ReaderBench – Automated Evaluation of Collaboration based on Cohesion and Dialogism Learning

Mihai Dascalu, Stefan Trausan-Matu, Danielle S. Mcnamara, Philippe Dessus

ReaderBench – Automated Evaluation of Collaboration based on Cohesion and Dialogism

Mihai Dascalu * Stefan Trausan-Matu * Danielle S. McNamara * Philippe Dessus

Abstract   As Computer Supported Collaborative Learning (CSCL) gains a broader usage, the need for automated tools capable of supporting tutors in the time-consuming process of analyzing conversations becomes more pressing. Moreover, collaboration, which presumes the intertwining of ideas or points of view among participants, is a central element of dialogue performed in CSCL environments. Therefore, starting from dialogism and a cohesion-based model of discourse, we propose and validate two computational models for assessing collaboration. The first model is based on a cohesion graph and can be perceived as a longitudinal analysis of the ongoing conversation, thus accounting for collaboration from a social knowledge-building perspective. In the second approach, collaboration is regarded from a dialogical perspective as the intertwining or synergy of voices pertaining to different speakers, therefore enabling a transversal analysis of subsequent discussion slices.

ReaderBench – Automated Evaluation of Collaboration based on Cohesion and Dialogism

Abstract   As Computer Supported Collaborative Learning (CSCL) gains a broader usage, the need for automated tools capable of supporting tutors in the time-consuming process of analyzing conversations becomes more pressing. Moreover, collaboration, which presumes the intertwining of ideas or points of view among participants, is a central element of dialogue performed in CSCL environments. Therefore, starting from dialogism and a cohesion-based model of discourse, we propose and validate two computational models for assessing collaboration. The first model is based on a cohesion graph and can be perceived as a longitudinal analysis of the ongoing conversation, thus accounting for collaboration from a social knowledge-building perspective. In the second approach, collaboration is regarded from a dialogical perspective as the intertwining or synergy of voices pertaining to different speakers, therefore enabling a transversal analysis of subsequent discussion slices.

Introduction

Computer Supported Collaborative Learning (CSCL) has gained a broader usage in multiple educational scenarios (Stahl, Koschmann, & Suthers, 2006). CSCL technologies facilitate the development of learning environments in which knowledge is collaboratively built and shared (Stahl, 2006), based on the inter-twining of collective and individual learning processes (Cress, 2013). Moreover, CSCL has become a viable alternative to classic learning environments as it can be employed in various settings, such as Massive Open Online Courses (MOOCs) or collaborative serious games (Hummel et al., 2011). At the same time, the need for automated tools capable of supporting and evaluating the involved actors has become more evident given the time-consuming processes involved in the analysis of multi-participant conversations (Holmer, Kienle, & Wessner, 2006). For example, Trausan-Matu (2010a) reported that the time required for a thorough

analysis greatly exceeds the actual duration of the chat session, rendering the manual evaluation process impossible for large corpora.

In terms of defining the variables of our analysis, **collaboration** can be perceived as a measure of interaction among participants centered on sharing ideas, fostering creativity for working in groups (Trausan-Matu, 2010b), and influencing others' points of view during the discussion. Complementary, participation represents the active involvement of members in ongoing CSCL conversations, which can be seen as independent processes that do not solely consist of collaboratively exchanging ideas with other participants. The number of uttered contributions can be considered the baseline for *participation*, whereas *collaboration* is reflected in the utterances addressed to other participants that bring a contribution to the knowledge building process. As such, our principal interest lies in automatically assessing collaboration within CSCL text-based, multi-participant interactions, and in particular, those performed within educational contexts. In order to achieve this goal, we propose two computational models based on *dialogism* and *cohesion,* two core concepts that can be used to highlight collaboration zones and become *signatures of collaboration* between different participants.

Dialogism

The concept of **dialogism** was introduced by Bakhtin (1981) and covers a broader, more abstract and comprehensive perspective of continual dialogue that exists in any type of text. Dialogism is centered on the dialogue reflected in "any kind of human sense-making, semiotic practice, action, interaction, thinking or communication" (Linell, 2009, pp. 5-6). This definition of dialogism, besides the intrinsic dialogue between individuals, may well be present in any text as "life by its very nature is dialogic … when dialogue ends, everything ends" (Bakhtin, 1984, p. 294). In addition, dialogue can be also perceived as 'internal dialogue within the self' or 'internal dialogue' (Linell, 2009, ch. 6), 'dialogical exploration of the environment' (Linell, 2009, ch. 7), 'dialogue with artifacts' (Linell, 2009, ch. 16) or 'dialogue between ideas' (Marková, Linell, Grossen, & Salazar Orvig, 2007, ch. 6). Regardless of context, discourse is modeled from a dialogical perspective as interaction with others, essentially towards building meaning and understanding.

Dialogism offers a well-grounded theoretical framing to automated discourse analysis and in particular, CSCL. Its key features are *multivocality* and *polyphony* (Koschmann, 1999), both tightly connected to the core concept of *voice*. In a nutshell, a **voice** expresses a distinct point of view, a position within the dialogue, and is reflected in concepts, utterances or events that will further influence the conversation (Trausan-Matu, 2010a). Therefore, a voice can be perceived as individual or collective perspectives on topics (Linell, 2009) that are socially generated and sustained in the "circulation of ideas" (François, 1993; Hudelot, 1994; Salazar Orvig, 1999). Individuals internalize and assimilate these ideas, and re-emit them as personal points of view or voices centered on the topics of the conversation. The overall conversation becomes analogous to a "voting" of uttered ideas, followed by an alignment to other individuals who share similar perspective (Linell, 2009).

Starting from the definition of voices, *multivocality* is centered on the multitude of meanings and the dialogue between multiple voices. Even further, *polyphony*, a central concept within our analysis, encapsulates multiple points of view or voices while focusing on their inter-animation, as well as the inter-relationships captured by their co-occurrence and overlap. Moreover, in addition to multivocality, polyphony is characterized also by a coherent achievement of the participating voices.

Following the perspective of Bakhtin (1981), the ***inter-animation of voices*** is generated by the influences between utterances, their interaction one with another, as well as one's reflection onto another (Trausan-Matu, Stahl, & Sarmiento, 2007). This process of voice inter-animation occurs progressively from simple repetitions to complex referential relationships between utterances. Moreover, aside from providing a theoretical starting point for developing tools to instruct thinking skills (Wegerif, 2006), dialogism and the underlying inter-animation of voices become key components for ensuring the success of a collaborative learning activity. To further illucidate the concepts of *polyphony* and *voice inter-animation*, Tables 1 and 2 present chat excerpts corresponding to different scenarios.

Table 1. Conversation sample highlighting a dense inter-animation of voices (e.g., "blog", "forum", "chat" technologies used to define the best "solution" in the context of the semantically related concepts of "customer – buy – sell – product – employee" that can be perceived as a background voice), as well as a high collaboration between participants.

Participant ID	Utterance ID	Text
1	176	a blog would also be useful for describing our latest products in order to convince the potential customers to buy
3	177	and posting some images for example
2	178	the forum is used to find more about something…for promotion blog is the best solution in my opinion.
4	179	and videos
1	180	a forum would be useful for offering solutions to some problems that our customers have
2	181	I agree…also other people can offer solutions not only us
4	182	agree but I meant that maybe we can discover that someone has a problem and post it on a forum and we can post back to offer our product as solution
1	183	and chat would also be great… as customers could have real-time feedback from the developers of the products they are using
2	184	or we can make a poll to find out what our customers will like to buy
4	185	what about new customers? What should we use for attracting them?

4

Participant ID	Utterance ID	Text
3	186	for selling the best solution is something like ebay but this is not our topic now
2	187	true ..but for using a chat the developers must be online.. so if they are not online….
1	188	from time to time we could also organize chats with customers and employees and ask for new ideas….
2	189	and in this case the best solution is a blog

Table 2. Conversation sample denoting a low inter-animation of voices as the dialogue is centered on only the "forum" voice that is presented in terms of "information – question – answer – communication" and "tools" voices, as well as low collaboration due to the monologue of one participant

Participant ID	Utterance ID	Text
3	219	so you tell us when you're done about the advantages
1	220	good communication tool
1	221	you can ask questions
1	222	you can give answers
1	223	you can find answers
1	224	unlike chats the information is well structured if the admin is smart and you can store it very well
1	225	good documentation tool
1	226	everything is stored and if a company wants to organize an information it can
1	227	available anytime and easy access with a good search engine
1	228	ease of use
1	229	anyone can add a post on a forum
1	230	anyone can access a forum

Text Cohesion

Besides dialogism, a key element of analysis in terms of discourse structure is ***cohesion***. Halliday and Hasan (1976) introduced the notion of cohesion as "relations of meaning that exist within the text, and that define it as a text." (Halliday & Hasan, 1976, p. 4). Cohesion provides overall unity and is used to establish the underlying structure of meaning. In other words, cohesion addresses the connections in a text based on features that highlight relations between constituent elements (words, sentences, or utterances). Overall, textual cohesion can be perceived as the sum of lexical, grammatical, and semantic relations that link together textual units. High cohesion usually models a consistent information flow, whereas cohesion gaps indicate in most cases topic changes corresponding to different discussion threads or off-topic contributions (see Tables 3 and 4).

Table 3. Conversation sample denoting a lower cohesion between adjacent contributions specific to brainstorming sessions – multiple topics and inter-twinned discussion threads can be observed (e.g., all discussion topics are clearly highlighted as voices that pertain to multiple users: wiki, forum, wave, chat)

Participant ID	Utterance ID	Text
1	223	wiki for documentation and faqs
3	224	and a forum for technical support
5	225	forum for technical support and maybe chat for live support
5	226	wave for collaboration brainstorming document sharing
3	227	chat for live support inside the company
4	228	yes live support is a good idea
2	229	we could also use chat for meetings ...

Table 4. Conversation sample denoting a high cohesion between more elaborated contributions centered on the benefits of "blog – blogger - blogging" and their corresponding "posts".

Participant ID	Utterance ID	Text
2	290	1. blogs can be well organized and therefore provide valuable information to the reader who wants to understand a certain topic
2	291	2. the blogger or the many bloggers blogging on one blog can add a personal touch to the information provided
2	292	3. there can be blogs for everything: for kids people about their life or business blogs - sharing information about a company technology etc.
2	293	4. blogs are a good support for providing conclusions and the steps taken in developing a technology
2	294	5. one can add tags to posts so that reader can select only the post him she likes or wants to read
2	295	6. users can comment at the end of every post - it's like having a forum at the end of every post
2	296	7. security is good also: only allowed users can post on a blog
2	297	and the blogger can also decide who to let to comment on his/her post

Transition toward Automated Computational Models

To date, only a few CSCL models based on dialogism have been proposed, and even fewer approaches provide automated analytic tools – for example, Dong's use of Latent Semantic Analysis (LSA) of design-team communication (Dong, 2005), *Polyphony* (Trausan-Matu, Rebedea, Dragan, & Alexandru, 2007), the *Knowledge*

Space Visualizer (Teplovs, 2008), and *PolyCAFe* (Trausan-Matu & Rebedea, 2010; Dascalu, Rebedea, Trausan-Matu, & Armitt, 2011; Trausan-Matu, Dascalu, & Rebedea, 2014). As a detailed comparison to other computational models is more suitable after providing an in-depth view of our models, the Discussion section highlights similarities and differences to three major approaches: the contingency graph (Medina & Suthers, 2009; Suthers & Desiato, 2012), transacts (Joshi & Rosé, 2007; Rosé et al., 2008), and Teplovs (2008) knowledge space visualizer.

In this paper, we propose two computational models integrated within our framework – *ReaderBench* (Dascalu, Trausan-Matu, & Dessus, 2013a). The first one, the dialogical voice inter-animation model described in the following section evaluates collaboration as an intertwining or overlap of voices pertaining to different speakers (Dascalu, Trausan-Matu, Dessus, & McNamara, 2015a). The second approach, the social knowledge-building model (Dascalu, Trausan-Matu, et al., 2013a; Dascalu, Trausan-Matu, Dessus, & McNamara, 2015b), represents a refinement of gain-based collaboration assessment (Trausan-Matu, Dascalu, & Rebedea, 2012) and takes full advantage of the cohesion graph (Trausan-Matu, Dascalu, & Dessus, 2012). In order to implement this model, we introduce the Cohesion Network Analysis (CNA) in section 3 as a means to score utterances and to analyze discourse structure within collaborative conversations. Both models are then used to assess the degree of collaboration between participants and to identify intense collaboration zones. Table 1 is a representative example of such an intense collaboration zone, which has, in that particular case, both a dense inter-animation of voices, as well as a high cohesion between contributions.

As an initial comparison between our two models, collaboration is regarded within the dialogical voice inter-animation model as the intertwining or overlap of voices pertaining to different speakers, therefore enabling a transversal analysis of subsequent discussion slices. On the other hand, the social knowledge-building model based on CNA (Dascalu, Dessus, et al., 2013) can be perceived as a longitudinal analysis accounting for collaboration from a social knowledge-building perspective. Afterwards, in Section 4, we validate the two computational models by comparing the predictions generated by *ReaderBench* with human annotations of collaborative conversations. In the end, we compare our models to other computational approaches, discuss their benefits and limitations, and conclude with future research paths. As an overview of the performed analyses, **Figure 1** presents the key concepts and methods of both computational models, as well as all of the automated indices used to predict collaboration, described in detail in results section.
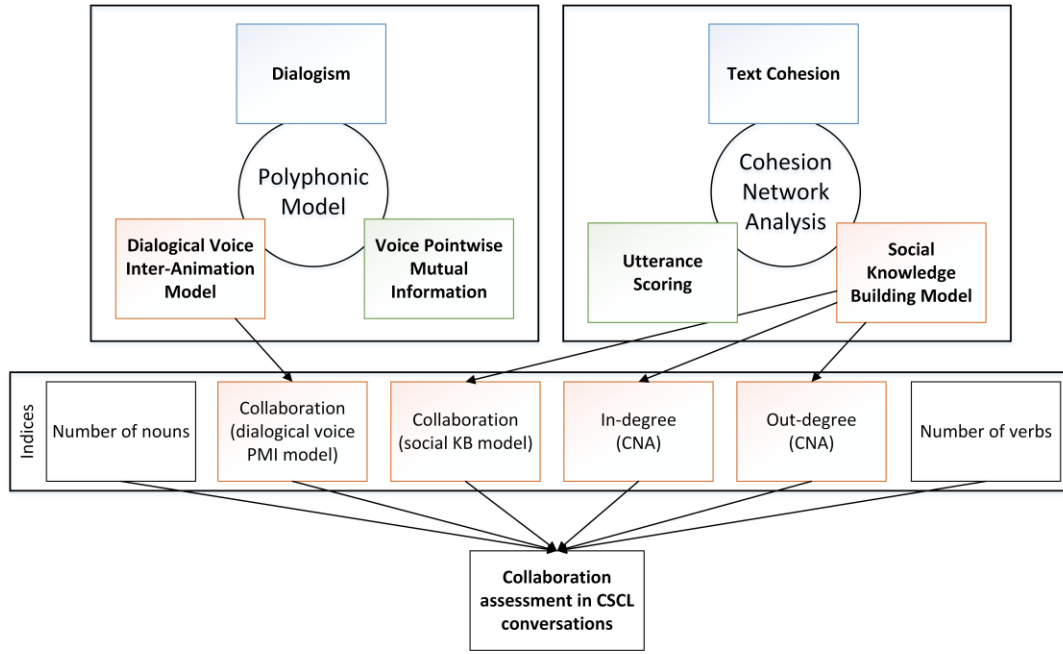
Figure 1. Visual representation of collaboration assessment based on both dialogical and social knowledge-building models.

From a more pragmatic perspective, this study represents an extension of the initial model (Dascalu, Trausan-Matu, et al., 2013a), which has now been further validated within an educational setting. Moreover, this paper represents an integrated view of dialogism (Dascalu, Trausan-Matu, et al., 2015a) and cohesion-based (Dascalu, Trausan-Matu, et al., 2015b) models which were previously presented separately. In contrast to simpler models which rely on counting the number of utterances exchanged between different speakers or the underlying links (Mislove et al., 2007), our models support the idea that dialogism and cohesion are salient predictors of collaboration. Therefore, signatures of collaboration emerge by modeling the interactions between participants through textual cohesion and voices' inter-animation. In addition, it is common for tutors to attempt to detect breaks in conversations that have limited or no collaboration or intense collaboration zones in learners' productions. Automated methods, such as those implemented in *ReaderBench* (Dascalu, Dessus, et al., 2013; Dascalu, 2014), provide crucial support to tutors in extracting such zones.

The Polyphonic Model and Collaboration Derived from Voice Inter-Animation

Philosophical Implications of Dialogism and the Polyphonic Model

One of the most important ideas of CSCL is that learning can be seen as a collaborative knowledge-building process (Bereiter, 2002; Scardamalia & Bereiter, 2006). Small groups of students interact (Stahl, 2006) and inter-animate (Trausan-Matu, Stahl, et al., 2007), rather than participate within knowledge transfer from the teacher to the learner. Moreover, if students receive tasks in their Zone of

8

Proximal Development (ZPD) (Vygotsky, 1978), the learning process may be seen as having two intertwining cycles: a personal one and a social knowledge building one (Stahl, 2006).

In order to properly introduce the polyphonic model presented in detail later on within this section, we must first present the three core and inter-dependent concepts of discourse analysis: utterances, voices and echoes. While *utterances* are defined as the main units of the analysis, *voices* may be considered to represent distinctive points of view emerging from the ongoing discussion. On the other hand, *echoes* represent the replication of a certain voice, the overtones and repetitions of the specific point of view that occur later on, with further implications in the discourse. Although the complexity of an utterance may vary greatly from a simple word to an entire novel (Bakhtin, 1986), our analysis adheres to Dong's perspective of separating utterances based on turn-taking events between speakers (Dong, 2009). Therefore, a new point of view or contribution from a different participant may divide the discourse by potentially modifying the inner, ongoing perspective of the current speaker. At a more fine-grained level, words, seen as the constituents of utterances, provide the liaisons between utterances and deepen the perspective of others' contributions into one's discourse. Obviously, utterances may contain more than a single voice, as well as alien voices to which the current voice refers (Trausan-Matu & Stahl, 2007). An *alien* voice is part of a turn uttered by a given participant that is later replicated in another one, marking therefore the transfer among different participants and their corresponding points of view with regard to the voice's central word.

In addition, if we consider the case of CSCL using instant messenger (chat), the collaborative knowledge construction in small groups necessitates the negotiation of participants' perspectives (Stahl, 2006). Any negotiation comprises both divergences and agreements among participants' opinions. In CSCL chats, students articulate personal beliefs (Stahl, 2006), they write utterances that contain ideas mediated by words. These utterances contain each student's personal ideas but they also contain others' ideas. We may say that they *revoice* others' utterances (Trausan-Matu et al., 2014). Following the musical metaphor introduced by Bakhtin (1981), during the chat conversations, the divergences and agreements among participants' opinions may be seen as dissonances and consonances among voices (Trausan-Matu, Stahl, et al., 2007).

The utility of the musical metaphor for CSCL may be more evident if we refer to the polyphony phenomenon, which was considered as an ideal model for collaborative sessions (Trausan-Matu, 2010a). Polyphony can be described as a group of voices jointly constructing a harmonious musical piece while each voice keeps its individuality. An important aspect of polyphony is that dissonances appear and are needed for assuring novelty, but these are eventually resolved. Therefore, conflicting views, various angles, and multiple perspectives can emerge, generating a truly collaborative conversation. However, as voices express ideas and opinions, the polyphony perspective can be used to perform a deep dialogical discourse analysis by summing up multiple voices co-occurring within the same discussion thread.

9

Starting from the polyphony phenomenon, Trausan-Matu and colleagues introduced a polyphonic model of CSCL (Trausan-Matu, Stahl, & Sarmiento, 2006; Trausan-Matu, Stahl, et al., 2007; Trausan-Matu, 2010a). The topics of discussion in students' CSCL chats can be seen as voices that inter-animate. Due to the specific individual features of each voice, differences appear manifested in dissonances that, for the sake of a coherent discussion, need to be resolved towards consonances, as in a polyphonic music piece. Each utterance contains both individual (inner) and alien (echoed) voices. The analysis of knowledge construction in groups should consider both these contributions. Therefore, the polyphonic model focuses on the notion of identifying voices in the analysis of discourse and building an internal graph-based representation, whether relying on the utterance graph (Trausan-Matu, Rebedea, et al., 2007) or the previously defined cohesion graph (Dascalu, Trausan-Matu, et al., 2013a). To this end, links between utterances are analyzed using repetitions, lexical and semantic chains, as well as cohesive links, and a graph is built in order to highlight discussion threads. Lexical and semantic cohesion between any two utterances can be considered the central liaison between the analysis elements within the graph.

Moreover, of particular interest is the multi-dimensionality of the polyphonic model (Trausan-Matu, 2013). First, following the conversation timeline, the *longitudinal* dimension is reflected in the explicit or implicit references between utterances. This grants an overall image of the degree of inter-animation of voices spanning the discourse. This polyphony provides a signature for collaboration, as the quality of interactions between multiple participants in a conversation is reflected within their voices. Second, *threading* affords the highlighting of voices' evolution in terms of the interaction with other discussion threads. Third, the *transversal* dimension is useful for observing a differential positioning of participants, when a shift of their point of interest occurs towards discussing other topics.

Finally, we must also emphasize an intrinsic problem that "it is indeed impossible to be 'completely dialogical', if one wants to be systematic and contribute to a cumulative scientific endeavor" (Linell, 2009, p. 383). The latter point of view also augments the duality between individual involvement and actual collaboration throughout a given CSCL conversation, as it is impossible to focus on both the animation with other participants' utterances and sustainably provide meaningful contributions. In the end, a balance needs to be achieved between individuals, without facing discourse domination.

Polyphonic Model

Until recently, the goals of discourse analysis in existing approaches oriented towards conversations analysis were to detect topics and links (Adams & Martell, 2008), dialog acts (Kontostathis et al., 2009), lexical chains (Dong, 2006), or other complex relations (Rosé et al., 2008). The polyphonic model makes use of advanced NLP techniques by taking full advantage of cohesion, integrates multiple semantic models (i.e., Latent Semantic Analysis – LSA, Latent Dirichlet Allocation – LDA and semantic distances from *WordNet*), as well as Social Network Analysis (Wasserman & Faust, 1994). Several computer-based support systems were

developed for assisting the polyphonic analysis: *Polyphony* (Trausan-Matu, Rebedea, et al., 2007), *PolyCAFe* (Trausan-Matu et al., 2014), and *ReaderBench* (Dascalu, Trausan-Matu, et al., 2013a; Dascalu, 2014), the latter being used within the current experiments.

The automated voice identification process starts by building lexical chains spanning throughout the conversation, which are afterwards merged into semantic chains by using the previously defined cohesion function (Dascalu, Trausan-Matu, & Dessus, 2013b). Due to the limitation of discovering lexical chains (Galley & McKeown, 2003) that only consider words with the same part-of-speech, the merge step is beneficial as it unites groups of concepts based on the degree of cohesion. In this context, we have proposed an iterative algorithm similar to an agglomerative hierarchical clustering algorithm (Hastie, Tibshirani, & Friedman, 2009) for merging lexical chains (Dascalu, Trausan-Matu, et al., 2015a). Groups of already clustered words are merged if the cohesion among them exceeds an imposed threshold. The empirically selected values for our experiments were of .75 for LSA and .85 for LDA, which best associated concepts pertaining to different lexical chains.

As semantic chains span across the discourse, the context generated by the co-occurrence or repetitions of tightly cohesive concepts is similar to the longitudinal dimension of voices. Echoes can be highlighted through cohesion based on semantic relationships between voice occurrences in different contributions, whereas attenuation is reflected in the considered distance between analytic elements. Moreover, by intertwining different semantic chains within the same textual fragment (sentence or utterance), we are able to better grasp the transversal dimension of voice inter-animation. Therefore, after manually selecting the voices of interest, the user can visualize the conversation as an overlap of co-occurring semantic chains that induce polyphony (see Figure 2). A voice is displayed within the interface as the three most frequent semantically related word lemmas. Its occurrences throughout the conversation are marked accordingly within the overall timeframe. Different speakers who utter a particular voice are demarcated with randomly assigned colors, consistent throughout a conversation for each participant. Each utterance may incorporate more than a single voice, as it may include, in addition to the current participant's voice, at least one other, an alien or echoed voice, re-uttered later on in the discourse after its first occurrence (Bakhtin, 1981; Trausan-Matu & Stahl, 2007). Overall, voices are reflected in the individual occurrences of the concepts from each semantic chain and, in return, are used to highlight the cohesive links that span throughout the discourse (Dascalu, Trausan-Matu, et al., 2013b).

Based on the previous rules of representation, the chart from Figure 2 follows the conversation timeline expressed in utterance identifiers and depicts the occurrences of five dominant voices, manually selected by the user for visualization purposes: a) use, application, technology; b) need, thing, want; c) chat, talk, debate; d) information, database, password; and e) forum, meeting, conference. Each of the five chat participants has a corresponding color and each voice occurrence reflects a speaker's assigned color.
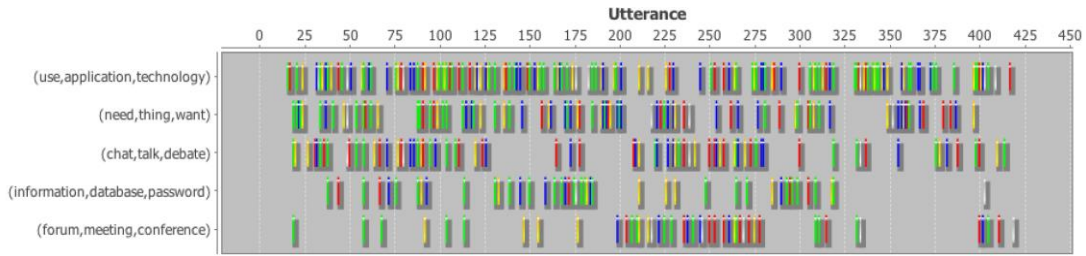
Figure 2. Chat voice inter-animation visualization covering participants' voices and implicit (alien) voices.

In order to better grasp the importance of each voice within the discourse, we have devised a series of indices, some inspired from 'rhythmanalysis' (Lefebvre, 2004) and 'polyrhythm' (The New Harvard Dictionary of Music, 1986). First, the number of contained words within each voice is used as a purely quantitative factor. Second, the cumulative scores of the analysis elements provide a broader qualitative perspective of the importance of the context of each voice's occurrences. Third, the recurrence of voices, inspired from rhythm analysis and seen as the distance between two analysis elements in which consecutive occurrences of the voice appear, is used to reflect the spread of each voice. Moreover, in accordance to Miller's law (Miller, 1956), we have applied a moving average (Upton & Cook, 2008) on the voice distribution for five datum points representing consecutive utterances. In other words, we have weighted the importance of each concept occurrence over 5 adjacent utterances, if no break in the discourse is larger than an imposed, experimentally determined threshold of one minute. Exceeding this value would clearly mark a stopping point in the overall chat conversation, making unnecessary the expansion of the singular occurrence of the voice over this break. The imposed values were experimentally determined, as there were extremely few explicit links manually added by the users that exceeded these thresholds. This step of smoothing the initial discrete voice distribution plays a central role in subsequent processing as the expanded context of a voice's occurrence is much more significant than the sole consideration of the concept uttered by a participant in a given contribution. In this particular case, entropy (Shannon, 1948) has been applied on the smoothed distribution in order to highlight irregularities of voice occurrences throughout the entire conversation.

By considering all of the previous indices used to estimate the importance of a voice, Table 5 presents an image of their correlations when considering a conversation of approximately 420 utterances. All 75 automatically identified voices, including the ones presented in Figure 2, are considered with the sole constraint that each voice include at least five word occurrences in order to have a quantifiable overall impact. Overall, all factors, besides recurrence, correlate positively and can be used to estimate the overall impact of a voice within the conversation. In contrast, recurrence is more specific and can be used to pinpoint whether the concepts pertaining to a voice are collocated or are more equally dispersed throughout the discourse. Nevertheless, small correlation values are acceptable as our aim was to identify meaningful factors that can be used to better characterize a voice's importance. Further evaluations need to be performed in order to determine the most representative factors, but our aim was to identify

12

specific measures that are generated as effects of different underlying assessment factors. For example, the use of the number of utterances in which the voices occurred or of statistics applied on the initial distribution would have been inappropriate as all of these indices would have been directly linked to the number of words within each semantic chain.

Table 5. Cross-correlation matrix between factors used to estimate the importance of voices (*$p < .05$; **$p < .01$).

Factor used to estimate the importance of voices	1	2	3	4	5
1. Number of words within the semantic chain	1				
2. Average utterance importance scores	.21	1			
3. Entropy applied on the utterance moving average	.80**	.23*	1		
4. Recurrence Average	-.59**	-.13	-.79**	1	
5. Recurrence standard deviation	-.55**	-.08	-.73**	.86**	1

As voice synergy emerges as a measure of co-occurrence of semantic chains, mutual information (Manning, Raghavan, & Schütze, 2008) can be used to quantify the global effect of voice overlapping between any pair of contiguous voices. Therefore, by computing the Pointwise Mutual Information (PMI) (Fano, 1961) between the moving averages of all pairs of voice distributions that appear in a given context, we obtain a local degree of voice inter-weaving or overlap. In order to better grasp the underlying reason of using PMI, we have presented in Figure 3 three progressive measures for synergy (Dascalu, Trausan-Matu, et al., 2013b).
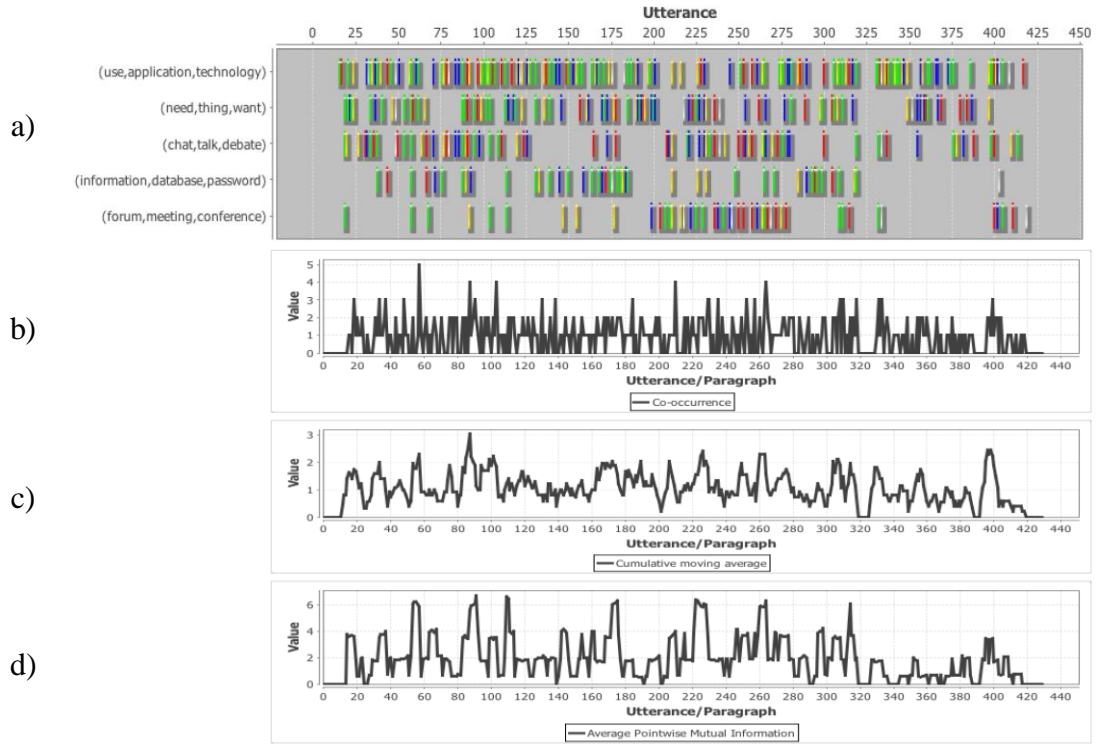
Figure 3. Evolution of voice synergy: a) Timeline evolution of voice occurrences (baseline for comparison); b) Number of co-occurrences; c) Evolution of cumulated moving average; d) Average Pointwise Mutual Information.

The first and the simplest estimator of overlap, the actual number of voices (co-) occurring, is misleading as we encounter a large number of singular values (meaningless, as only one voice is present) and double values, which are also not that interesting in observing the global trend. Also, the first spike with a value of 5 in Figure 3 is locally representative, but because it is isolated from the rest of the conversation, its importance should be mediated globally. The second estimation, the cumulated moving average, is better as the smoothing effect has a positive impact on the overall evolution. Nevertheless, it is misleading in some cases – for example, a spike is obtained around utterance 400 where the overall inter-animation of voices is quite low. The third estimator, the average PMI applied on the moving averages, best grasps the synergic zones (e.g., just before utterance 60 where we have four selected voices co-occurring, as well as around 90, 110, 220, and 260 due to the overlap of all five voices). Therefore, by observing the evolution of PMI using a sliding window that follows the conversation flow, we obtain a trend in terms of voice synergy that can be later on generalized to Bakhtin's polyphony (Bakhtin, 1984).

We opted to present the evolution of voice synergy as our computational model uses co-occurrence and overlap of voices within a given context. In order to emphasize further the effect of inter-animation that would induce true polyphony, we envisage the use of argumentation acts and discourse patterns (Stent & Allen, 2000). The latter approaches enable a deeper discourse analysis by highlighting the

14

interdependencies between voices and how a particular voice can shed light onto another.

Dialogical Voice Inter-Animation Model

In order to achieve genuine collaboration, the conversation must contain threads of utterances integrating *voices* that inter-animate in a similar way to counterpoint in polyphonic musical fugues (Trausan-Matu, Stahl, & Zemel, 2005; Trausan-Matu & Stahl, 2007). As collaboration is centered on multiple participants, a split of each voice into multiple viewpoints pertaining to different participants is required. A *viewpoint* consists of a link between the concepts pertaining to a voice and a participant through their explicit use within one's contributions in the ongoing conversation. We opted to represent this split in terms of implicit (alien) voices (Trausan-Matu & Stahl, 2007) (see Figure 4) because the accumulation of voices through transitivity in inter-linked cohesive utterances clearly highlights the presence of alien, echoed voices. In addition, this split presentation of semantic chains per participant is useful for observing each speaker's coverage and distribution of dominant concepts throughout the conversation.
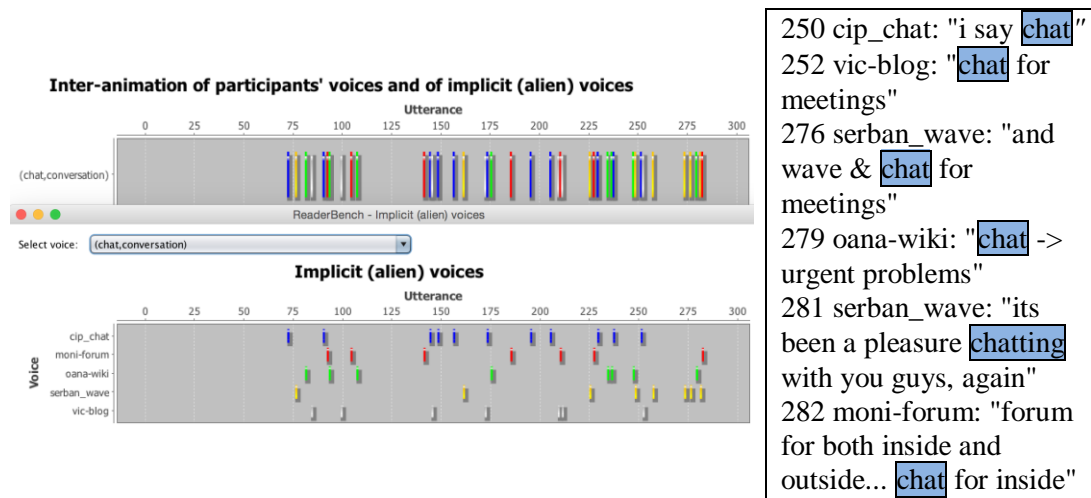

Figure 4. Chat-conversation voice split per participant, with examples from the last occurrences highlighting the voice's echo between different participants.

Afterwards, starting from the polyphonic model, collaboration is determined as the cumulated PMI value obtained from all possible pairs of contiguous voices pertaining to different participants (different viewpoints) within subsequent contexts of the analysis. From an individual point of view, each participant's overall collaboration is computed as the cumulated mutual information between an individual's personal viewpoint and all other participant viewpoints. In other words, by comparing individual voice distributions that span throughout the conversation, collaboration emerges from the overlap of voices pertaining to different participants.
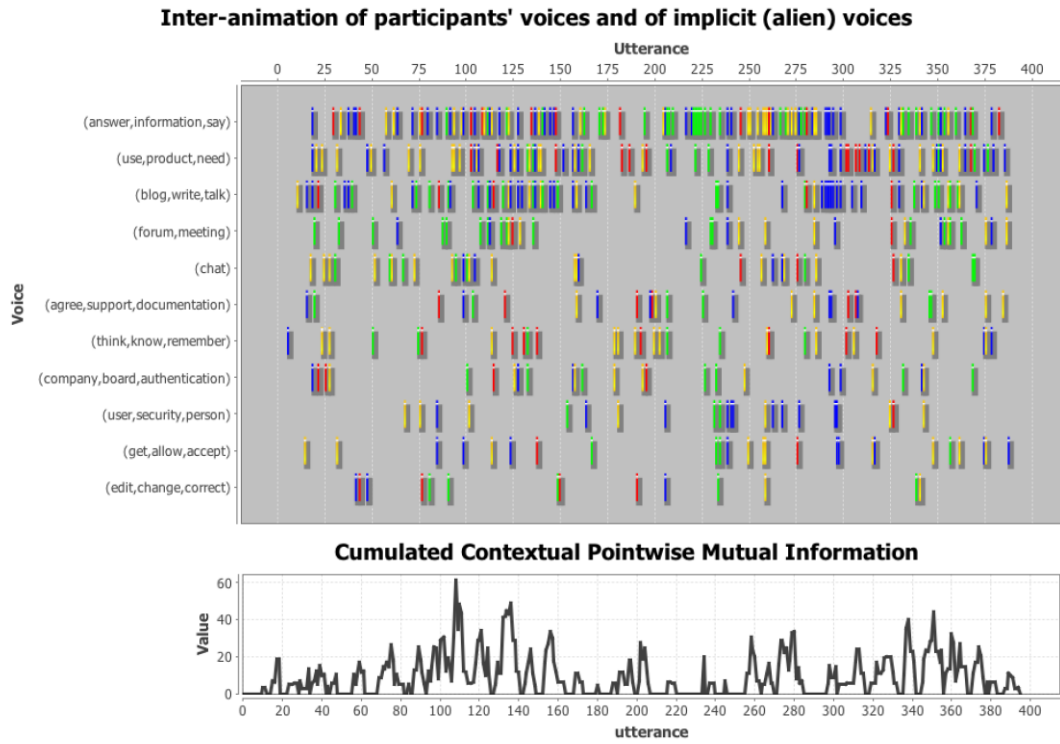
Figure 5. Collaboration evolution viewed as voice overlap between different participants (intertwining of different viewpoints).

The inter-animation frame from Figure 5 presents the voices with the longest semantic chain span throughout the conversation. Each peak of collaboration obtained through PMI corresponds to a zone with a high transversal density of voices emitted by different speakers (e.g., around utterances with the following identifiers: 110, 136, 225, 280, or 350). Two important aspects need to be mentioned. First, because the algorithm uses the moving averages and applies PMI on sliding windows, the user must also consider a five-utterance frame in which each individual occurrence is equally dispersed. Second, all of the voices from the conversation are considered (even those that have as low as three constituent words); this explains greater cumulative values encountered in the graph. As an example, Table 6 presents the chat sample centered on utterance 136 in which all conversation participants are engaged and multiple voices inter-animate.

16

Table 6. Conversation sample highlighting a dense inter-animation of voices pertaining to different participants (e.g., "wiki", "forum", "blog" and "knowledge").

Participant ID	Utterance ID	Text
2	134	"wiki wiki means rapidly in hawaiian language"
3	135	"the forum was the place where in roman times people used to come and talk business"
1	136	"and now the next best thing could be the blog - where someone shares it's knowledge"
2	137	"so it is a very quick way of letting others know what you have discovered"
4	138	"yes, but knowledge is stored in books"
4	139	"so a blog is not that needed"
3	140	"blogs are journals, good to say what you believe about one thing"

Cohesion Network Analysis and the Social Knowledge-Building Model

Discourse Structure and Cohesion Network Analysis

Cohesion is a central linguistic feature of discourse (McNamara, Louwerse, McCarthy, & Graesser, 2010) and is often regarded as an indicator of its structure. More specifically, cohesion can derive from various discourse connectors including cue words or phrases (e.g., 'but', 'because'), referencing expressions identified through co-reference resolution, as well as lexical and semantic similarity between concepts (Jurafsky & Martin, 2009; Raghunathan et al., 2010; McNamara, Graesser, McCarthy, & Cai, 2014). Semantic relatedness can be determined as semantic distances in lexicalized ontologies (Budanitsky & Hirst, 2006) or by using semantic models, such as LSA (Landauer & Dumais, 1997) or Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003).

Within our implemented model, cohesion is determined as an average semantic similarity measure of proximities between textual segments that can be words, phrases, contributions or the entire conversation. This semantic similarity considers, on the one hand, *lexical proximity*, identified as semantic distances (Budanitsky & Hirst, 2006) within *WordNet* (Miller, 1995). On the other hand, *semantic similarity* is measured through LSA and LDA semantic models trained on the Touchstone Applied Science Associates (TASA) corpus (http://lsa.colorado.edu/spaces.html, containing approximately 13M words) for the English version of our system using in the current experiments. Additionally, specific natural language processing (NLP) techniques (Manning & Schütze, 1999) are applied to reduce noise and to improve the system's accuracy: (a) the reduction of inflected forms to their lemmas, (b) the annotation of each word with its corresponding part of speech, and (c) stop word elimination. Additionally, individual word occurrences are adjusted for the term-document LSA matrix

through the use of term frequency-inverse document frequency (*Tf-Idf*) (Manning & Schütze, 1999).

Our previous studies (Dascalu, 2014) have shown that Wu-Palmer ontology-based semantic similarity (Wu & Palmer, 1994) combined with LSA and LDA models can be used to complement each other. Underlying semantic relationships are more likely to be identified if multiple complementary approaches are combined after normalization, reducing the errors that can be induced by using a single semantic model. To estimate cohesion using CNA, we combine information retrieval techniques (reflected by word repetition and term frequency) with semantic distance, estimated using ontologies (i.e., WordNet), LSA, and LDA. Cohesive links are defined as connections between textual elements that have high values for cohesion (i.e., a value that exceeds the mean value of all semantic similarities between constituent textual elements). In the end, a *cohesion graph* (Trausan-Matu, Dascalu, & Dessus, 2012; Dascalu, Trausan-Matu, et al., 2013a), which is a generalization of the *utterance graph* previously proposed by Trausan-Matu, Stahl, et al. (2007), is used to model all underlying cohesive links, providing a semantic, content-centered representation of discourse.

The cohesion graph is a multi-layered mixed graph consisting of three types of nodes (see Figure 6) (Dascalu, 2014). Starting from a *central* node, the entire conversation is split into *utterance nodes* (i.e., contributions per participant), which are divided into corresponding *sentence nodes*. *Hierarchical links* are enforced to reflect the inclusion of sentences into contributions, and of utterances within the entire conversation. *Mandatory links* are established between adjacent contributions and sentences, and are used to model information flow, rendering possible the identification of cohesion gaps within the discourse. In the particular case of chats, explicit links defined by users – such as those added by users in the *ConcertChat* (Holmer et al., 2006) graphical interface – are also included in the cohesion graph and are considered mandatory. Additional optional *relevant links* are added to the cohesion graph to highlight the semantic relatedness between distant elements. In our experiments, in order to reflect a high degree of similarity between the selected textual fragments, we opted to include only the cohesive links that have values exceeding the mean of all cohesion values by one standard deviation.
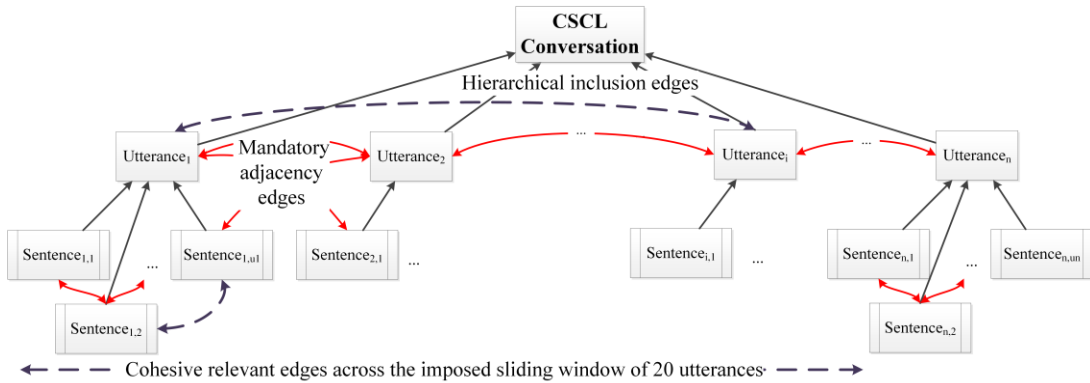

Figure 6. Cohesion graph generic representation.

In addition, due to the high number of contributions within a chat conversation, we opted to limit the search space for significant implicit cohesive links to 20

adjacent utterances. Rebedea (2012) has shown that links explicitly defined by users span a maximum of 20 utterances and are usually generated when a user feels that an implicit link is not obvious. Therefore, from a computational perspective in which the search space of similar utterances needs to be limited, we have adopted an equivalent window.

Cohesion-based Utterance Scoring

Within the CNA approach, we perform a content-centered analysis of utterances based on NLP and a cohesion-based discourse analysis. A central constituent for the evaluation process is the *utterance score* that reflects topics' coverage and the strength of the relatedness of each utterance to other contributions. Our approach can be compared to a purely quantitative approach that uses solely the number of contributions as a signal of collaboration. Here, we estimate an utterance's impact from the underlying concepts' relevance and cohesive links. Nevertheless, we cannot ignore the existing intrinsic link to the number of contributions, as more related words, even off-topic, determine the trend of the conversation.

In order to evaluate the importance of each utterance, we must first determine the value of its constituents or, more specifically, the relevance of each contained word. With regards to the process of evaluating each word's relevance in relation to its corresponding textual fragment (e.g., sentence, utterance, or entire conversation), there are several classes of factors that play important roles in the final analysis (Dascalu, Trausan-Matu, et al., 2015b) (see Table 7).

Table 7. Factors used to measure a word's relevance

Class	Descriptors
Statistical presence	Normalized term frequency used to reflect the specificity of each conversation
Semantic relatedness	Semantic similarity to the analysis element (sentence, utterance, entire conversation)
Semantic coverage	The importance of the semantic chain containing a particular word and its span throughout the entire conversation

The most straightforward factor consists of computing the *statistical presence* of each word. The next factor is focused on determining the *semantic relatedness* between a word and its corresponding textual fragment, whereas the last evaluates the *semantic coverage* of each concept. Semantic coverage is reflected by the length and the span of the semantic chains that contain semantically related concepts. This provides a reliable global estimate for the importance of each concept with regards to the entire conversation. Based on the previous classes of factors, the keywords of the conversation are determined as the words with the highest cumulative relevance based on their individual occurrences.

In terms of the scoring model, each utterance is initially assigned an individual score equal to the normalized term frequency of each word multiplied by its previously determined relevance (Dascalu, 2014). We measure to what extent each utterance conveys the main concepts of the overall conversation as an estimation of

on-topic relevance. Afterwards, these individual scores are augmented through cohesive links to other inter-linked textual elements by using the previously defined cohesion values as weights. Keywords reflect the local importance of each word, whereas cohesive links are used to transpose the local relevance upon other inter-linked elements.

Special attention is given in our approach towards utterances pertaining to the same speaker, considered as *inner* links, expressed as a continuation of the discourse that might potentially follow *alien* voices belonging to different participants. For some conversations, the importance of the links can be comparable in strength to the sum of all other out-going links, marking an individual behavior instead of collaboration, an aspect that we elaborate upon in the following section.

Social Knowledge-Building Model

The social knowledge-building model considers both personal and social knowledge-building (KB) processes (Bereiter, 2002; Scardamalia, 2002; Stahl, 2006). First, a personal dimension emerges by considering utterances by the same speaker, therefore modeling a kind of inner voice or continuation of the discourse. Second, inter-changed utterances with different speakers define a social perspective that models collaboration as a cumulative effect. This information exchange can also be perceived as "alien" voices that model the replication of the initial voice to different participants and their corresponding points of view with regards to the voice's central concept.

Our model is similar to some extent to the gain-based collaboration model (Trausan-Matu, Dascalu, & Rebedea, 2012) and marks a transition towards Stahl's model of collaborative knowledge-building (Stahl, 2006) by representing a conversation thread as our multi-layered cohesion graph. Whereas the previous section emphasized participatory analysis, our aim now shifts towards idea sharing, fostering creativity for working in groups (Trausan-Matu, 2010b) and influencing the other participants' points of view, thus enabling a truly collaborative discussion.
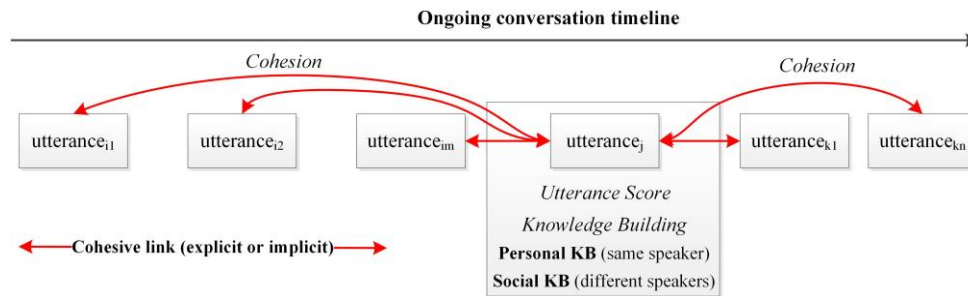

Figure 7. Slice of the cohesion graph depicting inter-utterance cohesive links used to measure personal and social knowledge-building effects (Dascalu, 2014).

As presented in Figure 7, the continuation of ideas or explicitly referencing utterances of the same speaker builds an inner dialogue or personal knowledge explicitly expressed in the discourse. In other words, personal knowledge building

addresses individual voices, more specifically participant voices and/or alien voices re-uttered by the speaker. In contrast, social knowledge building, derived from explicit dialog that by definition is between at least two different individuals, sustains collaboration and highlights external voices. Moreover, by referring to the dialogic model of discourse analysis, echoes are reflected by cohesion in terms of the information transferred between utterances. In addition, the echo attenuation effect considers the distance between the contributions and diminishes the strength of the cohesion link proportionally to the increase in distance.

Therefore, each contribution now has its previously defined importance score and a knowledge-building effect, both personal and social (see Figure 7). The personal effect is initialized as the utterance's score, whereas the social effect is zero. Later on, by considering all of the links from the cohesion graph, each dimension is correspondingly augmented. If the link is between utterances having 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 by the same amount of information (previous knowledge multiplied by the cohesion measure). As such, we measure collaboration as the sum of social knowledge-building effects, starting from each utterance score corroborated with the cohesion function.

We must also consider the limitations of our implemented model in terms of personal knowledge-building analysis. Through cohesion, collaboration emerges from social knowledge transfer and is perceived as the influence of one's contributions over other participants' discourse. In contrast, the approximation of personal knowledge-building represents an upper bound of the explicitly expressed information transfer between one's personal contributions. Similarly to the gain-based approach (Dascalu, Rebedea, & Trausan-Matu, 2010; Trausan-Matu, Dascalu, & Rebedea, 2012), we use a quantifiable approximation of inner dialogue, without being able to evaluate the overall cognitive and inference processes performed behind the scenes by the learner. Personal knowledge-building is seen as a reflection of one's thoughts expressed explicitly within the ongoing conversation as cohesive links between utterances 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.

Results

Validation Experiment

Our validation experiment is focused on the assessment of 10 chat conversations, selected from a corpus of more than 100 chats that took place in an academic environment. The 10 conversations were manually selected as being the most informative ones while covering most usage scenarios: combinations of highly

collaborative sections with monologues, on-topic discussions versus off-topic ones, equitable versus off-balanced involvement of participants, limited time-span versus extensive and long discussions. Within each chat, Computer Science undergraduate students from the fourth year undergoing the Human-Computer Interaction course at our university debated on the advantages and disadvantages of CSCL technologies (e.g., chat, blog, wiki, forum, or Google Wave). Each conversation involved four or five participants, with an equitable gender distribution, who previously knew each other by pertaining to the same class. Each participant first debated on the benefits and disadvantages of a given technology, and then proposed an integrated alternative that encompassed the previously presented advantages.

Afterwards, 110 fourth year undergraduate and master students were asked to manually annotate three chat conversations, grading the entire conversation and each participant individually on a 1-10 scale in terms of collaboration and, separately, participation. We opted to distribute the evaluation of each conversation due to the high amount of time required to manually assess a single discussion (on average, users reported 1.5 to 4 hours for a deep understanding) (Trausan-Matu, 2010a). Initially, for each conversation, we had on average 35 annotations, out of which raters with no variance and with a correlation lower than 0.3 in terms of intra-class correlations (ICC) with the other raters were disregarded. Most of the weak relationships to the other raters were, in most cases, due to erroneous or superficial evaluations. In the end, we had more than 20 ratings for each conversation. This resulted in an increased Cronbach's alpha from an average of 0.9 to a value of 0.96 (see Table 8). These high values demonstrate a very good agreement between rates and are justifiable by taking into consideration the high number of evaluations per conversation.

Raters were specifically instructed to evaluate collaboration as the exchange of ideas with *other* participants, not as the active involvement throughout the conversation. Raters had previous knowledge about each debated CSCL technology, but were unaware of the dialogical implications (e.g., polyphony) or of the automated models that would be later on enforced. In addition, raters were asked to identify intense collaboration zones as segments from the conversation with a high degree of collaboration among participants. These non-overlapping segments determined by each rater were defined as the start and end indexes of utterances among which participants actively collaborated. We opted not to request a rating per segment as from the overlap of more than 20 evaluations, collaboration peaks would emerge.

Table 8. Collaboration agreement among raters.

Conversation	Utterances	Participants	Duration (hours)	Initially assigned raters	Initial Cronbach's alpha	Final raters	Final Average ICC	Final Cronbach's alpha
Chat 1	339	5	2	37	.970	32	.954	.976
Chat 2	283	5	1.5	35	.821	23	.904	.945
Chat 3	405	5	2.5	40	.728	22	.953	.956
Chat 4	251	5	1.5	36	.907	24	.929	.956
Chat 5	416	5	1.5	34	.960	29	.951	.972
Chat 6	378	5	1.5	32	.957	26	.965	.975
Chat 7	270	5	1.5	35	.907	23	.920	.968
Chat 8	389	4	2	35	.923	26	.942	.967
Chat 9	190	4	1	36	.971	30	.897	.980
Chat 10	297	4	1.5	30	.864	20	.792	.936
Average	*321.8*	*4.7*	*1.65*	*35*	*.901*	*25.5*	*.921*	*.963*

With regards to the pre-processing phase of the chat conversation logs exported from *ConcertChat* (Holmer et al., 2006), all emoticons and non-dictionary words have been disregarded as typos were not represented in any semantic model space. In spite the fact that chats are considered in most cases a noisy text-based interaction medium, in our conducted experiments students retained an academic conduct as they were afterwards graded based on their involvement throughout the conversation. Moreover, although *ConcertChat* includes a second interaction space – a shared whiteboard –, no corresponding information was processed because learners were instructed to use the chat facility for brainstorming, without necessarily needing the whiteboard facilities. Therefore, we were faced with only a few typos, extremely limited slang and abbreviations, rendering adequate our approach of disregarding such words. Afterwards, natural language processing (NLP) techniques (Manning & Schütze, 1999) were applied to improve the system's accuracy: the reduction of inflected forms to their lemmas, part of speech tagging, and stop word elimination.

Validation of Collaboration Assessment

In order to have a broader analysis of collaboration, besides the two indices derived from the computational models presented in detail, we consider adequate to introduce additional indices of collaboration. First, we introduce *in-degree* and *out-degree* as Social Network Analysis (SNA) metrics applied on the interaction graph (Dascalu, Trausan-Matu, et al., 2013a; Dascalu, Trausan-Matu, & Dessus, 2014). This graph models the interaction between participants based on CNA, including the cohesion graph and on the utterance importance scores, as links reflect the cohesion similarity between the utterances of different participants. Second, the *number of nouns* is used as an estimator of the descriptive concepts expressed by each participant. Third, the *number of verbs* estimates each participant's commitment towards action and involvement with other participants. The simplest quantitative index mentioned in the Introduction section (number of exchanged utterances to other participants) is not feasible in this case because there are only a few explicit links added by users. All implicit links that are used to model the discourse are identified via CNA.

Pearson correlations (see Table 9) and non-parametric correlations (Spearman's Rho) (see Table 10) were determined between automated and human mean ratings for each conversation. As an interpretation of the results presented in Table 9 and Table 10, we can observe that predictions are accurate except for four conversations in which we could identify atypical behaviors highlighted in bold. In chats 2 and 10, similar rankings of collaboration for multiple participants highlight the difficulty in differentiating between participants due to similar involvement, therefore making the evaluation more prone to error. Chat 3 is overall off-balanced due to the focus on only one technology ("blog") which shifted the overall equilibrium with the other technologies that should have been debated. Chat 8 had specific zones in the conversation dominated by certain participants who misled the evaluation since monologue was not accordingly differentiated by raters in contrast to collaboration.

While there are reliable predictors of collaboration for each conversation, we must also consider that the overall evaluations are partially biased because some raters took into consideration quantitative factors to estimate collaboration (i.e., the number of utterances). Instead of focusing on the quality of the dialogue and on the way utterances pertaining to different participants inter-animate, quantity became the determinant factor for some raters.

Table 9 Pearson correlations between indices and mean rater collaboration (*p < .05; **p < .01).

Conversation	Social KB model	Dialogical voice PMI model	In-degree (CNA)	Out-degree (CNA)	Number of nouns	Number of verbs
Chat 1	.96**	.83	.98**	.92*	.89*	.96**
Chat 2	.68	**.41**	.71	.41	.74	.70
Chat 3	**.25**	.68	.54	.77	.79	.84
Chat 4	.66	.95*	.88*	.94*	.92*	.96*
Chat 5	.95*	.89*	.92*	.85	.83	.91*
Chat 6	.99**	.90*	.99**	.96**	.84	.95*
Chat 7	.96**	.47	.93*	.79	.75	.91*
Chat 8	.67	.53	.97*	.84	.73	.50
Chat 9	.82	.56	.84	.81	.80	.78
Chat 10	.84	**.28**	.85	.86	.69	.96*
Average	*.78*	*.65*	*.86*	*.81*	*.80*	*.85*

Table 10 Spearman correlations between indices and mean rater collaboration (*p < .05; **p < .01).

Conversation	Social KB model	Dialogical voice PMI model	In-degree (CNA)	Out-degree (CNA)	Number of nouns	Number of verbs
Chat 1	.90*	.90*	1.00**	.90*	.90*	.80
Chat 2	.60	**-.20**	.60	.20	.71	.40
Chat 3	**.30**	.50	.30	.80	.80	.80
Chat 4	.70	.90*	.90*	.90*	.90*	.98**
Chat 5	.90*	.70	.90*	1.00**	.82	.70
Chat 6	1.00**	.90*	1.00**	.90*	.60	.98**
Chat 7	.90*	.80	.90*	.80	.80	1.00**
Chat 8	**.40**	**.40**	1.00*	.40	.20	.20
Chat 9	.80	.40	.60	.60	.80	.80
Chat 10	.80	.60	.80	.80	.40	1.00**
Average	*.73*	*.59*	*.80*	*.73*	*.69*	*.77*

The indices were checked for multicollinearity (see Table 11) and all of the indices except the *Social KB model* were considered in further analyses, as this index was highly correlated with in-degree derived from CNA. We have opted to use in-degree because it has higher individual correlations per conversation and it better grasps collaboration in terms of social involvement.

Table 11 Correlation matrix among collaboration indices (*p < .05; **p < .01).

	1	2	3	4	5	6
1. Social KB model	1	.882**	.955**	.854**	.763**	.776**
2. Dialogical voice PMI model	.882**	1	.834**	.802**	.775**	.750**
3. In-degree (CNA)	.955**	.834**	1	.942**	.877**	.877**
4. Out-degree (CNA)	.854**	.802**	.942**	1	.943**	.906**
5. Number of nouns	.763**	.775**	.877**	.943**	1	.933**
6. Number of verbs	.776**	.750**	.877**	.906**	.933**	1

Overall, individual chat assessments support the reliability of the proposed qualitative indices in assessing collaboration, as well as the complementarity of the implemented indices—when one is skewed due to atypical behavior, the others compensate. Moreover, since our intent was to create a unitary predictive model for evaluating all conversations, we performed the same measurements after combining all individual ratings for all conversations (see Table 12). The later significant correlations support the adequacy of our proposed computational models. The lower values for the dialogical PMI model are justifiable, as the voice identification process requires further enhancements.

Table 12. Correlation between indices and mean rater collaboration for all conversations together (*p < .05; **p < .01).

Correlation measure	Dialogical voice PMI model	In-degree (CNA)	Out-degree (CNA)	Number of nouns	Number of verbs
R	.59**	.71**	.71**	.69**	.71**
Rho	.61**	.75**	.77**	.73**	.76**

A final stepwise regression analysis was calculated to determine the degree to which the automated indices predicted the human ratings of collaboration. This regression yielded a significant model, $F(1, 45) = 46.426$, $p < .001$, $r = .713$, $R^2 = .508$. One variable was a significant predictor in the regression analysis and accounted for 51% of the variance in the manual annotations of collaboration: number of verbs [$\beta = .713$, $t(1, 45) = 6.814$, $p < .001$]. This is understandable from the point of view of collaboration, as verbs induce action among participants. Moreover, regression analyses based on each collaboration model separately yielded significant models as well: $F_{CNA\ In\text{-}degree}(1, 45) = 45.960$, $p < .001$, $r = .711$, $R^2 = .51$ (extremely close to the step-wise model) and $F_{Dialogical\ voice\ PMI}(1, 45) = 24.533$, $p < .001$, $r = .594$, $R^2 = .35$.

Validation of the Identification of Intense Collaboration Zones

In addition to the estimation of collaboration based on both previous assessment models, *ReaderBench* automatically identifies *intense collaboration zones*. These zones are defined as utterance intervals in which participants are actively involved, collaborating and generating ideas related to the ongoing context of the discussion. With regards to the social knowledge-building model, these collaboration zones emerge as conversation segments with multiple cohesive links between different participants, therefore modeling the information transfer among them in a cohesive context. As a complementary view, the dense inter-animation of voices pertaining to different speakers also generates similar collaboration zones represented as voice overlap or co-occurrence.

From a computational perspective, the first step within our greedy algorithm (Dascalu, Trausan-Matu, et al., 2013a) that builds up intense collaboration zones consists of identifying social knowledge-building or voice PMI peaks as maximum local values. Afterwards, each peak is expanded sideways within a predefined slack

(experimentally set at 2.5% of the 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 five utterances are selected as intense collaboration zones.

In other words, after identifying the utterances with the greatest collaborative effect (highest social KB score or highest voice PMI pertaining to different speakers), the algorithm expands each zone to the left and to the right in a non-overlapping manner. 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 dialogism, cohesion and voice synergy bind utterances within an intense collaboration zone in terms of topic relatedness. For example, in Figure 5, we start with the maximum value of estimated collaboration around the utterance with ID 108 and we expand sideways, in the end obtaining the first intense collaboration zone - [87; 159]. All utterances within that interval have a high PMI score and denote voice overlap between different participants. Afterwards, the algorithm expands around utterances with IDs 375, resulting in the [311; 391] zone, as well as around 274, resulting in the third most important collaboration zone - [256; 282].

With regards to the validation experiment, all manual annotations were cumulated in a histogram that presented, for each utterance, the number of raters who considered it to be part of an intense collaboration zone. In the end, the same greedy algorithm was applied on this histogram in order to obtain an aggregated version. As presented in Table 13, there is good overlap in terms of accuracy measured as precision, recall, and *F1* score between the annotated collaboration zones and the two computational models. This indicates that the models are consistent with one another, but are also good estimators of the annotated zones, therefore demonstrating the feasibility of our two approaches. Moreover, the manual annotation process was a subjective and bias-prone task as there were no constraints imposed in terms of the overall coverage of these zones and the raters' perceptions of interaction among multiple participants.

Table 13. Evaluation of identification of intense collaboration zones.

Conversation	Overlap between annotated collaboration zones and Social KB model			Overlap between annotated collaboration zones and Voice PMI model			Overlap between Social KB model and Voice PMI model		
	P	R	*F1*	P	R	*F1*	P	R	*F1*
Chat 1	.72	.96	*.83*	.88	.80	*.83*	1.00	.68	*.81*
Chat 2	.64	.83	*.72*	.63	.61	*.62*	.92	.70	*.79*
Chat 3	.78	.96	*.86*	.84	.75	*.79*	1.00	.73	*.84*
Chat 4	.77	.81	*.79*	.78	.79	*.79*	.80	.77	*.79*
Chat 5	.64	.95	*.77*	.71	.83	*.77*	.91	.72	*.80*
Chat 6	.75	.88	*.81*	.75	.93	*.83*	.82	.86	*.84*
Chat 7	.64	.79	*.71*	.79	.62	*.69*	.94	.60	*.73*
Chat 8	.72	.80	*.76*	.75	.64	*.69*	.92	.71	*.80*
Chat 9	.89	.93	*.91*	.91	.64	*.75*	.86	.59	*.70*
Chat 10	.70	.85	*.77*	.73	.55	*.63*	.96	.59	*.73*
Average	*.73*	*.88*	*.79*	*.78*	*.72*	*.74*	*.91*	*.70*	*.78*

Discussion

Although constructed differently, both collaboration models are centered on dialogism and reflect cohesion. As voices are represented as points of view covering semantically related concepts, their recurrence reflects cohesive links within the discourse. Subsequently, the cohesive links from the cohesion graph represent the echoes of voices and model their span throughout the dialogue. Therefore, based on our results, we can consider cohesion as a binder between the utterances within an intense collaboration zone. Cohesion measures the topic relatedness between the utterances, whereas social interaction in a cohesive context determines collaboration. Moreover, the voice synergy effect between different participants captures a similar cohesive information flow in which alien voices shed light on each other. In other words, cohesion among the utterances of different speakers becomes a signature of collaboration within both models. In addition, the identified collaboration peaks and synergies build on text cohesion and voices' inter-animation become traces of dialogism and productive polyphony.

In order to better grasp the specificity of our analysis, we must also consider a comparison to other computational models of CSCL discourse, namely the contingency graph (Medina & Suthers, 2009; Suthers & Desiato, 2012) and transactivity (Joshi & Rosé, 2007; Rosé et al., 2008). First, the contingency graph is used as a representational foundation for abstract transcriptions and considers contingencies between events. As an analogy, our cohesion graph also considers *temporal proximity* while performing cohesion-centered and dialogical analyses in sliding conversation windows, as well as *semantic relatedness* that, in our case, is computed based on multiple semantic models.

Second, transactivity (Joshi & Rosé, 2007) can be perceived as a complementary approach to our information flow. In contrast to modeling information transfer

between participants through cohesion and voice inter-animation, transacts are used to represent the relationship between competing positions of different speakers similar to that of dialogue acts (Stolcke et al., 2000), but at a different semantic granularity. Therefore, we consider transacts as a potential extension of our two computational models that could be used to better reflect the synergy or juxtaposition of participants' points of view.

In terms of automated systems, the Knowledge Space Visualizer – KSV (Teplovs, 2008) might be considered to have many similarities to *ReaderBench*. However, while both systems envision the visualization of interactions between users through Social Network Analysis and semantic similarities, their respective approaches are fundamentally different. *ReaderBench* evaluates collaboration via a deep analysis of each conversation that employs multiple NLP techniques, including semantic distances, LSA and LDA. By contrast, KSV provides a more shallow perspective of individuals and links which can be structural (e.g., reply-to, build-on, reference, annotation, contains), authorial, or semantic (based only on LSA). In a nutshell, KSV was designed to provide an overview of interactions, with an emphasis on visualization, whereas *ReaderBench* makes use of in-depth discourse analysis.

There are also certain limitations of our models. Foremost, the models address only specific educational situations in which participants share, continue, debate, or argue certain topics or key concepts of the conversation. In other words, collaboration is particularly derived from idea sharing between participants who exchange cohesive utterances. It becomes evident that specific discourse markers or speech acts (e.g., confirmations or negations) (Austin, 1962; Searle, 1969) should also be considered for modeling collaboration. Moreover, as CNA and voice synergy capture cohesion through semantic similarity, additional discourse markers for identifying intertwined epistemic and argumentative moves, as well as social modes of interaction and consensus building (Weinberger & Fischer, 2006) need to be considered. But for our specific educational scenario presented in the validation experiments from Section 4, cohesion and voice synergy by themselves proved to be reliable predictors. As the students debated on specific topics, both textual cohesion and voice PMI highlighting the exchange or continuation of ideas represented a reliable estimator of the generated collaborative effect.

From a different perspective, the *ReaderBench* framework has also been used to assess the textual complexity of texts by providing a wide range of complexity indices covering surface, lexical, syntactic and semantic levels of discourse (Dascalu, Dessus, et al., 2014; Dascalu, Stavarache, et al., 2015). In future research, we will examine the assessment of learning and comprehension in the context of collaborative discourse using analogous indices adapted for chat conversation (characterized by short contributions). Moreover, key concepts from the *ConcertChat* shared whiteboard will be considered for as potential measures of relatedness to the extracted keywords from the conversation.

Overall, our models should not be perceived as rigid structures, but as adaptable ones that evolve based on the cohesion to other participants' utterances. Nevertheless, we must highlight additional limitations in terms of personal knowledge building, social knowledge transfer, noise within the experiment, and

underlying cognitive processes. As an initial assumption, we consider personal knowledge building as the reflection of one's thoughts continued into subsequent utterances through cohesive links. This is only partially valid because the written form expressed within the conversation can be substantially less representative than the processes and inferences performed in the learner's mind. Also, with regards to the dialogism model, further refinements of the automated identification of semantic chains need to be enforced in order to exclude less relevant voices identified at present.

From a higher level perspective built on top of cohesion, *coherence*—used to "jointly integrate forms, meanings, and actions to make overall sense of what is said" (Schiffrin, 1987, p. 39)—becomes a salient factor for collaboration. Furthermore, coherence can be considered a "semantic property of discourses, based on the interpretation of each individual sentence relative to the interpretation of other sentences" (van Dijk, 1977, p. 93). Moreover, coherence can be perceived as a generalization of cohesion due to its multiple additional perspectives (e.g., reader's skill level, background knowledge, and motivation, each helping to form the situation model) (Tapiero, 2007). Based on these definitions, collaboration that emerges from cohesion or voice inter-animation among the utterances of different speakers supports discourse coherence. Therefore, collaboration becomes an additional constituent specific to CSCL conversations that is required to achieve a coherent discourse.

This does not necessarily mean that collaboration determines coherence. However, the exchange of ideas and of points of view in a cohesive and dialogical manner greatly facilitates the processes of achieving a coherent mental representation, commonly called a situation model (van Dijk & Kintsch, 1983). To further argue this point, a monologue within a conversation is likely to be relatively coherent as it expresses only a participant's perspective, but it completely lacks collaboration. On the contrary, multiple participants could be actively involved in the conversation, collaborating one with another, but on different topics and generating nested sub-conversations. The overall effect would be of discourse segmentation due to multiple concurrent discussion threads, not to mention the frequent case of off-topic or irrelevant utterances, which further reduce discourse coherence. However, these contributions might nonetheless be considered stimulants for collaboration, and ultimately, coherence.

Starting from the definition provided by Graesser, McNamara, Louwerse, and Cai (2004, p. 193) that coherence is a "characteristic of the reader's mental representation of the text content", we further argue that, in the case of CSCL, we are dealing with a collective representation whose overall coherence is determined by the synergic effect of each individual's points of view or voices. Therefore, discourse coherence can be achieved collectively through collaboration and is built on cohesion that can become an indicator for collaboration if the exchange of information is performed between different participants.

Conclusions and Future Research Directions

Starting from a dialogic model of discourse centered on cohesion, we validated our system in terms of assessing collaboration by employing a longitudinal model

based on social knowledge-building and a different transversal model based on voice inter-animation. Within the social-knowledge building model, collaboration was evaluated using a bottom-up approach. Initially, the importance of an utterance was measured with regard to the overall discourse in terms of topics coverage wherein each contribution was assigned a corresponding score. Afterwards, collaboration was estimated as the impact on other speakers' utterances, therefore modeling information exchange between participants. In the second dialogical model, collaboration emerges from co-occurrences and the overlap of voices within a given context, emphasizing the tight inter-dependencies between collaboration and true polyphony.

Based on the performed analyses, we were able to extend the perspective of collaboration in terms of achieving a coherent representation of the discourse through the inter-animation of participants' points of view. Therefore, starting from dialogism as a framework of CSCL (Koschmann, 1999), we were able to model the exchange and sharing of ideas among participants in a conversation through specific computational linguistics. In conclusion, as the validations supported the accuracy of the models built on dialogism, we can state that dialogism derived from the overlapping of voices, as well as textual cohesion, can be perceived as a signature for collaboration.

In addition, our analyses have a broad spectrum of applications, extending from utterance cohesion towards group cohesion rooted in collaboration. For example, one line of our research will further examine the relations between student collaboration in forums and predicting their completion rate in MOOCs. We also envision the use of this dialogical perspective to assess narrative features of novels, highlighting different points of view pertaining to different characters. Still further, another set of experiments might focus on the assessment of students' self-explanations that can be perceived as a 'dialogue' between the author's text and students' thoughts viewed as echoes of the voices from the initial text. Overall, the range of potential applications for this approach is only limited by the presence of dialog in which collaboration emerges from the interactions between participants marked by textual cohesion and voices' inter-animation.

References

Adams, P.H., & Martell, C.H. (2008). Topic Detection and Extraction in Chat. In *IEEE Int. Conf. on Semantic Computing (ICSC 2008)* (pp. 581–588). Santa Clara, CA: IEEE.

Austin, J.L. (1962). *How to Do Things With Words*. Cambridge, MA: Harvard University Press.

Bakhtin, M.M. (1981). *The dialogic imagination: Four essays* (C. Emerson & M. Holquist, Trans.). Austin and London: The University of Texas Press.

Bakhtin, M.M. (1984). *Problems of Dostoevsky's poetics* (C. Emerson, Trans. C. Emerson Ed.). Minneapolis: University of Minnesota Press.

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

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Budanitsky, A., & Hirst, G. (2006). Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics, 32*(1), 13–47.

Cress, U. (2013). Mass collaboration and learning. In R. Luckin, S. Puntambekar, P. Goodyear, B. Grabowski, J. Underwood & N. Winters (Eds.), *Handbook of design in educational technology* (pp. 416–424). New York: Routledge.

Dascalu, M. (2014). *Analyzing discourse and text complexity for learning and collaborating, Studies in Computational Intelligence* (Vol. 534). Switzerland: Springer.

Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., & Nardy, A. (2014). Mining texts, learners productions and strategies with ReaderBench. In A. Peña-Ayala (Ed.), *Educational Data Mining: Applications and Trends* (pp. 335–377). Switzerland: Springer.

Dascalu, M., Dessus, P., Trausan-Matu, S., Bianco, M., & Nardy, A. (2013). ReaderBench, an environment for analyzing text complexity and reading strategies. In H. C. Lane, K. Yacef, J. Mostow & P. Pavlik (Eds.), *16th Int. Conf. on Artificial Intelligence in Education (AIED 2013)* (pp. 379–388). Memphis, USA: Springer.

Dascalu, M., Rebedea, T., & Trausan-Matu, S. (2010). A deep insight in chat analysis: Collaboration, evolution and evaluation, summarization and search. In D. Dochev & D. Dicheva (Eds.), *14th Int. Conf. on Artificial Intelligence: Methodology, Systems, Applications (AIMSA 2010)* (pp. 191–200). Varna, Bulgaria: Springer.

Dascalu, M., Rebedea, T., Trausan-Matu, S., & Armitt, G. (2011). PolyCAFe: Collaboration and utterance assessment for online CSCL conversations. In H. Spada, G. Stahl, N. Miyake & N. Law (Eds.), *9th Int. Conf. on Computer-Supported Collaborative Learning (CSCL 2011)* (pp. 781–785). Hong Kong, China: ISLS.

Dascalu, M., Stavarache, L.L., Dessus, P., Trausan-Matu, S., McNamara, D.S., & Bianco, M. (2015). Predicting Comprehension from Students' Summaries. In *17th Int. Conf. on Artificial Intelligence in Education (AIED 2015)* (pp. 95–104). Madrid, Spain: Springer.

Dascalu, M., Trausan-Matu, S., & Dessus, P. (2013a). Cohesion-based analysis of CSCL conversations: Holistic and individual perspectives. In N. Rummel, M. Kapur, M. Nathan & S. Puntambekar (Eds.), *10th Int. Conf. on Computer-Supported Collaborative Learning (CSCL 2013)* (pp. 145–152). Madison, USA: ISLS.

Dascalu, M., Trausan-Matu, S., & Dessus, P. (2013b). Voices' inter-animation detection with ReaderBench – Modelling and assessing polyphony in CSCL chats as voice synergy. In *2nd Int. Workshop on Semantic and Collaborative Technologies for the Web, in conjunction with the 2nd Int. Conf. on Systems and Computer Science (ICSCS)* (pp. 280–285). Villeneuve d'Ascq, France: IEEE.

Dascalu, M., Trausan-Matu, S., & Dessus, P. (2014). Validating the Automated Assessment of Participation and of Collaboration in Chat Conversations. In S. Trausan-Matu, K. E. Boyer, M. Crosby & K. Panourgia (Eds.), *12th Int. Conf. on Intelligent Tutoring Systems (ITS 2014)* (pp. 230–235). Honolulu, USA: Springer.

Dascalu, M., Trausan-Matu, S., Dessus, P., & McNamara, D.S. (2015a). Dialogism: A Framework for CSCL and a Signature of Collaboration. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine & S. Ludvigsen (Eds.), *11th Int. Conf. on Computer-Supported Collaborative Learning (CSCL 2015)* (pp. 86–93). Gothenburg, Sweden: 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)* (pp. 350–354). Poughkeepsie, NY: ACM.

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

Dong, A. (2006). Concept formation as knowledge accumulation: A computational linguistics study. *AIE EDAM: Artificial Intelligence for Engineering Design, Analysis, and Manufacturing, 20*(1), 35–53.

Dong, A. (2009). *The language of design: Theory and computation*. New York, NY: Springer.

Fano, R.M. (1961). *Transmission of information: A statistical theory of communication*. Cambridge, MA: MIT Press.

François, F. (1993). *Pratiques de l'oral. Dialogique, jeu et variations de figures du sens*. Paris: Nathan Pédagogie.

Galley, M., & McKeown, K. (2003). Improving word sense disambiguation in lexical chaining. In G. Gottlob & T. Walsh (Eds.), *18th International Joint Conference on Artificial Intelligence (IJCAI'03)* (pp. 1486–1488). Acapulco, Mexico: Morgan Kaufmann Publishers, Inc.

Graesser, A.C., McNamara, D.S., Louwerse, M.M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36*(2), 193–202.

Halliday, M.A.K., & Hasan, R. (1976). *Cohesion In English*. London: Longman.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd Ed.). New York, NY: Springer.

Holmer, T., Kienle, A., & Wessner, M. (2006). Explicit Referencing in Learning Chats: Needs and Acceptance. In W. Nejdl & K. Tochtermann (Eds.),

Innovative Approaches for Learning and Knowledge Sharing, First European Conference on Technology Enhanced Learning, EC-TEL 2006 (pp. 170– 184). Crete, Greece: Springer.

Hudelot, C. (1994). La circulation interactive du sens dans le dialogue. In A. Trognon, U. Dausendschön-Gay, U. Krafft & C. Riboni (Eds.), *La construction interactive du quotidien* (pp. 15). Nancy, France: Presses Universitaires de Nancy.

Hummel, G.K., Houcke, J.van, Nadolski, R.J., Hiele, T. van der, Kurvers, H., & Löhr, A. (2011). Scripted collaboration in serious gaming for complex learning: Effects of multiple perspectives when acquiring water management skills. *British Journal of Educational Technology, 42*(6), 1029–1041.

Joshi, M., & Rosé, C.P. (2007). Using Transactivity in Conversation Summarization in Educational Dialog. In *SLaTE Workshop on Speech and Language Technology in Education*. Farmington, Pennsylvania, USA.

Jurafsky, D., & Martin, J.H. (2009). *An introduction to Natural Language Processing. Computational linguistics, and speech recognition* (2nd ed.). London: Pearson Prentice Hall.

Kontostathis, A., Edwards, L., Bayzick, J., McGhee, I., Leatherman, A., & Moore, K. (2009). Comparison of Rule-based to Human Analysis of Chat Logs. In P. Meseguer, L. Mandow & R. M. Gasca (Eds.), *1st International Workshop on Mining Social Media Programme, Conferencia de la Asociación Española Para La Inteligencia Artificial* (pp. 12). Seville, Spain: Springer.

Koschmann, T. (1999). Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration. In C. M. Hoadley & J. Roschelle (Eds.), *Int. Conf. on Computer Support for Collaborative Learning (CSCL'99)* (pp. 308–313). Palo Alto: ISLS.

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. *Psychological Review, 104*(2), 211–240.

Lefebvre, H. (2004). *Rhythmanalysis: Space, Time and Everyday Life* (S. Elden & G. Moore, Trans.). London, UK: Continuum.

Linell, P. (2009). *Rethinking language, mind, and world dialogically: Interactional and contextual theories of human sense-making*. Information Age Publishing: Charlotte, NC.

Manning, C.D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval* (Vol. 1). Cambridge, UK: Cambridge University Press.

Manning, C.D., & Schütze, H. (1999). *Foundations of statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Marková, I., Linell, P., Grossen, M., & Salazar Orvig, A. (2007). *Dialogue in focus groups: Exploring socially shared knowledge*. London, UK: Equinox.

McNamara, D.S., Graesser, A.C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.

McNamara, D.S., Louwerse, M.M., McCarthy, P.M., & Graesser, A.C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*(4), 292–330.

Medina, R., & Suthers, D. (2009). Using a contingency graph to discover representational practices in an online collaborative environment. *Research and Practice in Technology Enhanced Learning, 4*(3), 281–305.

Miller, G.A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review, 63*(2), 81–97.

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

Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *7th ACM SIGCOMM conference on Internet measurement* (pp. 29–42). San Diego, CA: ACM.

The New Harvard Dictionary of Music. (1986). Cambridge, MA: Harvard University Press.

Raghunathan, K., Lee, H., Rangarajan, S., Chambers, N., Surdeanu, M., Jurafsky, D., & Manning, C.D. (2010). A Multi-Pass Sieve for Coreference Resolution. In *Conference on Empirical Methods in Natural Language Processing (EMNLP '10)* (pp. 492–501). Cambridge, MA: ACL.

Rebedea, T. (2012). *Computer-Based Support and Feedback for Collaborative Chat Conversations and Discussion Forums.* (Doctoral dissertation), University Politehnica of Bucharest, Bucharest, Romania.

Rosé, C.P., Wang, Y.C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing Collaborative Learning Processes Automatically: Exploiting the Advances of Computational Linguistics in Computer-Supported Collaborative Learning. *International Journal of Computer Supported Collaborative Learning, 3*(3), 237–271.

Salazar Orvig, A. (1999). *Les mouvements du discours: Style, réfèrences et dialogue dans des entretiens cliniques*. Paris, France: L'Harmattan.

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 Publishing.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 97–118). New York, NY: Cambridge University Press.

Schiffrin, D. (1987). *Discourse Markers*. London, UK: Cambridge University Press.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, UK: Cambridge University Press.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal, 27*, 379–423 & 623–656.

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

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge, UK: Cambridge University Press.

Stent, A.J., & Allen, J.F. (2000). Annotating Argumentation Acts in Spoken Dialogue. Rochester, New York, NY: University of Rochester. Computer Science Department.

Stolcke, A., Ries, K., Coccaro, N., Shriberg, J., Bates, R., Jurafsky, D., . . . Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics, 26*(3), 339–373.

Suthers, D., & Desiato, C. (2012). Exposing Chat Features Through Analysis of Uptake Between Contributions. In *45th Hawaii International Conference on System Sciences* (pp. 3368–3377). Maui, HI: IEEE.

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

Teplovs, C. (2008). The Knowledge Space Visualizer: A Tool for Visualizing Online Discourse. In *Workshop on A Common Framework for CSCL Interaction Analysis, ICLS 2008* (pp. 12). Utrecht, Netherland.

Trausan-Matu, S. (2010a). Automatic Support for the Analysis of Online Collaborative Learning Chat Conversations. In P. M. Tsang, S. K. S. Cheung, V. S. K. Lee & R. Huang (Eds.), *3rd Int. Conf. on Hybrid Learning* (pp. 383–394). Beijing, China: Springer.

Trausan-Matu, S. (2010b). 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. (2013). *From Two-Part Inventions for Three Voices, to Fugues and Creative Discourse Building in CSCL Chats*. Unpublished manuscript.

Trausan-Matu, S., Dascalu, M., & Dessus, P. (2012). Textual complexity and discourse structure in Computer-Supported Collaborative Learning. In S. A. Cerri, W. J. Clancey, G. Papadourakis & K. Panourgia (Eds.), *11th Int. Conf. on Intelligent Tutoring Systems (ITS 2012)* (pp. 352–357). Chania, Grece: Springer.

Trausan-Matu, S., Dascalu, M., & Rebedea, T. (2012). A system for the automatic analysis of Computer-Supported Collaborative Learning chats. In C. Giovannella, D. G. Sampson & I. Aedo (Eds.), *12th IEEE Int. Conf. on Advanced Learning Technologies (ICALT 2012)* (pp. 95–99). Rome, Italy: IEEE.

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. doi: 10.1007/s11412-014-9190-y

Trausan-Matu, S., & Rebedea, T. (2010). A polyphonic model and system for inter-animation analysis in chat conversations with multiple participants. In A. F. Gelbukh (Ed.), *11th Int. Conf. Computational Linguistics and Intelligent Text Processing (CICLing 2010)* (pp. 354–363). New York: Springer.

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/Prentice Hall.

Trausan-Matu, S., & Stahl, G. (2007). Polyphonic inter-animation of voices in chats. In *CSCL'07 Workshop on Chat Analysis in Virtual Math Teams* (pp. 12). New Brunwick, NJ: ISLS.

Trausan-Matu, S., Stahl, G., & Sarmiento, J. (2006). Polyphonic Support for Collaborative Learning. In Y. A. Dimitriadis, I. Zigurs & E. Gómez-Sánchez (Eds.), *Groupware: Design, Implementation, and Use, 12th International Workshop (CRIWG 2006)* (pp. 132–139). Medina del Campo, Spain: Springer.

Trausan-Matu, S., Stahl, G., & Sarmiento, J. (2007). Supporting polyphonic collaborative learning. *Indiana University Press, E-service Journal, 6*(1), 58–74.

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

Upton, G., & Cook, I. (2008). *A Dictionary of Statistics*. Oxford: Oxford University Press.

van Dijk, T.A. (1977). Coherence *Text and Context: Exploration in the Semantics and Pragmatics of Discourse* (pp. 93–129). London, UK: Longman.

van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

Vygotsky, L.S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.

Wegerif, R. (2006). A dialogical understanding of the relationship between CSCL and teaching thinking skills. *International Journal of Computer-Supported Collaborative Learning, 1*(1), 143–157.

Weinberger, A., & Fischer, F. (2006). A Framework to Analyze Argumentative Knowledge Construction in Computer-Supported Collaborative Learning. *Computers & Education, 46*, 71–95.

Wu, Z., & Palmer, M. (1994). Verb semantics and lexical selection. In *32nd Annual Meeting of the Association for Computational Linguistics, ACL '94* (pp. 133–138). New Mexico, USA: ACL.