ARTIST: A Learning Support System for Fostering Students' Argumentative Writing Skills

Thomas Huber
University of St. Gallen, Switzerland
thomas.huber@unisg.ch

Christina Niklaus
University of St. Gallen, Switzerland
christina.niklaus@unisg.ch

Abstract

We present ARTIST, a learning support system that can help students assess their argumentative writing and provide automated, individual feedback, thus improving their writing performance. It analyzes student-written argumentative texts by identifying argument components and their relationships. The resulting argumentative discourse structure is displayed in an interactive interface. In that way, the ARTIST tool provides immediate and personalized visual feedback on the quality of students' texts, supporting self-monitoring and reflection on how to improve their texts.

1 Introduction

Argumentative writing skills are essential to enable one to convey one's own understanding and critical thinking. Effort and training are needed to improve them. In many contexts however, writing skills are not promoted and training measures are not very effective (Thaiss and Zawacki, 2006; Stevenson and Phakiti, 2014). However lecturers often do not have time to provide individual feedback to each student. Generic responses hinder their learning progress. This is problematic, as argumentative writing is rarely done outside of schools when one is a student. To address this issue, recent advances in Natural Language Processing (NLP) and Artificial Intelligence (AI) are leveraged to analyze the writing quality of texts and to provide students with personalized and adaptive feedback, as well as to support gains in student's writing motivation and quality (Zhang, 2013; Rapp and Kauf, 2018; Strobl et al., 2019).

Whereas automated support for revisions on the *micro-level*, targeting factual knowledge (e.g., grammar, spelling, word frequencies) is well-represented in current literature, tools that support the development of writing strategies and encourage self-monitoring to improve *macro-level* text quality (e.g. argumentative structures, rhetorical moves) are still rare. Therefore, we propose an AI-enabled learning support system to assess students' argumentative writing and to automate feedback to individual students, thus supporting writing performance. This enables personalized learning. One of the most significant benefits of using AI in education is seen as a support tool for personalized learning and formative feedback (Stone et al., 2016; Zawacki-Richter et al., 2019). ARTIST contributes to this new emerging interdisciplinary research field as recent advances in AI emphasize the importance of better understanding of the human-machine power relationship in learning and problem solving (Wesche and Sonderegger, 2019; Raisamo et al., 2019; Seufert et al., 2020).

We make a video demonstration of ARTIST available at https://youtu.be/fOs2EcWd7fU and release the code at https://github.com/unisg-ics-dsnlp/artist-inlg2025.

2 Interface

ARTIST provides direct and indirect feedback through three main channels: (i) the Argumentation Dashboard, (ii) the Discourse Structure overview, which provides an analysis of the rhetorical structure and coherence relations of the text to help the user identify weaknesses, and lastly (iii) through direct, adaptive Improvement Suggestions. Figure 1 shows the Argumentation Dashboard.

Argumentation Dashboard Claims, major claims and premises are highlighted in different colours directly in the input and presented as a graph, showing the argumentative structure of the text. A detailed view shows how the individual components of each argument connect with each other. A sunburst diagram shows the proportion of how much of the text consists of argumentative components. Coherence and Persuasion scores are presented as a box plot based on the rating

Proceedings of the 18th International Natural Language Generation Conference: System Demonstrations, pages 10–13
October 29–November 2, 2025.

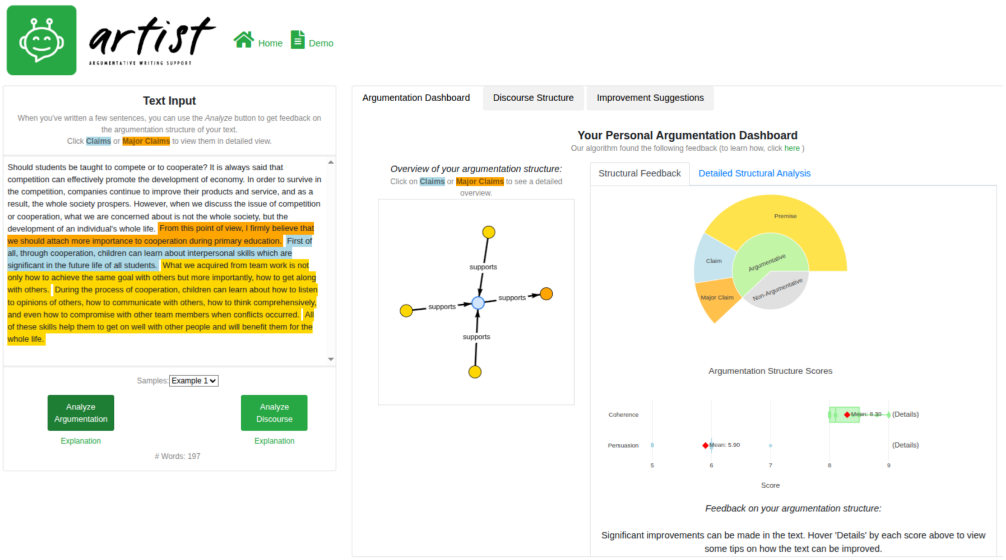

Figure 1: Screenshot of the ARTIST writing support system. A highlighted argument is shown, as well as the structure graph, the sunburst diagram of the distribution of components and the Coherence and Persuasion scores.

of multiple LLM raters following the approach introduced by Hu et al. (2024). To simulate a panel of experts the model is prompted 50 times with a temperature of 0.7. Scores are presented as a boxplot to provide feedback about the consistency of the scores to the user. We use plotly.js for this plot.

Discourse Structure The Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) parse tree of the input text is shown, with explanations for the more high-level discourse markers. The RST tree shows the text split into discourse units and how they relate to each other. By default a simplified version using the high level markers by Fraser (1996) is shown. The complex markers are mapped to *Elaborative*, *Inferential*, *Contrastive* and *Temporal*. Experienced users can switch to ‘Expert Mode‘ to see fine-grained labels. Users can select discourse units to highlight the relation in the text. Explanations for the labels are shown next to the graph. Figure 2 shows the Discourse Structure functionality.

Improvement Suggestions The user can request adaptive improvement suggestions for their text. These suggestions are made by an LLM, and adapt to the user's input.

3 Implementation Details

Backend The backend of ARTIST is a Python Django project.

LLM ARTIST supports using self-hosted LLMs. We use a Llama 3.3 70B instance running on 8 V100 GPUs. We want to emphasize that smaller models, with lower hardware requirements, are suitable alternatives. This includes small models like the Phi family of models, which are designed to be hardware efficient (Abdin et al., 2024a,b), and can be run locally on current consumer grade laptops. The LLM is used for the improvement suggestions feature, as well as to calculate the Coherence and Persuasion scores.

Visualization We use vis.js for the visualization of the argument structure, Plotly.js for the structural feedback, Cytoscape.js for the RST tree.

Discourse Structure Detection We use an updated version of RST parser by Feng and Hirst (2014b,a). The parser itself is unchanged, but we provide a Python package to make it easier to use. The updated package is available at `https://github.com/ThHuberResearch/feng-hirst-rst-parser`.

4 Evaluation

We evaluated successive prototypes of our argumentation feedback system through a series of controlled laboratory experiments and real-world classroom studies, demonstrating its effectiveness in improving students' argumentative skills. For instance, in a study with first-year students (n=80), we observed measurable gains in argumentation competency (Burkhard et al., 2023). In a comple-

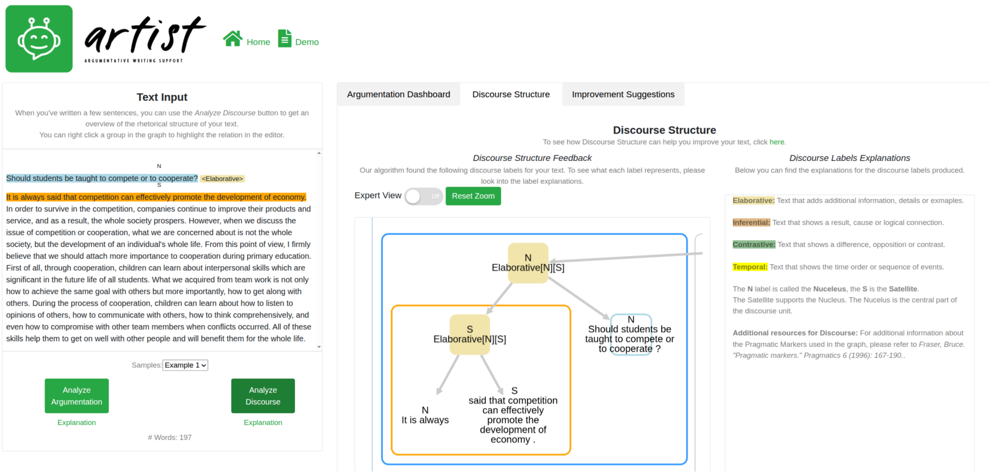

Figure 2: Screenshot of the Discourse Structure functionality of ARTIST. A subgraph of the discourse tree is highlighted, and the relation is shown in the argument on the left. The labels are explained to the right of the graph.

mentary study with 30 participants, we collected qualitative feedback on the tool's usability and perceived effectiveness (Htaw et al., 2024). Moreover, students (n=63) rated the quality of the feedback provided by open-source and proprietary LLMs positively. More precisely, they regarded the suggestions for improving their argumentative texts as helpful (7.51 vs. 7.65 on a 10-point Likert scale) (Gubelmann et al., 2024). Most importantly, students wrote more convincing essays with higher formal argument quality, producing on average 5.1 arguments with our tool compared to 3.2 with a baseline scripting tool (Wambsganss et al., 2020).

5 Example User Interaction

In the following, we describe a typical use case scenario of the ARTIST tool.

The user enters an argumentative text. They click *Analyze Argumentation*. This highlights their argument components, and shows their relationships in the dashboard, which helps find unsubstantiated claims in the text. The *Structural Feedback* sunburst diagram shows that a large portion of the text is non-argumentative. The Coherence and Persuasion scores are also rather low. Next, the user presses *Analyze Discourse*. This generates an RST parse tree, which shows the individual discourse units and their relation. The user is experienced, so they toggle *Expert View*, which provides fine-grained labels. They right-click a subgraph with an *Explanation* and it is highlighted in the text. The user realizes that a part of their text, intended to explain a certain point they were making, does not have this relation in the graph. They read the corresponding passage and note they did not properly elaborate their point. They revise the sentence and generate the graph again. The user analyzes the new graph and is satisfied. Lastly, they open the *Improvement Suggestions* tab, and request individual feedback. The feedback suggests to add concrete examples or evidence to further strengthen the argument. Based on the analyses provided by the tool the user improves their argument further. As the user keeps working with the tool, their overall argumentation skills improve.

References

Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, Alon Benhaim, Misha Bilenko, Johan Bjorck, Sébastien Bubeck, Martin Cai, Qin Cai, Vishrav Chaudhary, Dong Chen, Dongdong Chen, and 110 others. 2024a. Phi-3 technical report: A highly capable language model locally on your phone. *Preprint*, arXiv:2404.14219.

Marah Abdin, Jyoti Aneja, Harkirat Behl, Sébastien Bubeck, Ronen Eldan, Suriya Gunasekar, Michael Harrison, Russell J Hewett, Mojan Javaheripi, Piero Kauffmann, and 1 others. 2024b. Phi-4 technical report. *arXiv preprint arXiv:2412.08905*.

Michael Burkhard, Sabine Seufert, Reto Gubelmann, Christina Niklaus, and Patcharin Panjaburee. 2023. Computer supported argumentation learning: Design of a learning scenario in academic writing by means of a conjecture map. In *CSEDU (1)*, pages 103–114.

Vanessa Wei Feng and Graeme Hirst. 2014a. A linear-time bottom-up discourse parser with constraints and post-editing. In *Proceedings of the 52nd Annual*

Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 511–521, Baltimore, Maryland. Association for Computational Linguistics.

Vanessa Wei Feng and Graeme Hirst. 2014b. Two-pass discourse segmentation with pairing and global features. *Preprint*, arXiv:1407.8215.

Bruce Fraser. 1996. Pragmatic markers. *Pragmatics*, 6:167–190.

Reto Gubelmann, Michael Burkhard, Rositsa V. Ivanova, Christina Niklaus, Bernhard Bermeitinger, and Siegfried Handschuh. 2024. Exploring the usefulness of open and proprietary llms in argumentative writing support. In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky*, pages 175–182, Cham. Springer Nature Switzerland.

Mi Chan Htaw, Daria Pipa, Namkang Sriwattanarothai, Chailerd Pichitpornchai, Reto Gubelmann, Sabine Seufert, Christina Niklaus, and Siegfried Handschuh. 2024. Argumentative writing software: Perceptions of undergraduate students toward artist prototype. In *2024 IEEE 7th Eurasian Conference on Educational Innovation (ECEI)*, pages 92–96.

Zhe Hu, Hou Pong Chan, and Yu Yin. 2024. AMERICANO: Argument generation with discourse-driven decomposition and agent interaction. Association for Computational Linguistics.

William C Mann and Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.

Roope Raisamo, Ismo Rakkolainen, Päivi Majaranta, Katri Salminen, Jussi Rantala, and Ahmed Farooq. 2019. Human augmentation: Past, present and future. *International Journal of Human-Computer Studies*, 131:131–143.

Christian Rapp and Peter Kauf. 2018. Scaling academic writing instruction: Evaluation of a scaffolding tool (thesis writer). *International Journal of Artificial Intelligence in Education*, 28(4):590–615.

Sabine Seufert, Josef Guggemos, and Stefan Sonderegger. 2020. Digitale transformation der hochschullehre: Augmentationsstrategien für den einsatz von data analytics und künstlicher intelligenz. *Zeitschrift für Hochschulentwicklung*, 15(1):81–101.

Marie Stevenson and Aek Phakiti. 2014. The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19:51–65. Feedback in Writing: Issues and Challenges.

Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, and 1 others. 2016. Artificial intelligence and life in 2030. *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*, 52.

Carola Strobl, Emilie Ailhaud, Kalliopi Benetos, Ann Devitt, Otto Kruse, Antje Proske, and Christian Rapp. 2019. Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education*, 131:33–48.

C.J. Thaiss and T.M. Zawacki. 2006. *Engaged Writers and Dynamic Disciplines: Research on the Academic Writing Life*. Boynton/Cook.

Thiemo Wambsganss, Christina Niklaus, Matthias Cetto, Matthias Söllner, Siegfried Handschuh, and Jan Marco Leimeister. 2020. AL: An adaptive learning support system for argumentation skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA. Association for Computing Machinery.

Jenny S. Wesche and Andreas Sonderegger. 2019. When computers take the lead: The automation of leadership. *Computers in Human Behavior*, 101:197–209.

Olaf Zawacki-Richter, Victoria I Marín, Melissa Bond, and Franziska Gouverneur. 2019. Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1):1–27.

Mo Zhang. 2013. Contrasting automated and human scoring of essays. *R & D Connections*, 21(2):1–11.