



Universiteit
Leiden
The Netherlands

Computer-Based Generation of Learner-Sensitive Feedback to Argumentative Learner Texts

Kilsbach, Sebastian; Michel, Nadine; Boogaart, Ronny; Garssen, Bart; Jansen, Henrike; Van Leeuwen, Maarten; ... ; Reuneker, Alex

Citation

Kilsbach, S., & Michel, N. (2024). Computer-Based Generation of Learner-Sensitive Feedback to Argumentative Learner Texts. *Proceedings Of The Tenth Conference Of The International Society For The Study Of Argumentation*, 501-508. Retrieved from <https://hdl.handle.net/1887/4107831>

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/4107831>

Note: To cite this publication please use the final published version (if applicable).

Computer-Based Generation of Learner-Sensitive Feedback to Argumentative Learner Texts

SEBASTIAN KILSBACH & NADINE MICHEL

Sebastian Kilsbach
Department of Language Didactics
University of Paderborn
33098 Paderborn
North-Rhine-Westphalia, Germany
sebastian.kilsbach@uni-
paderborn.de

Nadine Michel
Department of Language Didactics
University of Paderborn
33098 Paderborn
North-Rhine-Westphalia, Germany
nadine.michel@uni-paderborn.de

ABSTRACT: What is learner-sensitive feedback to argumentative learner texts when it is to be issued computer-based? Learning stages are difficult to quantify. The paper provides insight into the history of research since the 1980s and a preview of what this automated feedback might look like. These questions are embedded in a research project at the Universities of Paderborn and Hannover, Germany, from which a software (project name ArgSchool) emerges that will provide such feedback.

KEY WORDS: AI, argumentation mining, discourse history, (automated, learner-sensitive) feedback

1. INTRODUCTION

What is learner-sensitive feedback in the context of argumentative texts that school students write? How must learner-sensitive feedback on argumentative texts be conceptualized if it is generated automatically? What determines precisely which learning stage a student has already reached? These are, among others, relevant questions the current DFG-funded project Computational Support for Learning Argumentative Writing in Digital School Education (453073654) aims at. In our research we will develop a program or software (project name ArgSchool) for analyzing argumentative student texts to generate computer-based feedback conducive to learning.¹

The main goal of the project is to develop algorithmic methods for the automated analysis of argumentative learner texts and the generation of computer-based learner-sensitive feedback. To this end, a multilevel annotation process has been undertaken and completed, and a quality rating (assessment) is also underway. In this paper, however, we focus on learning stage sensitivity. After all, this determines what is actually understood by ‘conducive learner feedback’.

¹ The project runs for three years (2021–2024) led by Prof. Sara Rezat (Paderborn University, Language Didactics) and Prof. Henning Wachsmuth (Leibniz University Hannover, Computational Linguistics).

More precisely, we will classify from a discourse-historical perspective the preliminary work involved in the question of classifying a learning stage – and what makes it hard to find clear solutions. We discuss why the text order patterns (Textordnungsmuster) postulated by Augst & Faigel (1986) live on in the history of ideas until today. We then move on to ways in which various forms of automated feedback can be implemented.

2. MEASURING LEARNING STAGES?

2.1 First steps

In the German-speaking world, certain discourse relations or constructions in written argumentative texts have been studied since the 1980s, initially with a focus on language acquisition (cf. Augst, Disselhoff, Henrich, Pohl & Völzing, 2007, p. 199). The longitudinal study by Augst & Faigel (1986) was groundbreaking on this matter. Students from the first to the twelfth grade and students were asked to take position on the question of whether homework should be abolished. For this purpose, all students wrote a letter in an imaginary communication scenario to a so-called Prof. Augst, who claims to support this position. Some of the argumentative texts were collected later between 1988 and 1990 (cf. Augst, 1995).

The authors distinguished four different text ordering patterns (Textordnungsmuster). They were thus able to prove that all of them show up in the texts beyond the longitudinal age cut but are subject to a percentage shift.

At first, the strongly subjectively shaped linear-developing text order pattern dominates, at the end, the linear-dialogical pattern, which abstracts strongly from the own perspective, is found more frequently. (Becker-Mrotzek & Böttcher, 2020, p. 224, translation: SK)

Especially in earlier grades, linear-developing texts dominate, which have a chronological structure and are strongly influenced by the writer's own experience. Material-systematic texts understand the writing task in a strongly descriptive way and are oriented to it (often list-like). Only later do the other two patterns dominate, whereby form-systematic approaches are still more strongly influenced by school writing conventions, but (therefore) already distinguish pro and con arguments. Only in linear-dialogic texts does the social function of argumentation take a prominent place, since alternative perspectives are now presented with consideration of the addressed reader (cf. Augst & Faigel, 1986, pp. 114– 121).

Augst & Faigel (ibid., p. 109), drawing on preliminary work by Chafe (1982), postulate that texts constitute “super summative units” of “planned communication.” From this, they derive individual phenomena of planned integration and examine (a.) cognitive and interactional structures of text order, (b.) the structure of introduction and conclusion, (c.) the use of text- and argument-structuring metacommunicative turns of phrase, and (d.) the ability to orient to the addressee and the objective writing perspective (ibid., p. 110). These criteria can be determined as precursor indicators of a quality assessment of argumentative text acquisition.

2.2 State of the art

Later studies could support and extend the approaches shown in Augst & Faigel (1986) several times. Coirier & Marchand (1994) indicated that an age of 13 or 14 is a critical stage in the acquisition of argumentative structures. Coirier & Andriessen (1999) identified concessive argumentation (and thus the decentralization of perspective outlined in Augst & Faigel) as an essential point of an elaborated argumentative text (EAT). Nevertheless, concessive reasoning remains a critical skill across ages (cf. Stein & Bernas, 1999; Flower, Long & Higgins, 2000; Ferretti, 2000; Steinhoff, 2007), and for this very reason it needs to be supported in acquisition.

Of central relevance to our research questions are the results Leita (2003) generated for the relevance of counterarguments in texts written in school (second, fifth, and eighth grade). Participants of all ages tended to solely produce arguments in support of their own point of view rather than also discussing counterarguments (*ibid.*, p. 280).

Based on this pilot study, a qualitative study followed in which different age groups (second graders, fifth graders, and college freshmen) were presented with four different argumentative text versions containing different elements: (1) Introduction and point of view, (2) introduction, point of view, and (pro) arguments for point of view, (3) introduction, point of view, (pro) arguments, and counterarguments, as well as (4) a non-argumentative text.

At the broadest level, data from this study support the conclusion that participants do not conceive of counterarguments as a valuable rhetorical strategy in a text, unless the writer is able to preserve the strength of his or her own position by replying to it. (*ibid.*, p. 298)

These results are in sharp contrast to what is taught at school. There and in general, the appropriate application (and refutation) of counterarguments is considered a high-quality feature (cf. Feilke, 2010, p. 154).

A decentralized perspective, and with it a greater consideration of concessive structures, is thus one important criterion for measuring a student's learning stage. Augst, Disselhoff, Henrich, Pohl & Völzing (2007) refined the ontogenetic stage model of the Augst & Faigel survey twenty years earlier. It was composed of the following components: "Subjective selection and associative ranking" – this follows the observations of Feilke (1995) and can be strongly observed in younger cohorts; "subjective selection and connective complexification" – emotional attitudes towards the subject matter recede, also structurally the texts show more classical argumentation elements; "objective selection and analytical sequencing" – discourse relations or constructions such as weighing become increasingly apparent on the basis of pro-con rows; "objective selection and synthetic integration" – arguments lead to a clear conclusion, on the textual level the writers are more stylistically confident (Henrich, 2007, pp. 201–203).

It can therefore be said that Augst & Faigel's survey has had repercussions to this day. Neumann (2007) also took this up in her study on discursive writing products. A very precise scaling was developed, in which text structure was only one component among many. Rezat (2011) focused on the writing task and was primarily interested in concessive argument structures (cf. Anskeit, 2019, p. 98).

In 2019, Anskeit's empirical study on Writing Arrangements at the Primary Level sheds light on the influence of the writing task and medium in the writing process at the primary level. In the case of text form, argumentative texts were surveyed in addition to descriptions. The goal was to explore the "influence of teaching different text forms on text production" (Anskeit, 2019, p. 175). In the argumentation task, the participants were asked about their favourite excursion destination (zoo or Westfalenpark?). This was only one of many variables. The text quality was then evaluated using dichotomous criteria and holistic global judgments (cf. *ibid.*, 191). Dichotomously coded were the following variables (discourse units): Positioning (intro), Addressing, Justification, Anticipation, Weighing, Positioning (conclusion) (cf. *ibid.*, 195–197). The analysis was supplemented with a global scale (1 to 6), with text first and text last versions in the sense of research design. (For "rating and processing" of the data analysis, cf. *ibid.*, 223).

Our studies are based on this preliminary work. In addition to the aforementioned multi-level annotation, we also include a quality rating. However, the question of how a specific learning stage can really be measured by means of operationalizable indicators has been directly or indirectly circled in all studies since Augst & Faigel, without producing any tangible results so far (one reason for this is, of course, the diverging research interests). Nevertheless, this results in a desideratum: The measurement of learner stages is essential to provide individualized, developmental feedback. We therefore proceed in the following by giving an insight into our preliminary considerations for the generation of feedback and point out our connection points to relevant preliminary work. The underlying question is as to what learner-sensitive feedback might look like in the ArgSchool software.

3. FEEDBACK – PRELIMINARY CONSIDERATIONS

As our next project step, we will generate feedback for the annotated corpus of argumentative learner texts. Based on the considerations about the learning stages of the students, the feedback is to be intended to support students in the sense of formative feedback during the ongoing writing process and is generated according to individual needs. The aim of the feedback is to improve the quality of the text and especially the argument structure to move from the current learning stage to the next higher one. "Formative feedback is defined [...] as information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning" (Shute, 2008, p. 154).

Providing feedback on writing is an important part of teaching students to improve their own writing skills and writing processes, as writing places high demands on students (cf. Graham, Harris & Hebert, 2011). The most important project-related task for us is therefore to generate feedback to argumentative student texts that is as learner-sensitive and conducive to learning as possible, so that they are supported in the best possible way in the acquisition of their competencies.

Based on the multi-level annotation of the student texts as well as the qualitative assessment, learning stages must first be established to be used as a basis for feedback. The feedback will not only be generated for surface features of the texts, but will be much more detailed and focused on the acquisition of specific patterns of argumentative texts, both on a structural and on a content level. Steendam, Rijlaarsdam, van den Bergh & Sercu (2014)

stated that weaker writers often consider only surface aspects, so-called “lower-order concerns” (ibid., p. 909), when revising their texts. Through feedback on their performance, they should gain the ability to look at their texts more globally and revise “higher-order concerns” (ibid.) such as content and structure. Thus, one of the goals of learning-sensitive feedback is the training of the writers’ competencies as well as the development of structural knowledge and quality aspects of content. This goal can also be supported by further research, since feedback on low-order concerns (e.g., grammar, spelling, text length) is easy to implement and does not require much effort on the part of the writer (cf. Wagner & Lachner, 2021, p. 5). In contrast, feedback on higher-order concerns (e.g., organization, structure, style, and cohesion) is thought to have a more significant impact on text quality (cf. Patchan, Schunn & Correnti, 2016; Wagner & Lachner, 2021). By providing information about substantive aspects of a writing product, writers can be encouraged to realize a revision that can sustainably improve text quality (cf. Wagner & Lachner, 2021, p. 5). Accordingly, feedback on higher-order concerns tends to aim at a sustainable improvement of writing performance and is thus relevant for continuous improvement in the sense of a formative assessment. The planned feedback program is intended to finally capture the structure of learner texts at different levels (based on the learned patterns through multi-level annotation) and assess text quality (based on the quality ratings). Both aspects are needed to assess the writer’s learning stage and to generate individualized feedback through which the next learning stage can be reached. It is therefore clear to which text levels feedback should be generated.

Another important aspect is the content of the feedback. In addition to the text levels to be considered, the content of the feedback must also be designed in a learning-sensitive manner and promote the development of the students’ writing skills. The model of Hattie & Timperley (2007, p. 87) provides a good basis for the content-based design of feedback that promotes learning. According to them, feedback conducive to learning should answer three questions: (1) Where am I going? (Feed Up), (2) How am I going? (Feed Back), and (3) Where to next? (Feed Forward). Feed Up shows the target state and should be clearly defined. In best case, the learning objective is clearly defined in the classroom, which is then referred to in the context of feedback. Thus, the writer receives a reminder of the set goals regarding the writing product. Feed Back, on the other hand, shows the current or actual state and shows where the learner is in the ongoing learning process. The step-by-step process of increasing individual performance in order to reach the defined goal is called Feed Forward and is intended to reduce the difference between the actual and the target state. In particular, the feed forward plays an important role in the success of feedback, since feedback without concrete information on how to achieve an improvement is often left unused and thus has no effect on the learning process (cf. Busse, Müller & Siekmann, 2022, p. 117). To activate the learner to implement or apply the feedback in the context of a revision, concrete indications on potential for improvement are therefore crucial. The three basic questions that any effective feedback should answer are amplified in the model by levels of feedback that should be considered: “There is a distinction between feedback about the task (FT), feedback about the processing of the task (FP), about self-regulation (FR), and about the self as a person (FS)” (Hattie & Timperley, 2007, p. 90). All levels can have effects on the learner and should be considered when formulating feedback.

In addition to the groundbreaking findings of Hattie & Timperley (*ibid.*), other factors may influence the effectiveness of feedback. Shute (2008) names, for example “(a) motive (the student needs it), (b) opportunity (the student receives it in time to use it), and (c) means (the student is able and willing to use it)” (*ibid.*, p. 175). Despite the many other factors that can be found in the research literature on feedback, the model of Hattie & Timperley (2007) will be used as a basis, which can then be expanded during the development process for feedback.

The final issue under consideration at this time is the design of the feedback and how it can be delivered. When thinking about the design of feedback, several options are possible. Embedding the feedback into the text, as recommended by Wagner & Lachner (2021), facilitates the revision and transfer process for students. In this variant, text passages can be explicitly highlighted and comments or feedback on this passage can be directly implemented by the student without requiring a complex transfer performance from him. On the other hand, it would also be possible to provide feedback in the form of an overall evaluation that names individual parts of the text or conveys holistic impressions. In this case, the transfer performance for the students is higher since the feedback text must be compared with their own writing product and references between the two must be established.

Additionally, there is the possibility to integrate visual feedback, as implemented in the C-SAW tool (cf. Benetos & Bétrancourt, 2020). In the current project, visual feedback could graphically represent the argumentative structures of the text (e.g., as a tree diagram) and point out existing as well as missing structures or content. The software (C- SAW) is not intended as a feedback tool, because it rather supports the ongoing writing process for argumentative texts and is about digital authoring support. The software includes text modules that are to be filled in by the writer and assembled into an overall text. In this way, writers should gain sample knowledge about text structure and recognize missing structures more easily or consciously reflect on why text modules are used. The software also works with a visualization of the existing and missing structures in the form of a graphic tree map. Such a visualization would also be conceivable for the ArgSchool software but should then be made available primarily for teachers who can draw conclusions about general support needs of the learning group based on the representation.

4. CONCLUSION

At the current stage, there are still many unanswered questions regarding the design of the feedback and a separate approach based on the previous project steps is being developed. The goal of generating developmental and learner-sensitive feedback in an automated way remains, but also offers some challenges on different levels. A great advantage of automated feedback is that individualized feedback can be obtained independently of a real teacher. Access to feedback is easier overall and can be ensured on a very regular basis. In addition to the feedback from the teachers, students could also receive feedback on their text at home and during the ongoing writing process. This could happen several times a week if the student is motivated to write and revises the text based on the feedback. Thus, the process would have a corresponding positive effect on the writing process.

Despite its advantages, automated feedback should be seen as a complement to human feedback by a teacher rather than a replacement. It must be critically questioned whether AI software can provide the same performance as a human feedback provider with the appropriate expertise. AIs are already far advanced, but they are not completely perfect when it comes to evaluating and analysing the content of texts. Especially the effective feedback on higher-order-concerns poses great challenges to the systems, because for an adequate evaluation not only the resulting text product but also the context to the set learning goals and the underlying writing task are necessary. Interpersonal communication (e.g., gestures or facial expressions) as well as metacognitive influencing factors often cannot be performed at all or only to a very limited extent in automated feedback. Thus, the development of learner-sensitive, developmental automated feedback is still in its infancy and requires further research, to which the ArgSchool software is a contribution.

Of particular importance for successful feedback are clearly defined learning objectives, which are best defined in class and communicated by the teacher in a suitable form. In gradually moving from the current state of the writing product to the defined learning goal, the ArgSchool program is designed to actively support learners and drive skill acquisition. In addition to supporting learners, teachers can also use the data to get a better overview of the group's learning progress and make any needs visible. Whether as a basis for human feedback, as automated feedback for students, or as an evaluation option for argumentative learner texts, the possible uses of the program can be manifold. Therefore, the ArgSchool program could ultimately make a valuable contribution to the expansion of digital teaching content and promote the argumentative skills of students in a learning-sensitive and development-oriented manner.

REFERENCES

- Anskeit, N. (2019). Schreibarrangements in der Primarstufe. Eine empirische Untersuchung zum Einfluss der Schreibaufgabe und des Schreibmediums auf Texte und Schreibprozesse in der 4. Klasse. Münster: Waxmann.
- Augst, G. (Ed.) (1995). Frühes Schreiben. Studien zur Ontogenese der Literalität. Essen: Die Blaue Eule.
- Augst, G., & Faigel, P. (1986). Von der Reihung zur Gestaltung. Untersuchungen zur Ontogenese der schriftsprachlichen Fähigkeiten von 13-23 Jahren. Frankfurt a.M.: Peter Lang.
- Augst, G., Disselhoff, K., Henrich, A., Pohl, T., & Völzing, P. (2007): Text – Sorten – Kompetenz. Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter. Frankfurt a.M.: Peter Lang.
- Becker-Mrotzek, M., & Grabowski, J. (2018). Textkorpus Scriptoria. In M. Becker-Mrotzek, & J. Grabowski (Eds.), FD-LEX (Forschungsdatenbank Lernertexte). Köln: Mercator-Institut für Sprachförderung und Deutsch als Zweitsprache. DOI: 10.18716/FD-LEX/861.
- Benetos, K.; Bétrancourt, M. (2020): Digital authoring support for argumentative writing: What does it change? *Journal of Writing Research*, Juni 2020, S. 264–290.
- Busse, V., Müller, N., & Siekmann, L. (2022): Wirksame Schreibförderung durch diversitätssensibles formatives Feedback. In: V. Busse, N. Müller & L. Siekmann (Eds.): Schreiben fächerübergreifend fördern. Grundlagen und Anregungen für Schule, Unterricht und Lehrkräftebildung (pp. 114–133). Hannover: Klett/Kallmeyer.
- Coirier, P., & Marchand, E. (1994). Writing argumentative texts. A typological and structural approach. In: *Writing. Current trends in European research* (pp. 162–181). Freiburg: Hochschul-Verlag.
- Coirier, P., & Andriessen, J. (1999). *Foundations of Argumentative Text Processing*. Amsterdam: University Press.
- Chafe, W. L. (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

- Feilke, H. (1995): Auf dem Weg zum Text. Die Entwicklung der Textkompetenz im Grundschulalter. In: G. Augst (Ed.), *Frühes Schreiben. Studien zur Ontogenese der Literalität* (pp. 69–88). Essen: Die Blaue Eule.
- Feilke, H. (2010). Kontexte und Kompetenzen – am Beispiel schriftlichen Argumentierens. In: P. Klotz, P. Portmann-Tselikas & G. Weidacher (Eds.), *Kontexte und Texte* (pp.147–167). Tübingen: Narr.
- Ferretti, R., MacArthur, C., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology*, 92, 694–702.
- Flower, L., Long, E., & Higgins, L. D. (2000). *Learning to Rival. A Literate Practice for Intercultural Inquiry*. Philadelphia: Erlbaum.
- Graham, S.; Harris, K.; and Hebert, M. A. (2011). *Informing writing: The benefits of formative assessment. A Carnegie Corporation Time to Act report*. Washington DC: Alliance for Excellent Education.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77, 81–112. DOI: 10.3102/003465430298487.
- Henrich, A. (2007). Einzelanalyse der Textsorte Argumentation. In: G. Augst, K. Disselhoff, A. Henrich, T. Pohl & P. Völzing (Eds.), *Text – Sorten – Kompetenz* (pp. 199–203). Frankfurt a.M.: Lang.
- Leitao, S. (2003). Evaluating and Selecting Counterarguments. *Studies of Children's Rhetorical Awareness. Written Communication*, 20(3), 269–306.
- Neumann, A. (2007). *Briefe schreiben in Klasse 9 und 11. Beurteilungskriterien, Messungen, Textstrukturen und Schülerleistungen*. Münster/New York/München/Berlin: Waxmann-Verlag.
- Patchan, M, Schunn, C., & Correnti, R. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology*, 108(8), 1098– 1120. DOI: 10.1037/EDU0000103.
- Rezat, S. (2011). Schriftliches Argumentieren. Zur Ontogenese konzessiver Argumentationskompetenz. *Didaktik Deutsch*, 31, 50–67.
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78, 153–189. DOI: 10.3102/0034654307313795.
- Steendam, E. v., Rijlaarsdam, G., van den Bergh, H., & Sercu, L. (2014). The mediating effect of instruction on pair composition in L2 revision and writing. *Instructional Science*, 42, 905–927.
- Stein, N., & Bernas, R. (1999). The Early Emergence of Argumentative Knowledge and Skill. In: J. Andriessen & P. Coirier (Eds.), *Foundations of Argumentative Text Processing* (pp. 97–116). Amsterdam: University Press.
- Steinhoff, T. (2007). *Wissenschaftliche Textkompetenz. Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten*. Tübingen: Niemeyer.
- Wagner, S., & Lachner, A. (2021). Feedback – Ja, klar?! *Digitale Medien zur Förderung von Schreibkompetenzen. Leseforum.ch*, 3, 2–14.