collaborative human/machine texts into writing classrooms across disciplines to help students better understand their writing processes, improve their writing practices, increase genre awareness, and participate in various discourse communities. We will also explore how algorithms can teach students about what we call "good" writing in certain fields. After exploring what makes writing uniquely "human" through a demo of the proposed pedagogical model, participants will receive templates for grafting digital tools onto preexisting writing curriculum and consider the ways in which we can more creatively and proactively teach writing using analytics.
Jolien Strous, Angeniet Kam	TU Write: integrating e-learning in a blended setting (Demo)	At Delft University of Technology, students learn professional and academic writing as a group. Students usually split up the work, leading to a partial fulfilment of the learning goals. TU Write was developed to tackle this problem: e-learning modules used in a blended setting. Students study the modules at home, execute individual writing tasks, and discuss their work during the lectures. Statistics show that TU Write modules connected to individual assignments are visited around four times more on average than other modules. However, this is no guarantee for the achievement of the learning goals. Therefore, we are currently looking for data driven ways to analyse the individual assignments. During this demo, visitors can experience TU Write, its formative exercises and blended setting. The demo is of interest to lecturers who intend to flip their classroom with blended learning. Visitors are welcome to share their experiences with data driven writing analysis.
Caro Struijke	Radboud in'to Languages Tool for Learners of Academic Writing (Demo)	To meet the needs of local and distant students and teachers looking to change their role in the classroom, Radboud in'to Languages have developed a digital repository of learning materials on English academic writing that can be used both in blended courses and by individual learners. This demo will introduce the materials and the NovoLanguage platform it runs on, share the insights we gained during development, and report on our teaching experience with this integrated tool. The tool gives teachers the flexibility to assemble learning tracks geared toward learners' writing level and study field. Learners are presented with theory, which is interspersed with interactive tasks. They receive immediate feedback and explanation on all tasks. Because tasks are based on research articles from their fields, learners receive maximal exposure to authentic examples reflecting their personal writing goals. So far, we have built materials for medical sciences, life sciences, psychology, and linguistics.

