Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

Dialogism Meets Language Models for Evaluating Involvement in CSCL Conversations

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Dascalu, Maria-Dorinela	University Politehnica of Bucharest	
Ruseti, Stefan	University Politehnica of Bucharest	
Dascalu, Mihai	University Politehnica of Bucharest	
McNamara, Danielle S.	Arizona State University	
Trausan-Matu, Stefan	University Politehnica of Bucharest	

Publication/Completion Date—(if *In Press,* enter year accepted or completed) 2022

Check type of content being submitted and complete one of the following in the box below:

- ☐ If article: Name of journal, volume, and issue number if available
- ☑ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

6th international conference on smart learning ecosystems and regional development, June 2021, Bucharest, F

DOI or URL to published work (if available) https://doi.org/10.1007/978-981-16-3930-2_6

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** National Center for Educat through **[Grant number]** R305A1801 to **Institution]** Arizona State U.The opinions expressed are those of the authors and do not represent views of the **[Office name]** National Center for Educa or the U.S. Department of Education.

Chapter 6
Dialogism Meets Language Models for Evaluating Involvement in CSCL Conversations

Maria-Dorinela Dascalu, Stefan Ruseti, Mihai Dascalu, Danielle S. McNamara, and Stefan Trausan-Matu

Abstract The use of technology as a facilitator in learning environments has become increasingly prevalent with the global pandemic caused by COVID-19. As such, computer-supported collaborative learning (CSCL) gains a wider adoption in contrast to traditional learning methods. At the same time, the need for automated tools capable of assessing and stimulating collaboration between participants has become more stringent, as human monitoring of the increasing volume of conversations becomes overwhelming. This paper introduces a method grounded in dialogism for evaluating students' involvement in chat conversations based on semantic chains computed using language models. These semantic chains reflect emergent voices from dialogism that span and interact throughout the conversation. Our integrated method uses contextual information captured by BERT transformer models to identify links in a chain that connects semantically related concepts from a voice uttered by one or more participants. Two types of visualizations were generated to depict the longitudinal propagation and the transversal inter-animation of voices within the conversation. In addition, a list of handcrafted features derived from the constructed chains and computed for each participant is introduced. Several machine learning algorithms were tested using these features to evaluate the extent to which semantic chains are predictive of student involvement in chat conversations.

M.-D. Dascalu · S. Ruseti · M. Dascalu (✉) · S. Trausan-Matu
University Politehnica of Bucharest, 313 Splaiul Independentei, 060042 Bucharest, Romania
e-mail: mihai.dascalu@upb.ro

M.-D. Dascalu
e-mail: dorinela.dascalu@upb.ro

S. Ruseti
e-mail: stefan.ruseti@upb.ro

S. Trausan-Matu
e-mail: stefan.trausan@upb.ro

M. Dascalu · S. Trausan-Matu
Academy of Romanian Scientists, Str. Ilfov, Nr. 3, 050044 Bucharest, Romania

D. S. McNamara
Department of Psychology, Arizona State University, PO Box 871104, Tempe, AZ, USA
e-mail: dsmcnama@asu.edu

Ó. Mealha et al. (eds.), *Ludic, Co-design and Tools Supporting Smart Learning Ecosystems and Smart Education*, Smart Innovation, Systems and Technologies 249,
https://doi.org/10.1007/978-981-16-3930-2_6

Keywords Dialogism · Computer-supported collaborative learning · Semantic chains · Language models

6.1 Introduction

Smart learning environments greatly benefit from the development of technology [1], by improving educational processes, reducing the time to perform certain tasks, increasing the availability of resources, and providing an ecosystem that stimulates creativity and the desire to learn. Thus, learning becomes easily accessible to all people from different parts of the world, and communication with from different cultures is at hand, while resources are readily available. Nevertheless, technology should remain an enabler, given a *people in place-centered* perspective [2] in which attractiveness, coupled with the adhesion, must transcend toward a sense of belonging. The need of such learning environments has dramatically increased during the COVID-19 pandemic [3]. The transition from physical classes to fully online environments was drastic and adaptation was required in a short amount of time; as such, smart learning environments eased this transition, while supporting both students and teachers. Since everything moved online, face-to-face class discussions transitioned to forums, chats, and video meetings.

In conjunction with the adoption of learning environments, computer-supported collaborative learning (CSCL) has become increasingly used in educational contexts due to its synergic effects among peers. Learners share their ideas and opinions, learn from each other, while having access to a wide range of materials. Chats and forums, the most commonly used CSCL environments, offer learners the opportunity to work together to solve problems and ask for help when encountering issues—more generally, students collaboratively build knowledge and share it among all participants [4]. Collective and individual learning processes intertwine one with another to create collaborative knowledge, which is spread among all participants [5].

In CSCL environments, technology empowers communication and collaboration throughout the learning process. However, from tutor's perspective, analyzing the resulting conversations is a time-consuming task due to their increased volume. Therefore, automated tools that analyze conversations and evaluate collaboration between participants have become a necessity. Moreover, collaboration and creativity among peers can be stimulated using automated processes [6].

Dialogism was first introduced by Bakhtin [7] as a philosophical theory focusing on the idea that everything is a continuous exchange and interaction between several voices [8]. From a dialogical perspective, discourse is modeled as a weave of interactions in natural language among people, with the essential goal of building meaning and understanding. Voices represent different points of view that spread throughout the discourse and influence it. Closely correlated with the concept of voice are multivocality and polyphony, which are key features in dialogism. Dialogism is considered a paradigm for CSCL [9], where voices (points of view) take the form of concepts or events that are propagated throughout the conversation by participants who share convergent or divergent perspectives [10]. Multivocality and polyphony are key features in dialogism and are closely correlated to the concept of voice.

Voices found in participants' contributions from CSCL conversations interact one with another and influence each other. Throughout a conversation, the inter-animation of the voices is a key component for the success of the collaboration. The interactions between the participants are reflected in their voices; therefore, polyphony can be an indicator of collaboration [11]. The evolution of voices throughout a conversation and their influence on other participants provide valuable insights into collaboration.

This paper introduces a method grounded in dialogism to evaluate students' involvement in chat conversations based on semantic chains identified using state-of-the-art language models, namely bidirectional encoder representations from transformers [BERT; 12].

The paper is structured as follows. The next section presents state-of-the-art methods and solutions. The third section introduces our method, the corpus used for evaluation, and the features derived from the identified semantic chains. The fourth section describes our results, together with the longitudinal and transversal visualizations of semantic chains (i.e., voices), followed by conclusions and future work.

6.2 State of the Art

The number of applications and plugins aimed to support learning processes is constantly growing, while the COVID-19 pandemic has further evidenced the power of technology in the learning processes. Due to the COVID-19 pandemic, most learning institutions had to operate entirely online. This would have been almost impossible 20 years ago due to the lack of technologies and auxiliary solutions. Smart learning environments, online communication platforms (Zoom, Microsoft Teams, Google Meet), and discussion channels (forums, chats) facilitated this transition. However, online collaboration is different from face-to-face discussions in a classroom. In online environments, emotional aspects of the conversation, reflected in facial expressions that a teacher can interpret and guide, are lost or at least diminished when video is enabled. Students lose focus more easily and jump from one topic to another, for example, by simply posting a picture or a link. Thus, following the ways in which students collaborate and maintain focus presents a far more tedious task for teachers.

The information that spreads during a conversation, the topics that are discussed and the immediate transitions from one topic to another, is key components in evaluating the collaboration between participants. CSCL environments aim to support students in the learning process by emphasizing collaboration using chats or forums. Dialogism, considered the theoretical framework of CSCL [9], was first introduced by Bakhtin [7] as a philosophical theory. According to Bakhtin, everything around us is a continuous change and an interaction between several voices. Multiple elements can be derived from the inter-animation of voices: convergence of points of view, potential conflicts, the way the information spreads, the change of a topic, or the taking over of another; all these elements lead to true polyphony [8].

Starting from Bakhtin's theory of polyphony and inter-animation of voices, Trausan-Matu et al. [13–15] proposed the polyphonic model of discourse analysis for chat conversations. The polyphonic model is based on the identification of voices and builds a graph-based representation of the conversation in which the inter-animation of voices generates convergences or divergences of points of view. Tracking events in a conversation in chronological order reflects a *longitudinal* dimension of discourse. The voices that propagate and inter-animate one with another reflect the interactions between the participants and their collaboration [11]. The exchange of ideas, the abandonment of one topic and the taking over of another, the divergences and convergences that may appear at a certain moment reflect a *transversal* dimension of the discourse [16].

From a computational point of view, voices represent semantic chains [17], which in turn can be generated using lexical chains [18] (i.e., sequences of words that are repeated or semantically related, including synonyms, hyponyms, or siblings). Jayarajan, Deodhare, and Ravindran [19] rely on nouns and compound nouns in identifying lexical chains. Traditional methods of identifying lexical relationships between words are based on Wordnet [20] or Roget's Thesaurus [21]. Mukherjee, Leroy, and Kauchak [18] used lexical chains identified from word repetitions, synonyms, and semantic relationships between nouns, to classify medical texts into two categories: easy and difficult. Ruas et al. [22] combined lexical chains with word embeddings to extract semantic relationships between words. Migrating to CSCL environments, the evaluation of collaboration and interactions between participants can be performed based on voice overlap [23]. The ways in which voices are emitted and propagate throughout the conversation, and the interconnection of the exposed points of view is all indicative of collaboration [23].

6.3 Method

In this study, we extend the method proposed by Ruseti et al. [24] to identify semantic chains in chat conversation and derive corresponding features from these chains predictive of student involvement.

6.3.1 Corpus

Our analysis is performed on the same chat conversations processed in detail by Dascalu et al. [11, 25]. This corpus consists of 10 chats selected from a corpus of more than 100 conversations which were scored by 4 raters. The conversations took place between four to five undergraduate students studying computer-human interaction who debated on the advantages and disadvantages of specific CSCL technologies. The students had known each other since previous courses. During the conversations, each participant was an advocate of a technology and tried to convince the other participants of the advantages of their chosen technology. Afterward, all participants had to come up together with a new solution which incorporated the discussed advantages.

6.3.2 Building Semantic Chains

Our method uses contextual information captured by BERT [12] to identify the links in a semantic chain. BERT is a transformer-based deep neural network that uses a mechanism of attention, which learns the contextual relationships between words from a text. BERT builds contextual representations of words by stacking multi-head attention layers.

A dataset derived from the TASA corpus (http://lsa.colorado.edu/spaces.html) containing potential links between words in a given context was automatically generated given simple rules, namely (a) repetitions of words with the same lemma; (b) synonymy, hypernymy, or sibling relationships using WordNet [20]; and (c) coreferences identified using spaCy (https://spacy.io). This dataset was used to train a model on how to effectively combine the different attention heads from BERT using a multi-layer perceptron in order to identify generalized semantic links [24].

The previously trained model was then used to identify all potential semantic links in a conversation, while accounting for all pairs of words. Starting from these links, semantic chains are generated as connected components in the graph obtained by interconnecting all links exceeding an imposed similarity threshold. The entire procedure is described in detail by Ruseti et al. [24].

6.3.3 Semantic Chain Features

The previous semantic chains computed with BERT are used to assess the involvement of students within the conversation. The chains cannot be directly used to build a prediction model; as such, a list of features was defined based on the constructed chains. Part of the features are inspired from LEX-1 [26] and are generally applicable to lexical chains, whereas the remaining ones are chat specific. All handcrafted features (see Table 6.1) are computed for each participant from the conversation.

The chat-specific features are designed to capture the interaction between participants based on the semantic chains that span across their contributions. Each semantic chain represents a distinct topic discussed; therefore, part of the features consider how chains are initiated by a participant, and how semantic chains are afterward continued by the same or other speakers. Even if a chain is only included in the

Table. 6.1 Feature description

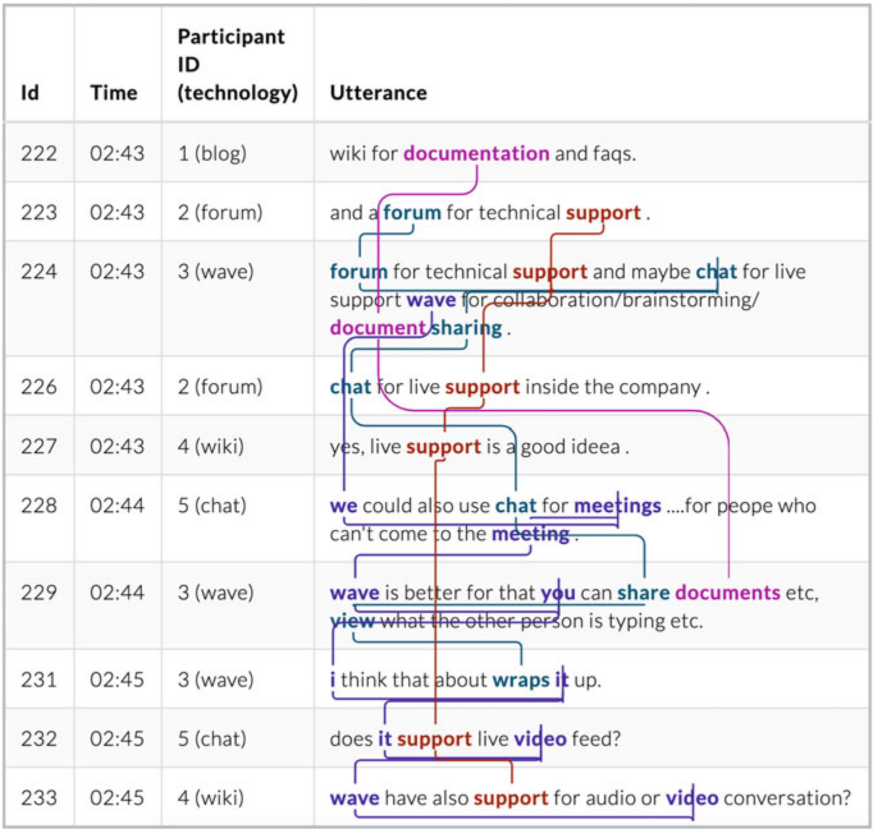
Feature name	Description
General features	
Chains	Count (#) and ratio of chains (i.e., how many semantic chains are used by a specific user divided by the overall count of semantic chains present in the conversation)
Large chains	Count (#) and ratio of large chains (i.e., semantic chains containing more than 4 words)
Varied chains	Count (#) and ratio of varied chains (i.e., semantic chains with more than one different lemma)
Large and varied chains	Count (#) and ratio of chains that are both large and varied
Chat-specific features	
Initiated chains	Count (#) and ratio of chains initiated by the participant
Independent chains	Count (#) and ratio of chains with only one participant
Avg. participants	Average participants per chain
Avg. words	Average words per chain for each participant
Continuations	Count (#) of backward links from the current participant to another participant
Avg. continuation length	Average words in the conversation for each backward link between different participants

contributions of a single participant, it still remains relevant for measuring participation because it denotes active involvement. Additional metrics are also considered to account for collaboration besides mere chain counts, for example, word occurrences per chain belonging to a specific participant (denoting topic coverage among participants), as well as the delay in the conversation before continuing a given voice (quantifying a pause in terms of topic continuation, measured as words in-between two occurrences from the same semantic chain in two contributions pertaining to different participants).

6.4 Results

6.4.1 Semantic Chains Visualizations

Interactive visualizations were introduced to highlight both a longitudinal propagation of voices (see Fig. 6.1) and transversal overlap of semantic chains between participants (see Fig. 6.2). The views were developed using Angular 6 (https://angular.io), while the links between words were drawn using SVG. The same chat excerpt from Fig. 6.1 was analyzed by Dascalu et al. [11]. Words are colored according to the semantic chain two which they belong as well as the corresponding links. Each row represents an utterance, enriched with following details: identifier, timestamp,



Id	Time	Participant ID (technology)	Utterance
222	02:43	1 (blog)	wiki for documentation and faqs.
223	02:43	2 (forum)	and a forum for technical support .
224	02:43	3 (wave)	forum for technical support and maybe chat for live support wave for collaboration/brainstorming/ document sharing .
226	02:43	2 (forum)	chat for live support inside the company .
227	02:43	4 (wiki)	yes, live support is a good ideea .
228	02:44	5 (chat)	we could also use chat for meetings ....for peope who can't come to the meeting .
229	02:44	3 (wave)	wave is better for that you can share documents etc, view what the other person is typing etc.
231	02:45	3 (wave)	i think that about wraps it up.
232	02:45	5 (chat)	does it support live video feed?
233	02:45	4 (wiki)	wave have also support for audio or video conversation?

Fig. 6.1 Longitudinal view of semantic chains within a conversation

and participant identifier, which was incrementally generated for anonymization and followed by the supported technology by each participant. Four semantic chains were identified in the conversation segment shown in Fig. 6.1: 1—concepts related to documentation (e.g., "documents," "documentation"); 2—concepts related to text-centered CSCL technologies (e.g., "forum," "chat," "share," "view"); 3—concepts related to CSCL technologies that integrate video (e.g., "wave," "meetings," "video"); 4—concepts related to actions facilitated by technologies (e.g., "support"). The identified semantic chains reflect the theme of the conversations, more specifically the advantages and disadvantages of the CSCL technologies considered.

In contrast with the initial findings of Dascalu et al. [11], our method identifies more semantic chains with more related words—for example, new semantic chains—concepts related to documentation and actions; more related words—"meetings" and "video" are related to "wave"; chats and forum are now aggregated together as language models grasp their similarity. For the same part of conversation, the initial results [11] did not identify the semantic chains related to documentation (fuchsia

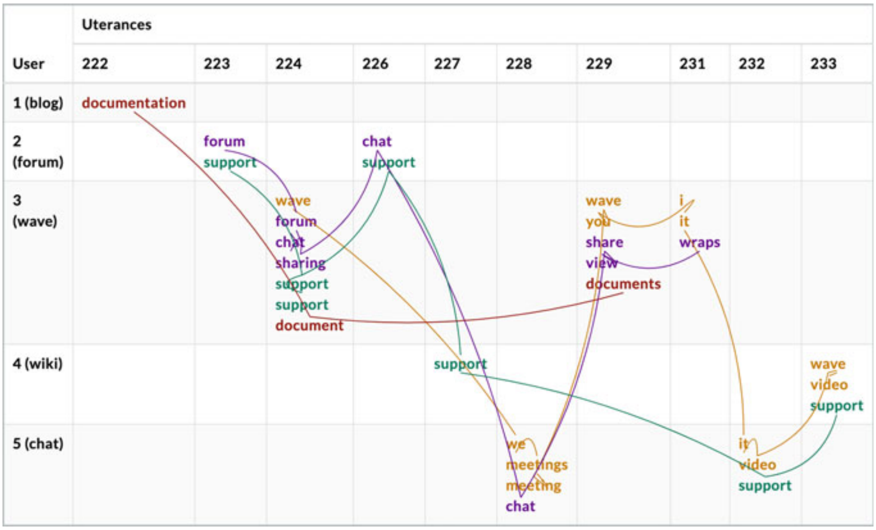


Fig. 6.2 Transversal view of semantic chains within a conversation

color) and actions (red color). Moreover, the initial results [11] did not include "meetings," "video," and "meetings" as related words to "wave," but the semantic chain contained only "wave." In addition, the new method identifies pronouns as part of the semantic chain (coreference resolution). Within the new method, "chat" and "forum" are now in the same semantic chain, while in the initial version [11], they were in separate chains. The sample denotes a higher cohesion between adjacent contributions, in contrast to previous results [11] in which a lower cohesion was argued for the same sample. Figure 6.2 presents a transversal view of the occurrences of semantic links whose underlying concepts are uttered by different participants. The words and the links are colored according to their corresponding semantic chain; colors differ between the visualizations because they are randomly selected.

6.4.2 Involvement Prediction

Several machine learning algorithms were tested using the previously introduced features to evaluate the performance of semantic chains in assessing involvement of students within the conversations. Since the dataset contained 10 distinct conversations, a tenfold cross-validation was performed, leaving one chat out for testing in each fold. Two different grades, namely participation (i.e., reflective of active involvement) and collaboration (i.e., interactions with peers) manually scored between 1 and 10, were predicted with two separate models. Table 6.2 shows a comparison of different models for the two tasks, and by taking into account the mean average error (MAE) on the two tasks, lower MAE values denote better models that are closer to

Table. 6.2 Model evaluation (values in bold mark the best performing model)

Model	Parameters	MAE participation	MAE collaboration
Inter-human agreement	–	0.907	0.819
Random forest	N trees = 10	**0.549**	0.655
Random forest	N trees = 100	0.551	**0.631**
SVR	RBF	0.662	0.674
SVR	Poly	1.391	1.397
Bayesian regression	–	0.625	0.658

the average human ratings. The human performance is approximated by computing the MAE of each rater compared to the average of all four ratings.

The human performance obtained on this dataset shows a MAE lower than 1 out of 10, which denotes a close, but not perfect agreement among raters (inherently, part were more relaxed, while the other were more fastidious); as such, 4 ratings were gathered for each participant and the system predicts the raters' average rating for both participation and collaboration. Random forest is the most predictive model reaching a MAE of 0.55; lowering the number of generated trees has a beneficial impact on performance, given the limited number of features and examples. Overall, the machine learning models seem to better capture the average scores of the raters, having the advantage of generalizing across all participants.

A subsequent analysis was performed to understand the importance of each feature for the two different tasks. The best model from the cross-validation for each task was trained on the whole dataset, and the Gini importance for each feature is presented in Table 6.3 [27]; higher values denote more relevant features, and the sum of all Gini feature importance scores is 1 for each prediction. The ratio of chains, denoting the coverage of semantic chains used by a participant in relation to the entire conversation, is by far the most predictive feature for participation with a score of 0.63. Continuations, reflective of links between two different participants (Gini importance of 0.35), coupled with counts and ratios of chains covered by the participant (Gini importance 0.16 and 0.15, respectively) are the best predictors for collaboration.

6.5 Conclusions and Future Work

CSCL environments help students learn by collaborating, sharing ideas and opinions, and finding the best solution together. CSCL technologies, such as chats and forums, are indispensable today, and they represent valuable sources for follow-up analyses.

Table. 6.3 Gini feature importance

Feature	Participation	Collaboration
# Chains	0.0047	0.1562
Chain ratio	**0.6294**	0.1457
# Large chains	0.0059	0.0235
Ratio of large chains	0.0064	0.0200
# Varied chains	0.0138	0.0457
Ratio of varied chains	0.0084	0.0631
# Large and varied chains	0.0024	0.0048
Ratio of large and varied chains	0.0275	0.0156
# Initiated chains	0.0036	0.0083
Ratio of initiated chains	0.0579	0.0274
# Independent chains	0.0034	0.0238
Ratio of independent chains	0.0881	0.0194
Avg. participants	0.0195	0.0262
Avg. words	0.0955	0.0307
# Continuations	0.0260	**0.3547**
Avg. continuation length	0.0076	0.0349

Bold denotes the most important feature for the considered predicted value

All activities moved online during the COVID-19 pandemic; as such, student assessment and monitoring become more challenging tasks for teachers. Collaboration between students in completing homework and their involvement in school activities are elements that must be taken into account in their assessment.

Within this paper, we introduce an automated method for evaluating students' involvement in chat conversations using dialogism as a paradigm and language models for identifying semantic chains. This study is an improvement on the analysis performed by Dascalu et al. [11]. Our method relies on contextual information captured by BERT to identify semantic links that are grouped into chains. Longitudinal and transversal visualizations were generated to highlight occurrence patterns of semantic chains, their propagation and co-occurrence throughout the conversation. In contrast to the previous results, our new method identifies more semantic chains with more related words, while also considering coreference resolution. A list of features was introduced, containing both general metrics applicable to the constructed semantic chains and chat-specific features, all computed for each participant. Several machine learning algorithms were tested using the defined features, and the best models based on random forest were capable to accurately predict participation and collaboration with a MAE of around 0.55 on a 10-point scale.

In terms of future work, we envision applying this dialogical model on other datasets and performing related analyses. For example, dialogism can be used to monitor student engagement in online courses, and the introduced features can be employed to predict dropout or course grades. Even more, by evaluating the introduced semantic chains, discussions and corresponding contributions can be cataloged as being course-specific, administrative, or off-topic, and an automated guidance mechanisms can be introduced, while targeting creativity stimulation.

Acknowledgements The work was funded by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS—UEFISCDI, project number TE 70 PN-III-P1-1.1-TE-2019-2209, ATES—"Automated Text Evaluation and Simplification." This research was also supported in part by the Institute of Education Sciences (R305A180144) and the Office of Naval Research (N00014-19-1-2424). The opinions expressed are those of the authors and do not represent views of the IES or ONR.

References

1. Chen, N.-S., Cheng, I.-L., Chew, S.W.: Evolution is not enough: revolutionizing current learning environments to smart learning environments. Int. J. Artif. Intell. Educ. **26**(2), 561–581 (2016)
2. Giovannella, C.: Smart learning eco-systems: "fashion" or "beef"? J. e-Learn. Knowl. Soc. **10**(3), 15–23 (2014)
3. Dhawan, S.: Online learning: a panacea in the time of COVID-19 crisis. J. Educ. Technol. Syst. **49**(1), 5–22 (2020)
4. Stahl, G.: Group Cognition. Computer Support for Building Collaborative Knowledge. MIT Press, Cambridge, MA (2006)
5. Cress, U.: Mass collaboration and learning. In: Luckin, R., Puntambekar, S., Goodyear, P., Grabowski, B., Underwood, J., Winters, N. (eds.) Handbook of Design in Educational Technology, pp. 416–424. Routledge, New York (2013)
6. Stamati, D., Dascalu, M., Trausan-Matu, S.: Creativity stimulation in chat conversations through morphological analysis. University Politehnica of Bucharest Scientific Bulletin Series C—Electr. Eng. Comput. Sci. **77**(4), 17–30 (2015)
7. Bakhtin, M.M.: The Dialogic Imagination: Four Essays. The University of Texas Press, Austin and London (1981)
8. Bakhtin, M.M.: Problems of Dostoevsky's Poetics. University of Minnesota Press, Minneapolis (1984)
9. Koschmann, T.: Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration. In: International Conference on Computer Support for Collaborative Learning (CSCL'99), pp. 308–313. ISLS, Palo Alto (1999)
10. Trausan-Matu, S.: Automatic support for the analysis of online collaborative learning chat conversations. In: 3rd International Conference on Hybrid Learning, Vol. LNCS 6248, pp. 383–394. Springer, Beijing, (2010)
11. Dascalu, M., Trausan-Matu, S., McNamara, D.S., Dessus, P.: ReaderBench—automated evaluation of collaboration based on cohesion and dialogism. Int. J. Comput.-Support. Collab. Learn. **10**(4), 395–423 (2015)
12. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
13. Trausan-Matu, S., Stahl, G., Zemel, A.: Polyphonic Inter-animation in Collaborative Problem Solving Chats. Drexel University, Philadelphia (2005)
14. Trausan-Matu, S., Dascalu, M., Rebedea, T.: PolyCAFe—automatic support for the polyphonic analysis of CSCL chats. Int. J. Comput.-Support. Collab. Learn. **9**(2), 127–156 (2014)
15. Trausan-Matu, S.: The polyphonic model of collaborative learning. In: Mercer, N., Wegerif, R., Major, L. (eds.) The Routledge International Handbook of Research on Dialogic Education, pp. 454–468. Routledge, Abingdon, UK (2020)

16. Trausan-Matu, S., Stahl, G., Sarmiento, J.: Supporting polyphonic collaborative learning. E-serv. J. Indiana Univ. Press **6**(1), 58–74 (2007)
17. Dascalu, M., Trausan-Matu, S., Dessus, P.: Voices' inter-animation detection with ReaderBench—modelling and assessing polyphony in CSCL chats as voice synergy. In: 2nd International Workshop on Semantic and Collaborative Technologies for the Web, in conjunction with the 2nd International Conference on Systems and Computer Science (ICSCS), pp. 280–285. IEEE, Villeneuve d'Ascq, France (2013)
18. Mukherjee, P., Leroy, G., Kauchak, D.: Using lexical chains to identify text difficulty: a corpus statistics and classification study. IEEE J. Biomed. Health Inform. **23**(5), 2164–2173 (2018)
19. Jayarajan, D., Deodhare, D., Ravindran, B.: Lexical chains as document features. In: Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I (2008)
20. Miller, G.A.: WordNet: a lexical database for English. Commun. ACM **38**(11), 39–41 (1995)
21. Kipfer, B.A., Chapman, R.L.: The Concise Roget's International Thesaurus. HarperTorch (2003)
22. Ruas, T., Ferreira, C.H.P., Grosky, W., de França, F.O., de Medeiros, D.M.R.: Enhanced Word Embeddings Using Multi-Semantic Representation Through Lexical Chains. Information Sciences (2020)
23. Dascalu, M., Trausan-Matu, S., Dessus, P., McNamara, D.S.: Dialogism: A Framework for CSCL and a Signature of Collaboration. In: 11th International Conference on Computer-Supported Collaborative Learning (CSCL 2015), Vol. 1, pp. 86–93. ISLS, Gothenburg, Sweden (2015)
24. Ruseti, S., Dascalu, M.-D., Corlatescu, D.-G., Dascalu, M., Trausan-Matu, S., McNamara, D.S.: Exploring dialogism using language models. In: 22nd International Conference on Artificial Intelligence in Education (AIED 2021). Springer, Utrech, Netherlands (Online) (in press)
25. Dascalu, M., McNamara, D.S., Trausan-Matu, S., Allen, L.K.: Cohesion network analysis of CSCL participation. Behav. Res. Methods **50**(2), 604–619 (2018)
26. Somasundaran, S., Burstein, J., Chodorow, M.: Lexical chaining for measuring discourse coherence quality in test-taker essays. In: Proceedings of COLING 2014, the 25th International conference on computational linguistics: Technical papers, pp. 950–961 (2014)
27. Menze, B.H., Kelm, B.M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., Hamprecht, F.A.: A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinform. **10**(1), 1–16 (2009)