THE ANNALS OF "DUNĂREA DE JOS" UNIVERSITY OF GALATI
FASCICLE III, 2015, VOL. 38, NO. 1, ISSN 2344-4738, ISSN-L 1221-454X
ELECTROTECHNICS, ELECTRONICS, AUTOMATIC CONTROL, INFORMATICS

RHETORICAL STRUCTURE ANALYSIS FOR ASSESSING COLLABORATIVE PROCESSES IN CSCL

Mohammad Hamad Allaymoun[1], Stefan Trausan-Matu [1, 2, 3]

[1]*University Politehnica of Bucharest, Computer Science Department, Bucharest, Romania*

[2]*Romanian Academy Research Institute for Artificial Intelligence, Bucharest, Romania*

[3]*Academy of Romanian Scientists*

Abstract: This paper presents a research on using rhetorical structures for assessing collaborative processes in Computer-Supported Collaborative Learning (CSCL) chats. For this purpose, the ideas of Bakhtin's dialogism theory and Trausan-Matu's polyphonic model are used, starting from the identification of the threads of repeated words from chats. Cue phrases and their usage in linking the identified threads are also considered. The results are presented in statistical tables and graphics that ease the understanding of the collaborative process, helping teachers to analyze and assess students' collaborative chats. It also allows students to know and understand the interactions and how it contributes to the conversation.

Keywords: Computer Supported Collaborative Learning; Collaboration Assessment; Chat Conversations; Rhetorical structures; Cue phrases.

1. INTRODUCTION

Computer-Supported Collaborative Learning (CSCL) is a learning paradigm in which students are grouped together and learn through discussions benefiting from the assistance of computer tools (Allaymoun, 2014; Stahl, 2006; Trausan-Matu, 2010). In fact, the increasing use of the Internet and information technology (IT) contributed mainly to the development of CSCL and led the researchers to further concentrate in this field. Therefore, CSCL gains a broader usage as a viable alternative to classic educational scenarios (Dascalu, *et al.*, 2014; Stahl, 2006; Trausan-Matu, 2010; Trausan-Matu, *et al.*, 2014). CSCL provides both learners and teachers with effective educational technology tools enabling them to achieve the educational goals within the technical framework. Moreover, there are tools that help powerfully in the generation of ideas and participation among learners and teachers. These tools also contribute to configure virtual classrooms for learners, so as to allow the sharing of knowledge and interaction among themselves to achieve several goals including: lessons' explanation, educational discussion, solving exercises and homework, etc. This collaborative process is a way in which co-learners exchange information, discuss different perspectives, take on diverse roles and coordinate their efforts in solving a joint task by using computer-based tools (Allaymoun, 2014; Stahl, 2006).

Undoubtedly, chat is regarded as one of the favorite tools in CSCL, which requires online synchronous textual interaction between students. It is also used in CSCL for building the collaborative knowledge in virtual learning environments (Stahl, 2006). Two of the most encouraging reasons for using chat are the easiness to deal with it by participants as well as its availability at any time and in any place.

This paper was recommended for publication by Viorel Minzu

Consequently, CSCL needs tools that help to achieve the assessment of the conversation. Effectively, the developed model analyzes the chats relying on the Mikhail Bakhtin's ideas of dialogue and polyphony (Bakhtin, 1984; Bakhtin, 1993; Bakhtin, 1986). and Stefan Trausan-Matu's polyphonic model (Trausan-Matu, *et al.*, 2005; Trausan-Matu, 2010a; Trausan-Matu, 2010b; Trausan-Matu, *et al.*, 2014).

These ideas are applied in the model by considering the most frequent words that constitute threads discussed in the chat and the links between them made by cue phrases ("but", "so", "nevertheless", "and", ...) (Allaymoun, 2015). This paper attempts to design and implement an integrated content analysis toolkit. For doing so, it relies on incorporating the quantitative statistics for participation and rhetorical structure analysis, chat analysis and graphics support, so as to facilitate assessing the collaborative process in CSCL.

2. THE POLYPHONIC MODEL AND THE RHETORICAL STRUCTURE

Dialogue is the process by which two or more participants discuss a specific topic, so that each speaker seeks to inform the recipient of his ideas. This action allows the speaker to choose consciously or unconsciously the words that best fit his goals in order to communicate and transfer his idea to all parties. Dialogue has been considered by Mikhail Bakhtin as present in any human language activity. He even raises the idea of dialogism to a fundamental philosophical category: "Any true understanding is dialogic in nature" (Voloshinov, 1973).

In addition, Bakhtin's theory also tries to explain how ideas are passed on through different voices. In fact, passing ideas through dialogue means repeating words, and the repetition of certain words may be considered that it reflects the importance with which the speaker wishes to transfer his ideas to the other participants.

Stefan Trausan-Matu's polyphonic model adopted the ideas of Mikhail Bakhtin's dialogic theory. It uses tools specific to Natural Language Processing (NLP) for the analysis of conversation. As a matter of fact, one important idea of this model is the extraction of repeated words, as it makes easier to detect threads of discussion in the conversation and that can be used effectively for detecting the structures specific to the polyphonic model of CSCL.

The Rhetorical Structure Theory (RST) is a way to describe the structure of texts and it has been found applicable to a wide variety of text's types. Moreover, RST can be used to analyze texts with the goal to assign roles to each part of the text. These roles are also incorporated in a method that links each part into a larger part, which will finally lead to link the whole text into one section (Mann, and Thompson, 2000). De facto, among the various characteristics of RST is that it deals directly with the coherence relations instead of corresponding linguistic expressions. RST analysis is based on the assumption that some units in the text are more central than the rest of the text, whereas the other units are given to support the reader's belief in them. Thus, the central units are labeled as nuclei, while the supporting units are labeled as satellites (Mann, and Thompson, 1983; Mann, and Thompson, 1985; Mann, and Thompson, 1992).

Table1: Set of pairs of Cue Phrases and Rhetorical Relation.

Rhetorical Relation	**Cue phrases**
Conjunction	thus, and, as well, at the same time
Contrast	however, but, despite, yet, nevertheless, nonetheless
Extending	further, in addition, this is, so, also, nor
Sequence	first, second, third, next, later
Explanation	because, since, such as, in order to, although, accordingly
Exemplify	for example, example
Conclusion	hence, accordingly, in conclusion, consequently
Amplify	moreover, that is to say, in particular, in fact

Cue phrases are words and phrases that specifically indicate the presence of rhetorical structures and, in general, of discourse structure in both text and speech. Additionally, cue phrases can be used as a sufficiently exact indicator of the boundaries between textual units and of the rhetorical relations that are shared between them (Marcu, 2000). Therefore, each cue phrase signals a rhetorical relation between the two units. Table 1 shows a set of cue phrases and the associated rhetorical relation (Allaymoun, 2015).

3. THE PROPOSED MODEL

In this section, we explain the model's design in more details. We also explain the basic processes and stages designated to complete the task. Fig. 1 shows the model's design of collaborative assessment, which it mainly comprises three components.

3.1. Preparation component

The preparation component allows the users to import data in an XML format (Trausan-Matu, 2010) and then prepare the data of chat before analysis, considering: the chat utterance number, the speaker nickname, the speaker text, the referenced utterance text (if it exists), the number of the referenced utterance, and the time stamp.

A standard format for the content of chat is generated from different input formats for chats. The purpose is

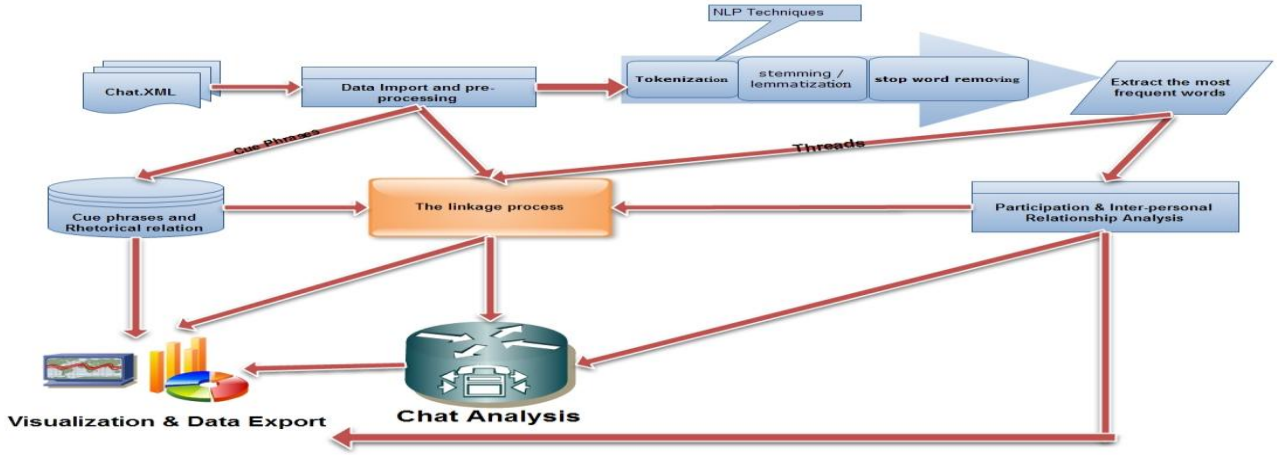

Figure 1. Design of the Proposed Model.

to be able to process various formats of chats, which, after the standard format is obtained, will be analyzed and data will be extracted.

The final goal of this component is to extract from the chat discussion threads cue phrases and the types of rhetorical relations and store them finally in a database. For these aims, we are using NLP techniques (tokenization, stemming and lemmatization, and stop words removing – we used Stanford NLP tools - http://nlp.stanford.edu) (Allaymoun, 2014), in order to identify the discussion topics from the most frequent words (after stemming and eliminating stop words). Repeated words-topics are indicators of the discussion threads in the chat (Trausan-Matu, 2010 a; Trausan-Matu, 2010 b; Trausan-Matu, *et al.*, 2014). Ultimately, the output of this component is as follows: standardized chat format, cue phrases, rhetorical relation types, and threads.

3.2. Component's Analysis

This module is based on several analyses of the data coming from the previous component to get relevant results for a study. This helps in effective collaborative and automatic assessment.

3.2.1 Linkage Process

This process is linking the discussion threads by the cue phrases, both extracted from the chat by the previous component. So, the process of linking between threads and cue phrases in the chat is as follows:

1. Linking threads that appear within one utterance of the chat;

2. Linking threads that appear in sequences of speakers' utterances;

3. Linking the threads that appear in the utterance text and in its references.

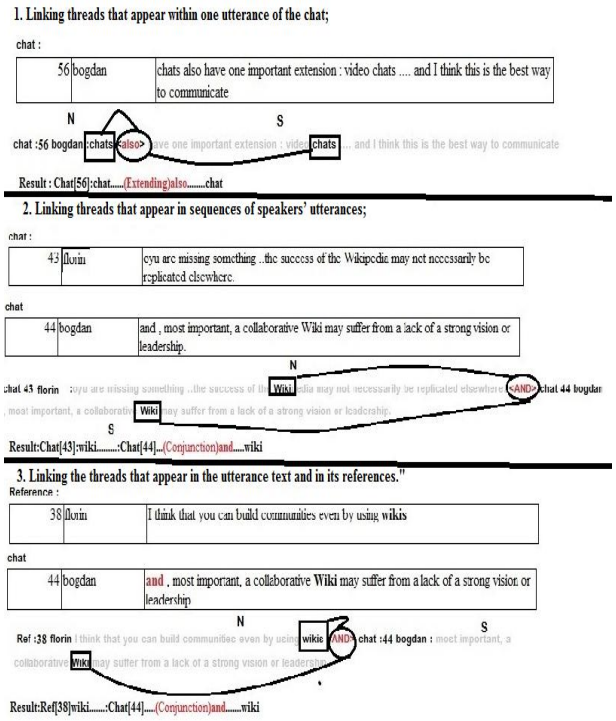

Figure 2. Linking Threads by Cue Phrases.

Fig. 2 shows examples of how threads are linked by cue phrases. Thus, the output of the process is the associative relationships between threads that were discussed among the participants. Accordingly, the idea of this model is to assess collaboration by identifying the links of all participants' discourse threads by rhetorical structures.

3.2.2 Participation and Inter-personal Relationship Analysis

The model provides basic statistical results for each participant (thread no., threads, utterance no.,

references no.). In fact, knowing the threads and the frequency of any participant utterances in chat helps to analyse the individual impact in knowledge building. Therefore, the analysis of the individual contributions for each participant helps to know his contribution's volume comparing to the volume of the chat.

3.2.3 Chat Analysis

Chat Analysis is one of the most important stages in this component, since here is performed the analysis of data collected from the previous phases. In this stage, it is the time to compute and integrate the results and relationships of the data collected (collaborative, individual, and threads). The model uses typical statistical formulas such as the mean, variance, standard deviation, correlation, etc. The purpose is to provide as much data as possible in order to facilitate studying and analyzing the information for learners and teachers. Besides, this also helps the collaborative and individual assessment in chats.

3.3. Visualization and Data Export component

This component consists of several modules devoted to provide data visualization (graphs, statistical tables) emanated from the analysis's results and to export the multiple analysis results in the format of tables for facilitating users to explore the collaborative and individual contributions and the collaborative regions of chats. For example, the component displays the statistical results about chats such as correlations, descriptive data, statistical data, rank and percentage. Also, graphs indicate individual and collaborative contributions and threads. In addition, they allow displaying the collaborative assessment of chat or group chat as well as displaying the comparison's results between chats.

4. DISCUSSION

In this section, we will provide examples of how the model components in chat analysis contribute to get statistical results and graphical presentations. In addition, the model helps to analyze the results of collaborative learning among the participants in the chat and assessment.

The model used the students' chats performed in a research conducted in the K-Teams Laboratory at University Politehnica of Bucharest. These chats consisted of groups of 4 to 5 participants discussing the advantages and disadvantages of CSCL technologies at the Human-Computer Interaction course they attended. Also, each student had to focus on one specific collaborative technologies (chat, blog, wave, wiki, and forum) (Trausan-Matu, 2010 b; Trausan-Matu, *et al.*, 2014; Dascalu, *et al.*, 2015a; Dascalu, *et al.*, 2015b). A chat environment that facilitates the referencing of previous utterances was used. The benefit of using these chats carried out in a collaborative environment in our model is to obtain results easier to study and assess collaboration.

For this section we are considering the results emanated from the analysis of one chat and discuss the results in detail. In doing so, the purpose is to assess the collaborative and individual participation as well as the follow-up the work of all model's components in order to discuss the results of these components and review its outputs. Table 2 shows the results of a chat of 5 participants after having the initial analysis. This table also presents the threads discussed in the chat (blog, forum, chat, wave, wiki) extracted by the NLP techniques, the number of interventions, the number of references to previous utterances and the number of repeated words (number of utterances), that are forming threads per each participant. By these results, we illustrate the involvement of each participant, so that any participant can see his (her) contribution to the chat.

Table 2: Chat statistics

Participant involvement and interaction			Threads in chat					
Participant Name	No. Interventions	No.Reference	forum	wiki	chat	blog	wave	Total
Daniil	124	102	11	24	18	7	9	69
Dragos	90	83	4	4	9	2	8	27
Bobricl	78	70	9	1	15	2	0	27
Flavius	32	26	4	1	4	3	1	13
Andrei	92	43	6	3	13	0	0	22
Total	416	324	34	33	59	14	18	158

Table 3 presents the results obtained from the process of linking threads by cue phrases and also indicates the number of repeated words in threads identified by the linking process. Consequently, these results support the assessment of collaboration using the model we considered. These linkages may also further be used for the identification of rhetorical relations. In the chat we analyzed, the identified rhetorical relations were as follows: Conjunction (18), Contrast (7), and Extending (3).

Table 3: results obtained from the process of linking threads

Participant involvement and interaction			Threads in linkage					
Participant Name	No. Interventions	No.Reference	forum	wiki	chat	blog	wave	Total
Daniil	124	102	3	10	7	2	5	27
Dragos	90	83	0	0	2	0	2	4
Bobricl	78	70	6	0	3	0	4	13
Flavius	32	26	3	1	3	1	0	8
Andrei	92	43	2	1	3	0	0	6
Total	416	324	14	12	18	3	11	58

Fig. 3 is a diagram showing the evaluation of the contribution of each participant in the discussion in terms of the number of words (utterances) in threads

before and after the linkage process (that means before and after considering linkages by cue phrases).

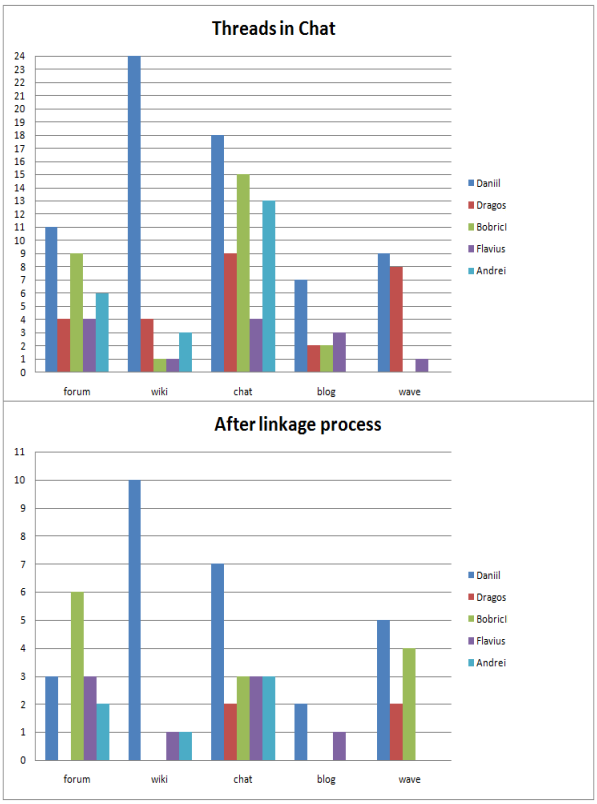

Figure 3. Contribution of each participant in the discussion according to the number of threads.

Fig. 4 shows the graphical representation that reveals the results of the chat analysis. It displays the sequence of appearances of utterances in threads in chat for all participants before the process of linking. In addition, this figure indicates the threads connected with each other by cue phrases. for all participants. As well, by this figure several outcomes can be concluded:

1. The collaborative process among the participants starts at utterance No 63 and ends at utterance No 397,

2. Collaborative density (collaboration regions) in chat [utterance no.83, utterance no.124] [utterance no.226, utterance no.277].

The statistical analyses for the conversation being discussed are on two axes: the first axis is to assess individual participation by calculating correlations between participant's contributions and the average of threads for each participant. Thus, the Correlations and Individual contributions vs. No. of interventions equals (0.97). This indicates the ratio of correlation between the threads and the participants in the considered chat, which is evidenced of the participant's involvement. The second axis is the collaborative assessment by calculating correlations between participant's contributions and threads resulting from the linkage process. Consequently, the correlations and number of threads in linkage process vs. Number of threads in interventions equals (0.99).Therefore, the correlation between threads in chat and threads in linking process indicates the percentage of collaborative assessment.

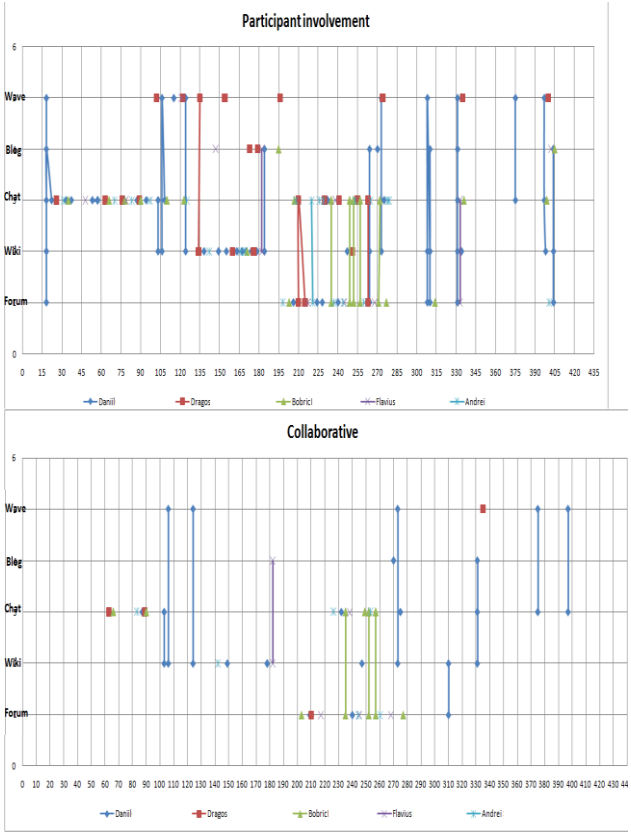

Figure 4. The Results of the Chat Analysis.

5. CONCLUSIONS

This paper presented an implementation of the polyphonic model for analyzing the content of chats by considering cue phrases that indicate rhetorical structures, in order to help in assessing the collaborative process in CSCL chat, additionally to the benefits of the integration of existing tools to get indicative results. In general, the most important functions of the chat is to exchange and share ideas among participants, which is an indicator for collaboration. Accordingly, the model relies on the idea of linking themes together among participants by rhetorical structures, by which they can estimate the collaboration in chat and view the results of chat analysis graphically and statistically. Thus, the collaboration and participant's involvement can be studied.

The model uses NLP tools for analyzing the chat texts without knowing in advance the topics of the discussions. These topics are identified starting from

repeated word stems, after eliminating stop words and are then used to extract and identify threads automatically.

The model provides teachers with information (graphs and statistics) to help them know the main threads discussed in the chat and for collaboration assessment. In addition, the model presents the sequence of discussion threads among the participants. As for the students, they can know the extent of their contribution and participation in the discussion of threads in chat.

The novel idea of the present paper is to extend the implementation of the polyphonic mode (Trausan-Matu, *et al.*, 2005; Trausan-Matu, 2010a; Trausan-Matu, 2010b; Trausan-Matu, *et al.*, 2014). by considering rhetorical structures. Previous implementations of it (Trausan-Matu, 2010a; Trausan-Matu, 2010b; Trausan-Matu, *et al.*, 2014; Dascalu, *et al.*, 2014; Dascalu, *et al.*, 2015a; Dascalu, *et al.*, 2015b) detected in a similar way topics and discussion threads and performed some other processing on them.

6. REFERENCES

Allaymoun, H.M., (2014), Analysis of Discourse Threading and Inter-animation in CSCL Chats, *Conference on eLearning & Software for Education*, Issue 1.Romania.

Allaymoun, H.M.,(2015), Graphical Visualization of Rhetorical Structures in Chats, *Proceedings of the 11th International Scientific Conference eLearning and software for Education*, Bucharest.

Bakhtin, M.M., (1984), *Problems of Dostoevsky's Poetics*, Theory and History of Literature Series, vol. 8, Minneapolis, (translated by C. Emerson).

Bakhtin, M.M., (1993), *Problems of Dostoevsky's Poetics*. University of Minnesota Press, Minneapolis.

Bakhtin, M.M., (1986), *Speech Genres & Other Late Essays*, University of Texas Press, Austin.

Dascalu, M., Trausan-Matu, S., and Dessus, P, (2014), Validating the Automated Assessment of Participation and of Collaboration in Chat Conversations. In S. Trausan-Matu, K.E. Boyer, M. Crosby and K. Panourgia (eds.), *ITS 2014*, Springer, Honolulu, USA, 230–235.

Dascalu, M., Trausan-Matu, S., Dessus, P., & McNamara, D.S., (2015a), Dialogism: A Framework for CSCL and a Signature of Collaboration. *In 11th Int. Conf. on Computer-Supported Collaborative Learning (CSCL 201*5). Gothenburg, Sweeden: ISLS.

Dascalu, M., Trausan-Matu, S., Dessus, P., & McNamara, D.S. (2015b), Discourse cohesion: A signature of collaboration. *In 5th Int. Learning Analytics & Knowledge Conf. (LAK'15).* Poughkeepsie, NY: ACM.

Mann, William C. and Sandra A. Thompson, (1983), *Relational Propositions in Discourse*, ISI: Information Sciences Institute of University of Southern California, Los Angeles, ISI/RR-83-115, 1-28.

Mann, William C. and Thompson, S.A., (1985), Assertions from Discourse Structure, in (eds,). *Proceedings of the Eleventh Annual Meeting of Linguistics Society*, Berkeley, CA: Berkeley.

Mann, William C. and Sandra A. Thompson, (1992), *Relational Discourse Structure: A Comparison of Approaches to Structuring Text by 'Contrast'* In S. J. Hwang and W. R. Merrifield (eds,). Language in Context: Essays for Robert E. Longacre Dallas: SIL.

Mann, William C. and Sandra A. Thompson, (2000), Toward a theory of reading between the lines: An exploration in discourse structure and implicit communication. the *7th IPrA International Pragmatics Conference*.

Marcu, D., (2000), *The Theory and Practice of Discourse Parsing and Summarization*. MIT Press, Cambridge, MA, USA.

Spada, H., Meier, A., Rummel, N., Hauser, S., (2005), A new method to assess the quality of collaborative process in CSCL, *Proceedings of the International Conference on Computer-support for Collaborative Learning*.

Stahl, G. (2006), *Group cognition. Computer support for building collaborative knowledge*. Cambridge: MIT Press.

Trausan-Matu, S., Stahl, G., and Zemel, A., (2005), Polyphonic Inter-animation in Collaborative Problem Solving Chats, Research Report, Drexel University, Philadelphia.

Trausan-Matu, S., (2010a), The Polyphonic Model of Hybrid and Collaborative Learning, In Wang, F.,L., Fong., J., Kwan, R.C., *Handbook of Research on Hybrid Learning Models: Advanced Tools, Technologies, and Applications*, Information Science Publishing, Hershey, New York, pp 466-486.

Trausan-Matu, S., (2010b), Automatic Support for the Analysis of Online Collaborative Learning Chat Conversations, in Philip M. Tsang, Simon K. S. Cheung, Victor S. K. Lee, Ronghuai Huang (Eds.): *Hybrid Learning, Third International Conference, ICHL 2010*, Beijing, Lecture Notes in Computer Science 6248, Springer, pp. 383-394

Trausan-Matu, S., Dascalu, M., & Rebedea, T, (2014), PolyCAFe – Automatic support for the analysis of CSCL chats. *International Journal of Computer-Supported Collaborative Learning*, 9(2), 127–156.

Voloshinov, V. (1973), *Marxism and the Philosophy of Language*, New York Seminar Press.