Cohesion-based Analysis of CSCL Conversations: Holistic and Individual Perspectives

Mihai Dascalu, Stefan Trausan-Matu, Philippe Dessus

HAL Id: hal-01470952
https://hal.science/hal-01470952

Submitted on 18 Feb 2017

Cohesion-based Analysis of CSCL Conversations: Holistic and Individual Perspectives

Mihai Dascalu, Stefan Trausan-Matu, University "Politehnica" of Bucharest,
313 Splaiul Independentei, 060042 Bucharest, Romania
Email: mihai.dascalu@cs.pub.ro, stefan.trausan@cs.pub.ro
Philippe Dessus, LSE, UPMF Grenoble-2 & IUFM–UJF Grenoble-1, France,
philippe.dessus@upmf-grenoble.fr

Abstract: Although Computer Supported Collaborative Learning (CSCL) technologies have gained an increasing role in educational environments, there are few automatic systems that address involvement, knowledge-building and collaboration in order to support tutors in the time consuming process of analyzing conversations. We propose a cohesion-based analysis model integrating multiple natural language techniques, an intervention scoring mechanism and a comprehensive collaboration assessment method, derived from social knowledge-building, reflected at utterance level through cohesion. Furthermore, by combining a holistic perspective of the entire conversation with a more fine grained view focused on each participant, we obtain a thorough evaluation of chat conversations with focus on topics modeling, participant interaction and collaboration. In order to sustain our model, we have performed a preliminary validation study that proves that our analysis is consistent with tutor evaluations.

Introduction

Learners collaborating through forums or chats use language as a mediator for building knowledge, more specifically for developing one's view about the material, for negotiating meaning with each other and for defining new cultural artifacts (Stahl, 2006). The activity of reading is the cornerstone of every CSCL situation: learners are typically engaged in reading to acquire new pieces of knowledge; to be aware of the way their peers get acquainted about this knowledge; to understand what is required by their teachers or tutors. They can even carefully re-read their own utterances before and after posting.

Cohesion, i.e., relations of meaning within a text, is one of the most important features driving the understanding of read texts (Tapiero, 2007). Cohesion gaps within textual materials, e.g., online conversation, may yield comprehension difficulties, mostly for low-knowledge learners (McNamara et al., 1996), but also be viewed as topic change along discussion threads (Dascalu et al., 2010a). However, the cohesion-based assessment of learning materials or learners' utterances is a very demanding task and systems that automatically analyze utterances during learning sessions are of particular interest (Fujita & Teplovs, 2009; Trausan-Matu et al., 2007). Such systems often make an intensive use of Latent Semantic Analysis (Landauer & Dumais, 1997), which has been proven to capture cohesion-based textual relationships (Graesser et al., 2004).

In this paper we introduce *ReaderBench*, a multi-purpose system that proposes several cohesion-based views on the learners' contributions to the discussion and presents them upon two perspectives: *individual*, centered on each participant, and *holistic*, covering the entire discussion. The first perspective allows learners to individually understand the extent to which they contribute to a given discussion, while the latter is useful for teachers and tutors to assess its fruitfulness in terms of knowledge-building and collaboration.

Moreover, our aim is to provide a comprehensive approach for *assessing collaboration* throughout a multi-participant chat conversation, as a refinement of the initial information gain assessment of collaboration (Dascalu et al., 2010b; Trausan-Matu et al., 2012). Our implemented model automatically identifies intense collaboration zones and extrapolates the concepts of personal and social knowledge-building (Bereiter, 2002; Scardamalia, 2002; Stahl, 2006) to a finer grained dimension, more specifically to each utterance. In this particular context, cohesion is the bridge between utterances, enabling information transfer and cumulative knowledge-building from both *personal* (utterances from the same speaker) and *social* perspectives (utterances from different speakers). Additionally, our model of analysis goes one step further towards defining an implemented *polyphonic model* of a discussion (Trausan-Matu et al., 2007) by also considering an analogy to key dialogic concepts (Bakhtin, 1986). In order to provide a quick equivalence that will be detailed later on, *topics* emerging from the conversation and pertaining to different speakers are actual *voices*, whereas *cohesion* induces the *echo cancelation* effect of each utterance upon later interventions within the discussion.

The following section provides an insight into our analysis model, with emphasis on building the cohesion graph and on the multi-layered scoring mechanism. Afterwards, the individual assessment section is centered on the identification of topics and modeling the participants' interaction throughout the conversation. Later on, we shift the point of interest towards the assessment and evolution of collaboration. Eventually, a preliminary validation experiment is presented and the relevance of the results is debated.

Cohesion-based Conversation Analysis

Cohesion, seen as the relationship at lexical and semantic levels between two segments of texts, is defined within our model in terms of *recency*, *lexical proximity* and *semantic distances*. Therefore, cohesion is a symmetric function that combines: 1/ the inverse distance between the analyzed textual elements within a fixed window of analysis that defines a local context, 2/ lexical proximity as a relaxed condition of having identical stems within both text fragments and 3/ sematic distances determined from three perspectives.

The first semantic dimension of cohesion is represented by semantic distances within lexicalized ontologies. As *ReaderBench* was designed to cover both English and French, we used besides *WordNet* (Miller, 1995), a transposed and serialized version of *WOLF* (*Wordnet Libre du Français*, http://alpage.inria.fr/~sagot/ wolf.html) in order to measure similarity between concepts. Our experiments showed that Wu-Palmer semantic distance performs best as it scales correspondingly to the [0; 1] interval and it works properly for both English and French languages. Although WOLF is rather incomplete and mixes word glosses, expressing them partially in English and partially in French, it is currently the best alternative for having a multi-lingual analysis platform.

Going deeper within our cohesion estimation function, the other two semantic dimensions are represented by Latent Semantic Analysis (LSA) (Landauer, & Dumais, 1997) and Latent Dirichlet Allocation (LDA) (Blei et al., 2003). Whereas LSA is used for measuring similarity between discourse fragments through cosine similarity after projecting concepts in a pre-trained vector space, LDA exploits the Jensen-Shannon dissimilarity (Schütze & Manning, 1999) of posterior topic distributions in order to grasp inconsistencies or discrepancies between texts. Therefore, from a computational point of view, we have combined Information Retrieval specific techniques, mostly reflected in word repetitions and normalized number of occurrences, with semantic models (LSA and LDA) in order to provide a comprehensible and measurable value of cohesion between textual fragments.

After defining the cohesion function between different analysis units within the discourse, the next step consists of automatically building the cohesion graph, which plays a central role within our analysis. We propose a multi-layered graph consisting of the following hierarchy: *document* (entire chat conversation or reading material) > *block* (utterance or paragraph within the initial text) > *sentence*. Within the previous structure, we enforce hierarchical links determined by inclusion functions (sentences within a block, blocks within the document) and mandatory links obtained from: 1/ *explicit links* defined by participants within the user interface – in our case, *ConcertChat* – or 2/ *adjacent blocks* of text. Later on, cohesion is determined between all sentences of each block (intra-block cohesion) and between blocks within a predefined sliding window of analysis (inter-block cohesion measurements that within our experiments considered a window of 20 utterances). The resulting links were pruned after the sum of mean and standard deviation of cohesion values in order to keep only relevant links within the discourse model.

Our aim was to provide a generalized and customizable model for assessing different types of discourses in terms of cohesion: multi-participant chat conversations, on one hand, and texts in general, on the other. More specific to CSCL conversations, we opted for Dong's perspective (1995) of separating utterances based on turn-taking events between speakers. Although most participants' utterances consist of one sentence (in some cases, elliptical sentences are quite frequent), we preferred to create a scalable model, which can be later on easily adjusted to forums, for example, in which interventions entangle multiple sentences.

An important component in the evaluation process of each participant and of collaboration throughout the conversation is our bottom-up intervention scoring method. Firstly, specific natural language techniques are applied on each utterance: tokenizing, splitting, part of speech tagging, parsing, stop words elimination, dictionary-only words selection, stemming, lemmatizing, named entity recognizer and co-reference resolution (Manning & Schütze, 1999). Afterwards, each sentence is assigned an individual score of each remaining concept equal to its term frequency – inverse document frequency (Manning & Schütze, 1999) multiplied by its cosine similarity from the LSA vector space to the corresponding sentence, denoting the local significance of each word. Furthermore, cohesion with a generic title or with a list of predefined key concepts is also measured and added to the score of each sentence. Due to the nature of our experiments conducted in an academic environment, each conversation had a specific set of concepts, defined by the tutor, which had to be covered by the participants. In other words, we measure through cohesion the degree in which each participant addresses each concept and whether the conversation stays on-topic.

Secondly, at utterance or block level, individual sentence scores are weighted by cohesion measures and summed up in order to define the intra-block score. By going further into our discourse decomposition model (document > block > sentence), inter-block cohesion values are used to augment the previous intra-block scores by also considering all the links from the previously defined cohesion graph. In the end, all block scores are combined in order to determine the overall conversation score. Figure 1 depicts the scores for each intervention in brackets, immediately after each participant's intervention. The proposed scoring mechanism is effective because it considers the local importance of each intervention and it also combines these individual scores through cohesion to other semantically related elements within the discourse structure.

Individual Participant Assessment

Our aim regarding the individual assessment of participants is to focus on covered topics and interaction modeling. *Topics*, seen as key concepts from the conversation, play a leading role in obtaining a view of the overall discussion, but also in observing each participant's points of interest. Tightly connected to the cohesion graph and the previous scoring mechanism, topics are extracted per entire conversation and per participant. The *relevance* of each concept mentioned within the discussion is determined by combining a multitude of factors: 1/ individual *term frequency – inverse document frequency* of each word, 2/ *semantic similarity through LSA* to the entire conversation, corresponding utterances and sentences 3/ *cosine similarity* between the distribution of the word in the predetermined topics from *LDA* and the conversation's overall distribution over the same classes and 4/ the *scores* of each *sentence* and *utterance* in which the word is present. Within this evaluation, the initial natural language processing, with emphasis on lemmatization, played an important role in standardizing the inputs. Moreover, the scores of each concept are normalized by the length of the lexical chains in which they occur; the previous lexical chains are determined from the disambiguation graph modeled through semantic distances from *WordNet* and *WOLF* (Galley & McKeown, 2003). In addition, as an empirical improvement, filtering the previous list of topics by corresponding parts of speech and selecting solely nouns provided more accurate results in most cases, as nouns tend to better grasp the conceptualization of the discussion.

In this context, key topics together with their corresponding lexical chains can be considered voices that spread throughout the conversation, while cohesion simulates echoes of voices to other inter-linked utterances. Moreover, by changing the focus on a specific participant, we can observe the strength of a voice as being directly proportional to the relevance of previously identified topics.

Figure 1 presents the main user interface from *ReaderBench* in which a chat conversation has been loaded from the XML file exported from *ConcertChat*, the scores of each intervention are displayed in brackets and the automatically identified topics for a specific participant are presented in the right sidebar.
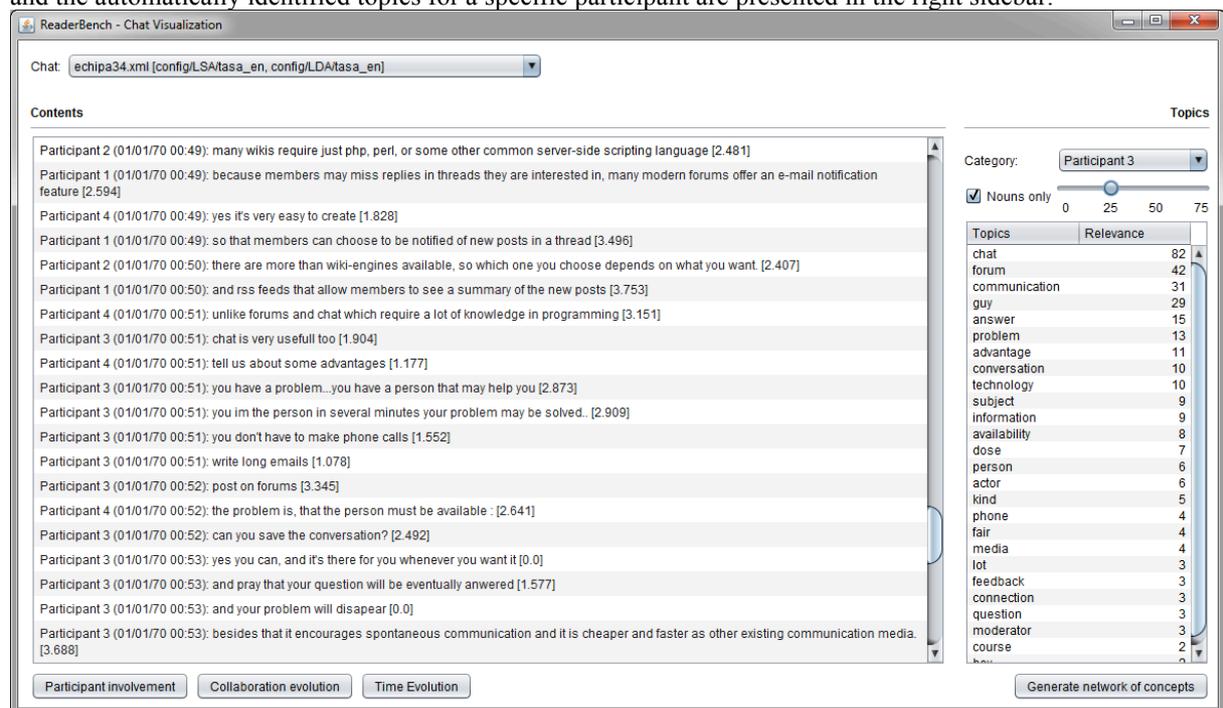
Figure 1. *ReaderBench* Main User Interface.

Following the transition from a global view of discourse to a user centered perspective, the visualization component of the conceptual space for each chat participant as a mind-map is based on cosine similarities computed within our LSA vector space. Terms central to a given discussion may not appear in any utterance but, nonetheless, be worth displaying for comprehension's sake. We thus enriched the previously identified topics list with *inferred concepts*, not mentioned within the text, for which three measures are computed and summed in the end: 1/ the similarity to the entire conversation vector, 2/ the similarity to the weighted topics vector determined from the previous list of concepts displayed within the user interface and 3/ the similarity to the closest concept within the text, because tight correlation to a concept already present within the text increases the proximity to the overall discussion. These three dimensions have as empiric argumentation the method in which experts selected inferred concepts from manually assessed conversations: similarity to a concept or idea from the text, but also to the whole text (and the entire text is represented as the document vector and as an extrapolation of the already identified topics).

Figure 2a presents the conceptualization space of a user including initial topics and the automatically inferred concepts. From a computational point of view, we determine all similarities between the previous words, considered nodes within the network, and select as edges solely the ones that are above the threshold selected from the interface (30% in this particular case). The actual size of each node is directly proportional to a Social Network Analysis metric derived from the previous graph (e.g. betweenness). All multidimensional scaling and visualization of the graph as a force-based layout are performed by Gephi (Bastian et al., 2009).
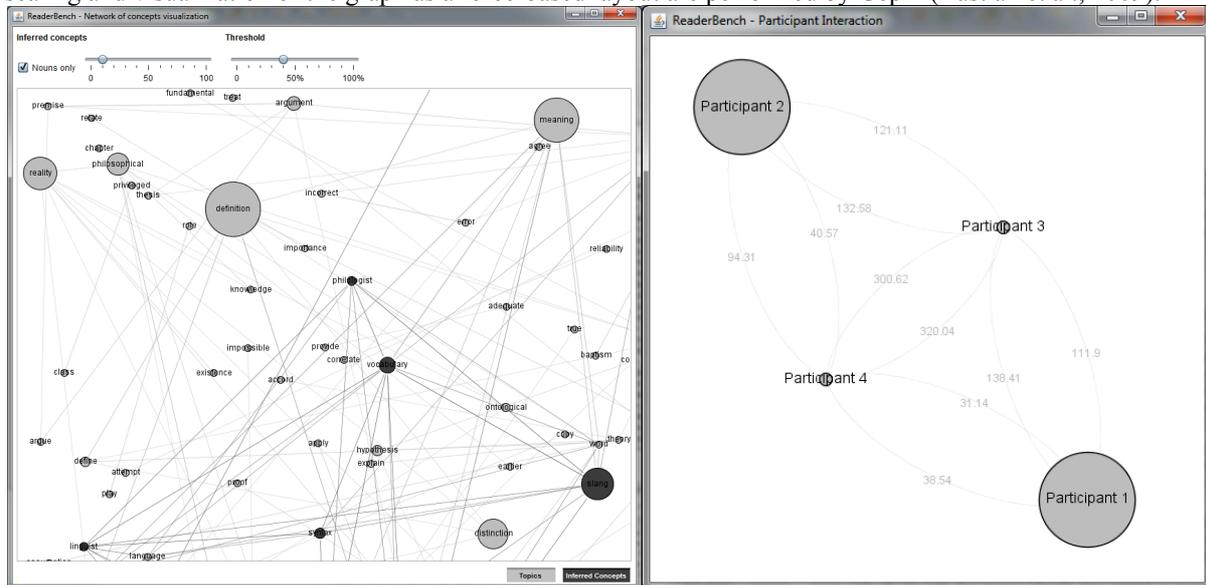
Figure 2. a. Concepts Visualization and b. Individual View of the Interaction Graph.

Besides the identification of topics for each participant in the discussion, *ReaderBench* also supports participant interaction modeling covering two major dimensions: a preliminary *quantitative* one, generated by the actual number of interactions per participant, and a deeper *qualitative* dimension, obtained after summing up intervention scores. Internally, an interaction graph is built with participants as nodes and the weight of links equal to the sum of interventions scores multiplied by the cohesion function with the referred element of analysis, extracted from the cohesion graph. Therefore, by performing social network analysis on the previous participant interaction graph, the scale of analysis is shifted towards an individual perspective (see Figure 2b).

Collaboration Assessment

In order to achieve genuine collaboration, which can be seen as true polyphony in terms of Bakhtin's dialogic theory (1984), the conversation must be generated by a dense intertwining of voices, derived from key concepts of the conversation and covering all participants (Trausan-Matu & Rebedea, 2009). Moreover, by enforcing a personal and social knowledge-building process (Bereiter, 2002; Scardamalia, 2002; Stahl, 2006) at utterance level and by using cohesion as a bridge between utterances seen as an overlap of voices, the actual information transfer through cohesion between units of analysis obtains two valences. Firstly, a *personal dimension* emerges by considering utterances with the same speaker, therefore modeling an inner voice or a continuation of the discourse. Secondly, inter-changed utterances having different speakers define a *social perspective* that models collaboration as a cumulative effect.

Therefore, each intervention now has its previously defined *importance score* and a *knowledge-building (KB) effect*, both personal and social. The personal effect is initialized as the intervention's score, whereas the social effect is zero. Later on, by considering all the links from the cohesion graph, each dimension is correspondingly augmented: if the link is between utterances with the same speaker, the previously built knowledge (both personal and social) from the referred utterance is transferred through the cohesion function to the personal dimension of the current utterance; otherwise, if the pair of utterances is between different participants, the social knowledge-building dimension of the currently analyzed utterance is increased with the same amount of information (previous knowledge multiplied by the cohesion measure). In other words, continuation of ideas or explicitly referencing utterances of the same speaker builds an inner dialogue or personal knowledge, whereas the social perspective measures the interaction with other participants, encourages sharing of ideas, fostering creativity for working in groups (Trausan-Matu, 2010b) and influencing the other participants' points of view during the discussion, thus enabling a truly collaborative discussion.

In this manner we can actually measure collaboration through the sum of social knowledge-building effects, starting from each intervention's score corroborated with the cohesion function. Moreover, personal knowledge-building addresses individual voices (participant voices or implicit/alien voices covering the same

speaker), while social knowledge-building, derived from explicit dialog (that by definition is between two entities), sustains collaboration and highlights external voices. By referring to the dialogic model of discourse analysis, besides voices that are derived from the topics of the conversation in correlation to each participant's point of view, echoes are reflected by cohesion in terms of the information transferred between utterances, whereas the attenuation effect diminishes the strength of the cohesion link with the increase in distance between the analysis elements.

Nevertheless, we must also consider the limitations of our implemented model in terms of personal knowledge–building. Collaboration clearly emerges from social knowledge transfer through cohesion as the influence of one's intervention over other participants' discourse. In contrast, the approximation of personal knowledge–building rather represents an upper bound of the explicitly expressed information transfer between one's personal interventions. Similarly to the gain–based approach (Dascalu et al., 2010b; Trausan-Matu et al., 2012), we use a quantifiable approximation of inner dialogue, although limited in terms of underlying cognitive processes. Personal knowledge–building is seen as a reflection of one's thoughts expressed explicitly within the ongoing conversation as cohesive links between interventions of the same chat participant. But this reflection does not necessarily induce personal knowledge–building, only a cohesive discourse. Therefore, we can consider that the computed value of personal knowledge–building is a *maximum value* of the explicit personal knowledge–building effect, modeled during the discourse through cohesive links.

In addition to the estimation of personal and social knowledge-building effects for each utterance and the modeling of their corresponding evolution throughout the conversation (Figure 3), *ReaderBench* automatically identifies *intense collaboration zones* that are intervals of utterances in which participants are actively involved, collaborate and generate new ideas related to the ongoing context of the discussion. The first step within our greedy algorithm for detecting intense collaboration zones consists of identifying social knowledge–building peaks as maximum local values. Afterwards, each peak is expanded sideways within a predefined slack (experimentally set at 2.5% of the number of utterances). This slack was important due to our focus on the macro–level analysis of collaboration and due to the possible intertwining of multiple discussion threads. In the end, only zones above a minimum spread of 5 utterances are selected as intense collaboration zones. In other words, after identifying the utterances with the greatest social knowledge-building effect, the algorithms expands each zone to the left and to the right, in a non-overlapping manner to previously identified zones, by considering solely utterances above the mean social knowledge-building value and that are in the previously defined slack. If in the end, the zone covers more than the specified minimum spread, it is considered an intense collaboration zone. From a different point of view and highly related to social knowledge-building, cohesion binds utterances within an intense collaboration zone in terms of on-topic relatedness.
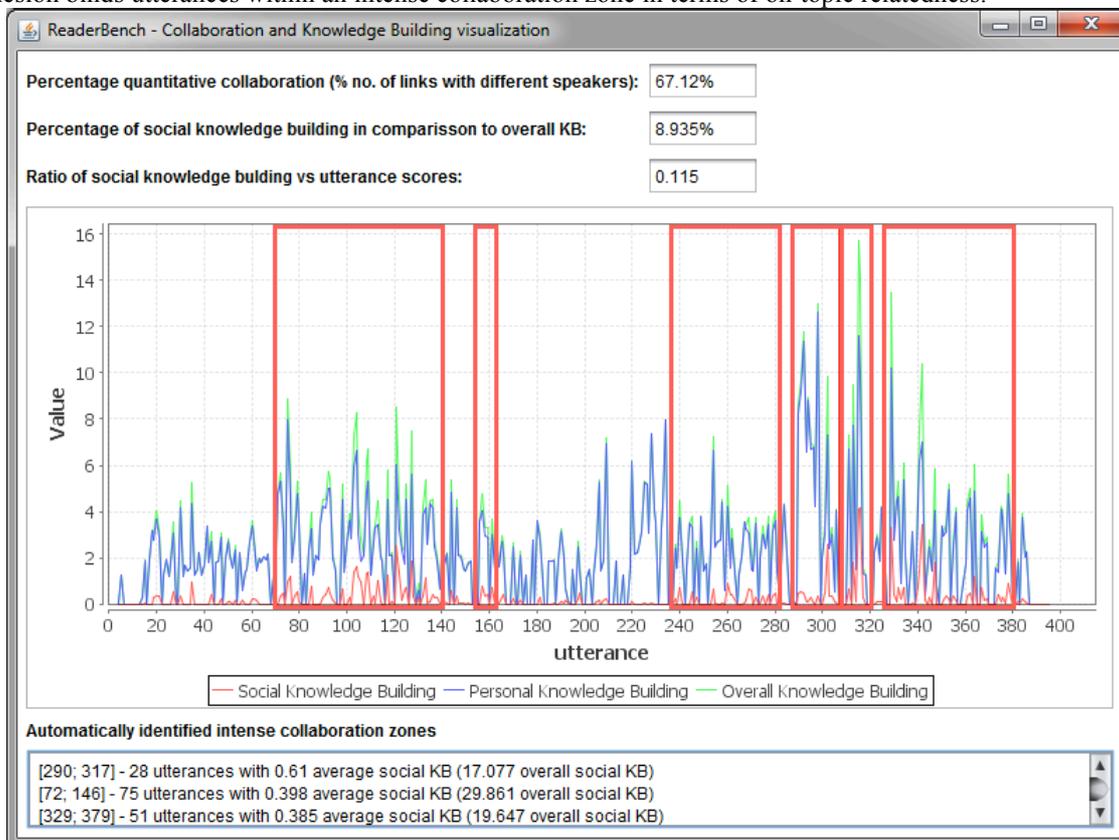

Figure 3. Collaboration Assessment and its Evolution in Time.

From a holistic perspective, three factors were implemented in order to best characterize the overall collaboration within the discussion. Firstly, *quantitative collaboration* is determined as the percentage of links from the cohesion graph having different speakers in comparison to the number of links automatically identified. Although rough as estimation, this measurement provides good insight with regards to the actual information exchange between participants. Secondly, the *overall social knowledge-building score* is compared to the overall knowledge-building effect. Thirdly, the *ratio* between the *overall social knowledge-building score* and the *overall intervention importance scores* is computed for highlighting the amount of information that is transferred through collaboration in comparison to what was withheld initially within each utterance. Low values for the latter factors are normal due to the fact that cohesion, after normalization in the [0, 1] interval, usually integrates low similarity values between utterances of a chat conversation in terms of LSA and LDA.

First Validation Study

Preliminary experiments were conducted in order to validate the dialogic model used for evaluating chat conversations, with emphasis on participant involvement and collaboration assessment. Three chat conversations conducted in an academic environment, with students from the 4th year undergoing the Human-Computer Interaction course and debating on CSCL technologies, were manually assessed by 4 tutors. More specifically, each student had to focus on a CSCL technology (chat, wiki, blog or forum), to present and debate on its benefits in specific use case scenarios generated throughout the conversation. These three conversations (Team 4, Team 34 and Team 36) were selected for detailed analysis after an overview of approximately 50 discussions engaging more than 200 students. Although high discrepancies were noticed in terms of the quality of the content, the involvement and the collaboration of its participants, these conversations were considered representative for the entire sample and the preliminary evaluations were conducted only on these conversations due to the high amount of time it takes to manually assess a single chat conversation (2 to 4 hours for a deep understanding of involvement and of collaboration).

Additionally, the *time evolution modeling* interface depicted in Figure 4 was developed in order to facilitate the manual evaluation of chats in terms of intense collaboration zones. In this context, the presentation of the conversation follows the timeline and models the inter-twining of utterances, based on the cohesion graph. This component is useful for manually identifying: 1/ *breaks within the conversation*, *zones with limited or no collaboration*, due to the fact than within a specific time-frame we have a monologue of a participant, without any interventions from other users, and 2/ *zones with high collaboration* due to the dense inter-animation of utterances between different participants. In the particular case presented in Figure 4, all utterances with identifiers between 30 and 50 belong a single user, within a limited time-frame, therefore making the social knowledge-building effect zero. Afterwards, as multiple participants get involved in the ongoing discussion, collaboration increases.
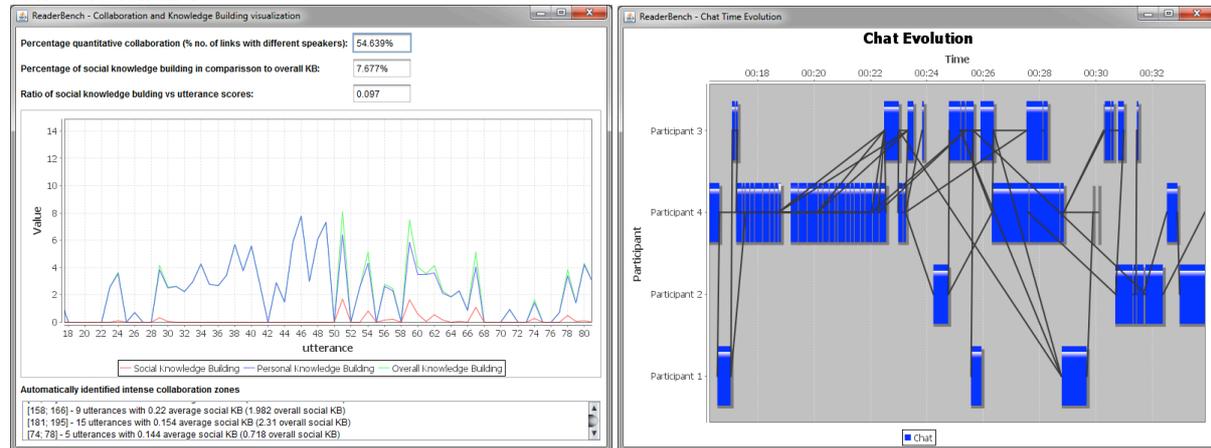

Figure 4. Time slice of a conversation highlighting cohesion links and a monologue within the conversation.

Table 1 presents the correlation between different evaluation factors extracted from *ReaderBench* and the final grades assigned by the experts. Although the participant's identifiers coincide, each conversation had different students attending it. Moreover, in order to ensure the equitability of our analysis, the correlations between the factors automatically determined by the system and the average values of the grades manually assigned by the experts were computed after combining the participants' scores from all conversations.

As an interpretation of the results presented in Table 1, we can observe in Team 4 conversation a discrepancy, as the involvement of the participants from a personal point of view was good, while the actual collaboration throughout the conversation is highly unbalanced. Team 34 conversation has the lowest scores in

all the factors, whereas Team 36 conversation, that was considered the best by the tutors in terms of both quality and involvement, has the highest scores assigned by the system.

By analyzing each factor's correlation, it becomes quite clear that the tutors emphasized on the quality of the interventions, not on the mere number of utterances. As the Intraclass Correlation Coefficient (ICC) was .61 on single measures and .86 on average, results in terms of intervention importance scores and social knowledge-building correlate extremely well with the average expert grades. Additionally, the personal knowledge-building dimension is better correlated with the expert's grades; this sustains the empiric observation that the personal effect reflected in each participant's involvement has a higher impact in the expert's evaluation than the social dimension, through which collaboration is attained. In other words, it is easier to grasp one's active participation than his social interaction from the perspective of collaboratively building knowledge.

Table 1: Correlation between manual and automatic participants' evaluations.

Participant Name	Number of Utterances	Overall Interventions Score	Overall Personal KB	Overall Social KB	Expert Grade				
					1	2	3	4	**Avg.**
Team 4									
Participant 1	90	215.34	262.19	27.63	9.0	10.0	8.0	9.5	**9.13**
Participant 2	61	99.78	114.91	15.39	7.0	7.0	8.0	6.5	**7.13**
Participant 3	120	220.92	248.99	12.76	8.0	9.0	8.5	10.0	**8.88**
Participant 4	118	205.79	239.96	29.19	8.0	6.5	9.0	8.0	**7.88**
Team 34									
Participant 1	23	60.13	77.91	11.42	6.0	6.0	7.0	6.0	**6.25**
Participant 2	34	64.17	72.58	6.66	5.0	7.0	7.0	6.0	**6.25**
Participant 3	73	120.03	139.51	13.22	8.0	7.0	8.0	7.0	**7.50**
Participant 4	60	124.89	141.12	4.55	7.0	4.0	7.0	5.5	**5.88**
Team 36									
Participant 1	54	160.89	200.18	37.95	9.0		9.5	8.0	**8.83**
Participant 2	67	201.57	251.98	49.46	10.0		8.0	10.0	**9.33**
Participant 3	119	327.45	412.78	58.83	9.0		8.0	8.5	**8.50**
Participant 4	57	161.39	200.60	40.33	9.0		9.5	8.0	**8.83**
Average correlation	**.64**	**.83**	**.84**	**.77**					

In terms of intense collaboration zones, manual annotations and automatically identified zones are presented in Table 2.

Table 2: Overlap between manual and automatic identification of intense collaboration zones.

Conversation	Number of utterances	Manually annotated collaboration zones	Automatically identified collaboration zones
Team 4	389	[90; 160], [320; 360]	[19; 43], **[72; 146]**, [157; 166], [240; 280], [290; 317], **[329; 379]**
Team 34	190	[90; 120], [170; 178]	[51; 67], [74; 78], **[94; 121]**, [127; 131], **[158; 166]**, **[181; 195]**
Team 36	297	Relatively uniform distribution	[38; 58], [69; 122], [136; 157], [202; 221], [286; 290]

By analyzing all previous results, besides the global evaluation factors that correlate well with the human evaluations, the following *indicators of bad collaboration* (mostly in Team 34 conversation) were observed: the high number of automatically identified zones containing 1 to 3 utterances which were not considered intense collaboration zones in the end, the low average value of social knowledge-building effects and no automatically identified collaboration zone with a wide spread (over 50 utterances, although it was the shortest conversation of the three). In contrast, *conversations with good collaboration* (namely Team 36 which had the best overall collaboration) have more balanced collaboration zones, higher average values of social knowledge-building and a better distribution and coverage over the entire conversation.

Further investigation and analysis of results is required on an extended sample of discussions, but by taking into consideration the subjectivity of the tasks at hand, preliminary results are quite promising and prove that our implemented dialogic model provides measures relevant in terms of the evaluation.

Conclusions

ReaderBench has been designed to support tutors and learners in various educational scenarios that involve the use of chats or forum discussions for solving different tasks, with emphasis on collaborative problem solving or debating. Based on the results of our first validation study, we strongly believe that *ReaderBench* can be used to support tutors in the time consuming process of manually assessing chat conversations.

Overall, starting from a dialogic model of discourse, our integrated analysis system can be used to identify cohesion gaps between utterances, to analyze participants' involvement and to evaluate collaboration individually and holistically, through the process of social knowledge-building. The collaboration assessment model can be considered a cornerstone as it induces dialogistic approach and emphasizes the social knowledge-building perspective of collaboration in CSCL environments. In extent, by combining two different perceptions of cohesion, CSCL participants can use this system to assess to what extent utterance cohesion reflects group cohesion, as an outcome of collaboration depicted from the interaction graph.

References

Bakhtin, M. (1984). *Problems of Dostoevsky's poetics*. Minneapolis, MN: University of Minnesota Press.

Bakhtin, M. (1986). *Speech genres and other late essays*. Austin: University of Texas.

Bastian M., Heymann S., & Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. *Proc. Int. AAAI Conference on Weblogs and Social Media*.

Bereiter, C. (2002). *Education and Mind in the Knowledge Age*. Mahwah: Lawrence Erlbaum Associates.

Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(4-5), 993–1022.

Dascalu, M., Trausan-Matu, S., & Dessus, P. (2010a). *Utterances assessment in chat conversations*. Research in Computing Science, 46, 323-334.

Dascalu, M., Rebedea, T., & Trausan-Matu, S. (2010b). *A Deep Insight in Chat Analysis: Collaboration, Evolution and Evaluation, Summarization and Search*. In D. Dicheva and D. Dochev (Eds.), Proceedings of AIMSA 2010, LNCS 6304, Springer, 2010, 191-200.

Dong, A. (2005). The latent semantic approach to studying design team communication, *Design Studies, 26*(5), 445-461.

Fujita, N., & Teplovs, C. (2009). Automating the analysis of collaborative discourse: Identifying idea clusters. In C. O'Malley, D. Suthers, P. Reimann & A. Dimitracopoulou (Eds.), *Proceedings of the Int. Conf. CSCL 2009* (pp. 162-164). Rhodes: ISLS.

Galley, M., & McKeown, K. (2003). Improving Word Sense Disambiguation in Lexical Chaining. *Proc. 18th International Joint Conference on Artificial Intelligence (IJCAI'03)*, Acapulco, Mexico.

Graesser, A. C., McNamara, D. S., Louwerse, M., & Cai, Z. (2004*)*. Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36*, 193–202.

Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychol. Rev.*, *104*(2), 211–240.

McNamara, D., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1–43.

Miller, G. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, *38*(11), 39–41.

Manning, C., Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Mass: MIT Press.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith & C. Bereiter (Eds.), *Liberal Education in a Knowledge Society* (pp. 67-98). Chicago: Open Court.

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

Tapiero, I. (2007). *Situation models and levels of coherence*. Mahwah: Erlbaum.

Trausan-Matu, S., Rebedea, T., Dragan, A., & Alexandru, C. (2007). *Visualisation of learners' contributions in chat conversations*. In J.Fong & F.L. Wang(Eds.), Blended learning (pp. 217-226). Singapour: Pearson.

Trausan-Matu, S., & Rebedea, T.(2009) *Polyphonic Inter-Animation of Voices in VMT*. In Stahl, G. (Ed.), Studying Virtual Math Teams (pp. 451-473), Boston, MA: Springer US.

Trausan-Matu, S. (2010). Computer Support for Creativity in Small Groups using chats. *Annals of the Academy of Romanian Scientists, Series on Science and Technology of Information*, 3(2), 81-90.

Trausan-Matu, S., Dascalu, M., & Rebedea, T. (2012). *A System for the Automatic Analysis of Computer-Supported Collaborative Learning Chats*. In proceedings of ICALT 2012, IEEE 2012, 95-99.

Acknowledgments