Chapter 9
ReaderBench (3) – Involvement and Collaboration Assessment through Cohesion and Dialogism

Although participants' involvement in chat environments has been studied in previous systems, as mentioned in Overview of Empirical Studies, *ReaderBench* has brought a series of remarkable improvements in terms of collaborative learning:

- Emphasis and better support of the dialogical and polyphonic model previously proposed in *PolyCAFe* with new visualizations and evaluation factors.
- Refinement of the initial collaboration assessment model (Trausan-Matu et al. 2012b; Dascalu et al. 2010a) based on the social knowledge-building effect, through the use of the cohesion graph (Trausan-Matu et al. 2012a; Dascalu et al. 2013b).
- A novel collaboration evaluation model based on the overlapping effect of voices seen as semantic chains (see *7.5* Dialogism and Voice Inter-Animation) pertaining to different participants.
- The validation of the evaluation mechanics on a long-term discussion group, seen as an aggregation of multiple threads across a longer timespan, and not only the assessment of individual chat conversations (Nistor et al. 2013a; Nistor et al. 2013b, submitted; Nistor et al. 2013c).

9.1 Participant Involvement Evaluation

Besides the identification of topics in the discussion for each participant, significant for pinpointing out the covered concepts, *ReaderBench* also supports participant interaction modeling covering a deeper qualitative dimension, obtained by considering the utterance scores (see *7.4* Cohesion-based Scoring Mechanism). Internally, an interaction graph is built with participants as nodes and the weight of links equal to the sum of interventions scores multiplied by the cohesion function with the referred element of analysis, extracted from the cohesion graph. Therefore, by performing social network analysis (see *3.2* Social Network Analysis) on the

M. Dascalu, *Analyzing Discourse and Text Complexity for Learning and Collaborating*, Studies in Computational Intelligence 534,
DOI: 10.1007/978-3-319-03419-5_9, © Springer International Publishing Switzerland 2014

previous participant interaction graph, the scale of analysis is shifted towards an individual perspective, centered on each of the participants. In the end, the size of each node in the interaction graph is directly proportional to its corresponding betweenness score (Brandes 2001; Bastian et al. 2009). Due to the fact that for chat conversations we are dealing in most cases with a complete graph in which the betweenness score for all nodes is 0, participants are displayed as points (see Figure 50). As cohesive links can exist between utterances pertaining to the same speaker, the visualization also includes the inner links equal to the importance of the utterances expressed as a continuation of the discourse, pertaining to the same participant; for some conversations, these values can be comparable in strength to the sum of all other outgoing links, marking an individual behavior instead of collaboration. Similar mechanics, when employed on a larger discussion group or community obtained from an aggregation of multiple conversations (chat sessions or forum discussion threads), become more meaningful and provide a clearer global perspective of the interactions between participants (see *9.3* Long-term Discussion Groups Evaluation). Moreover a clear separation must be made: personal involvement is expressed as the cumulative utterance importance scores, whereas the interaction graph reflects the exchange of information through cohesive links, making the two perspectives complementary one to another.

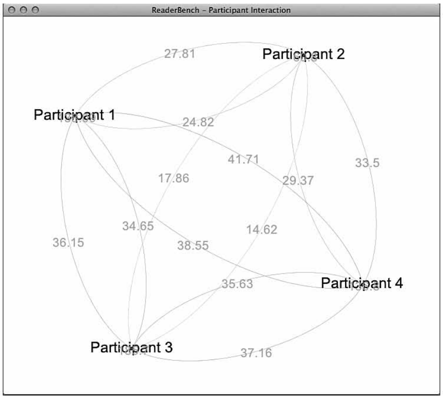

Fig. 50 *ReaderBench* (3) Participant centered view of the interaction graph. The strength of the link between two speakers is reflected in the cumulated effect of each intervention measured through its importance score and reflected in cohesion. In case of a chat conversation with a reduced number of participants, it is most likely to obtain in the end a complete graph, in which the betweenness scores for all nodes are equal to 0, implicitly reducing their diameter to 0

Moreover, an evolution graph of each participant's involvement throughout the conversation, similar to the visualizations provided by *Polyphony* (Trausan-Matu et al. 2007a) and *A.S.A.P.* (Dascalu et al. 2008a) (see *5.1 A.S.A.P.* – Advanced System for Assessing Chat Participants) is also generated (see Figure 51.a), useful for observing interaction patterns. For example, zones with a high slope for one participant are usually in the detriment of the involvement of others and represent areas of the conversation dominated by one participant. On the contrary, comparable growths of multiple participants in a given area induce an equitable involvement and possibly, although not mandatory, collaboration seen as building collaborative knowledge among multiple participants. In the particular case presented in Figure 51.b, all the utterances from the conversation transcript with the identifier from 220 up to 235 pertain solely to Participant 3, from 242 to 261 only two interventions do not belong to Participant 4, whereas Participant 1 completely dominated the discussion between 288 up to 300. Therefore the generated graph is clearly useful for highlighting zones with differential involvement of participants in the conversation.

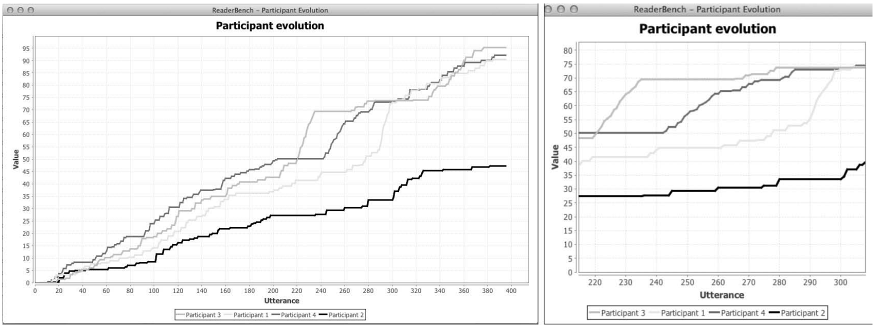

Fig. 51 *ReaderBench* (3) Participants' involvement evolution graph. a. global view of the entire discussion; b. expansion segment of a. around utterance 260.

Following the transition from a global view of the discourse to a user-centered perspective, a similar visualization component of the conceptual space for each chat participant as a mind-map, based on semantic similarities between concepts, is generated (see Figure 52 and *7.3* Topics Extraction). Terms central to a given discussion may not appear in any utterance but, nonetheless, be worth displaying for comprehension's sake. We thus enriched the previously identified participant's topics list with inferred concepts, not mentioned within the text, but the actual visualization component has a lowered threshold (30% in this particular case) as more diverse concepts are used throughout the conversation, with a smaller overall cohesion in comparison to reading materials. Moreover, as the identified list of topics per participant is much more dispersed and has a lower intrinsic cohesion in comparison to a reading material, we opted for eliminating the visualization of the list of inferred concepts as it was misleading; therefore, the inferred concepts are only displayed within the network (see Figure 52).

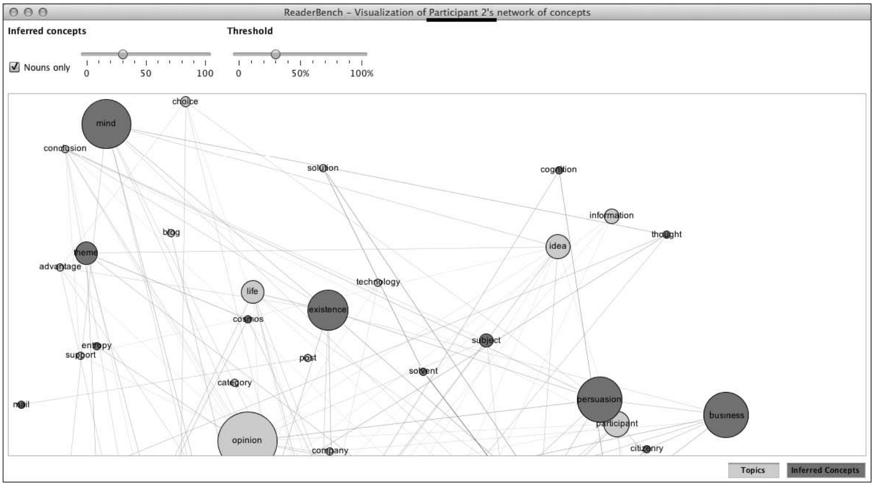

Fig. 52 *ReaderBench* (3) Network of concepts generated for a specific participant

9.2 Collaboration Assessment

In order to thoroughly assess collaboration, we have proposed two computational models. The first model (Dascalu et al. 2013b) based on the effect of social knowledge-building, is a refinement of the gain-based collaboration assessment (Dascalu et al. 2010a; Trausan-Matu et al. 2012b) (see *6.4.2* Collaboration Assessment) and takes full advantage of the cohesion graph (Trausan-Matu et al. 2012a). The second is a novel approach that evaluates collaboration as an intertwining or overlap of voices pertaining to different speakers. The main difference between the two is that the first focuses on the ongoing conversations, therefore on its longitudinal dimension, whereas the later considers subsequent slices of the conversation, the synergy of voices, in other words the transversal dimension. By applying a greedy algorithm (Cormen et al. 2009) on both approaches, the overlap between the identified intense collaboration zones is remarkable.

9.2.1 Social Knowledge-Building Model

The actual information transfer through cohesive links from the cohesion graph obtains two valences by enforcing a personal and social knowledge-building process (Scardamalia 2002; Bereiter 2002; Stahl 2006b) at utterance level. Firstly, a *personal dimension* emerges by considering utterances with the same speaker, therefore modeling an inner voice or continuation of the discourse. Secondly, inter-changed utterances having different speakers define a *social perspective* that models collaboration as a cumulative effect. Although similar to some extent to the gain-based collaboration model (Dascalu et al. 2010a; Trausan-Matu et al. 2012b), the transition towards Stahl's model of collaborative knowledge-building (see Figure 4) and the use of the multi-layered cohesion graph instead of the utterance graph are the main differentiators when addressing this computational knowledge-building model that enables a deeper and a more generalized analysis of collaboration in CSCL conversations.

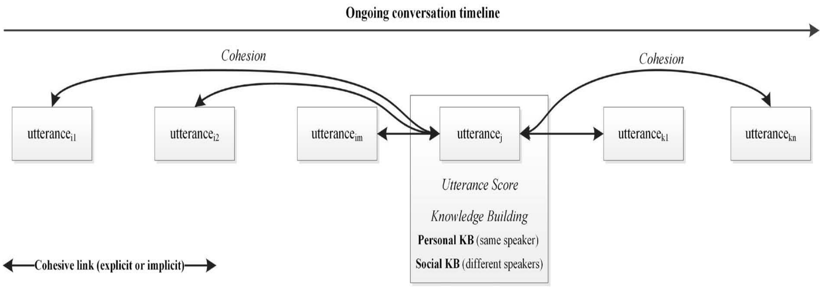

Fig. 53 *ReaderBench* (3) Slice of the cohesion graph depicting inter-utterance cohesive links used to measure personal and social knowledge-building effects

Therefore, each intervention or utterance now has its previously defined importance score (see *7.4* Cohesion-based Scoring Mechanism) and a knowledge-building (KB) effect, both personal and social (see Figure 53). The personal effect is initialized as the intervention's score, whereas the social effect is zero. Later on, by considering all the links from the cohesion graph, each dimension is correspondingly augmented: if the link is between utterances with the same speaker, the previously built knowledge (both personal and social) from the referred utterance is transferred through the cohesion function to the personal dimension of the current utterance; otherwise, if the pair of utterances is between different participants, the social knowledge-building dimension of the currently analyzed utterance is increased with the same amount of information (previous knowledge multiplied by the cohesion measure). In other words, continuation of ideas or explicitly referencing utterances of the same speaker builds an inner dialogue or personal knowledge, whereas the social perspective measures the interaction with other participants, encourages sharing of ideas, fostering creativity for working in groups (Trausan-Matu 2010b) and influencing the other participants' points of view during the discussion, thus enabling a truly collaborative discussion.

In this manner we can actually measure collaboration through the sum of social knowledge-building effects, starting from each intervention's score corroborated with the cohesion function. Moreover, personal knowledge-building addresses individual voices (participant voices or implicit/alien voices covering the same speaker), while social knowledge-building, derived from explicit dialog (that by definition is between at least two entities), sustains collaboration and highlights external voices. By referring to the dialogic model of discourse analysis, besides voices that are derived from the semantic chains in correlation to each participant's point of view, echoes are reflected by cohesion in terms of the information transferred between utterances, whereas the attenuation effect diminishes the strength of the cohesion link with the increase in distance between the analysis elements (see *7.2* Cohesion-based Discourse Analysis).

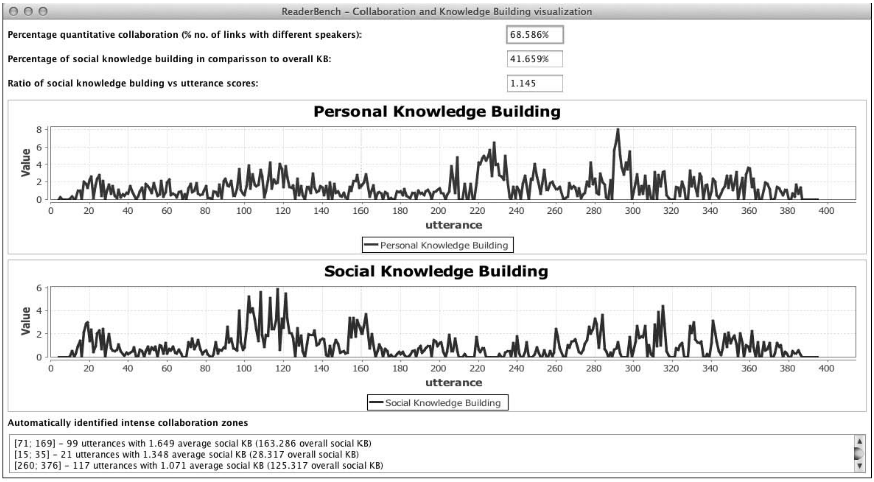

Fig. 54 *ReaderBench* (3) Collaboration assessment and its evolution in time. The interface introduces from top to bottom: a. the 3 overall collaboration factors as an overview of the conversation; b. individual graphs depicting the personal and social knowledge-building evolution throughout the entire discussion; c. the automatically identified intense collaboration zones with their corresponding span and cumulated social knowledge-building effect.

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

In addition to the estimation of personal and social knowledge-building effects for each utterance and the modeling of their corresponding evolution throughout the conversation (see Figure 54), *ReaderBench* automatically identifies *intense* *collaboration zones* that are intervals of utterances in which participants are actively involved, collaborate and generate new ideas related to the ongoing context of the discussion. The first step within our greedy algorithm (Cormen et al. 2009) exploited in order to build up intense collaboration zones consists of identifying social knowledge-building peaks as maximum local values. Afterwards, each peak is expanded sideways within a predefined slack (experimentally set at 2.5% of the number of utterances); this slack was important due to our focus on the macro-level analysis of collaboration and due to the possible intertwining of multiple discussion threads. In the end, only zones above a minimum spread of 5 utterances are selected as intense collaboration zones.

In other words, after identifying the utterances with the greatest social knowledge-building effect, the algorithms expands each zone to the left and to the right, in a non-overlapping manner to previously identified zones, by considering utterances above the mean social knowledge-building value and that are in the previously defined slack. If in the end, the zone covers more than the specified minimum spread, it is considered an intense collaboration zone. From a different point of view and highly related to the process of identifying social knowledge-building, cohesion binds utterances within an intense collaboration zone in terms of on-topic relatedness.

From a holistic perspective addressing the conversation viewed as a whole, three factors were implemented in order to best characterize the overall collaboration within the discussion (see Figure 54). Firstly, *quantitative collaboration* is determined as the percentage of links from the cohesion graph having different speakers in comparison to the number of links automatically identified. Although rough as estimation, this measurement provides good insight with regards to the actual information exchange between participants. Secondly, the *overall social knowledge-building score* is compared to the overall knowledge-building effect. Thirdly, the *ratio* between the *overall social knowledge-building score* and the *overall utterance importance scores* is computed for highlighting the amount of information that is transferred through collaboration in comparison to what was withheld initially within each utterance.

9.2.2 Dialogical Voice Inter-Animation Model

In order to achieve genuine collaboration, the conversation must contain a dense intertwining of voices derived from key concepts and covering multiple participants of the conversation (Trausan-Matu and Rebedea 2009; Trausan-Matu in press). In order to obtain a computational model, a shift of perspective is required, from the voice synergy effect, towards the participant's point of view. As collaboration is centered on multiple participants, a split of each voice into multiple viewpoints pertaining to different participant is required (see Figure 55). A viewpoint consists of a link between the concepts pertaining to a voice and a participant, through their explicit use within one's interventions in the ongoing conversation. Moreover, we opted to present this split in terms of implicit (alien) voices (Trausan-Matu and Stahl 2007), as the accumulation of voices through transitivity in inter-linked cohesive utterances clearly highlights the presence of alien 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 discussion.

In addition, in order to identify the voice overlaps now pertaining to different participants, we changed from an ongoing longitudinal analysis of the discourse, presented in the previous section, to a transversal analysis of a context consisting of five adjacent utterances (with a possible shortening of the window, if the pause between adjacent utterances is greater than the imposed threshold) (see *7.5* Dialogism and Voice Inter-Animation). Subsequently, in order to evaluate collaboration following the conversation's timeline, we used a sliding window that models through its replication the overlap of voices pertaining to different participant in different contexts. More specifically, we use a cumulated value of pointwise mutual information (PMI) obtained from all possible pairs of voices pertaining to different participants (different viewpoints), within subsequent contexts of the analysis (see Figure 56). In the end, a similar process of identifying intense collaboration zones based on the greedy algorithm described in the previous section is applied.

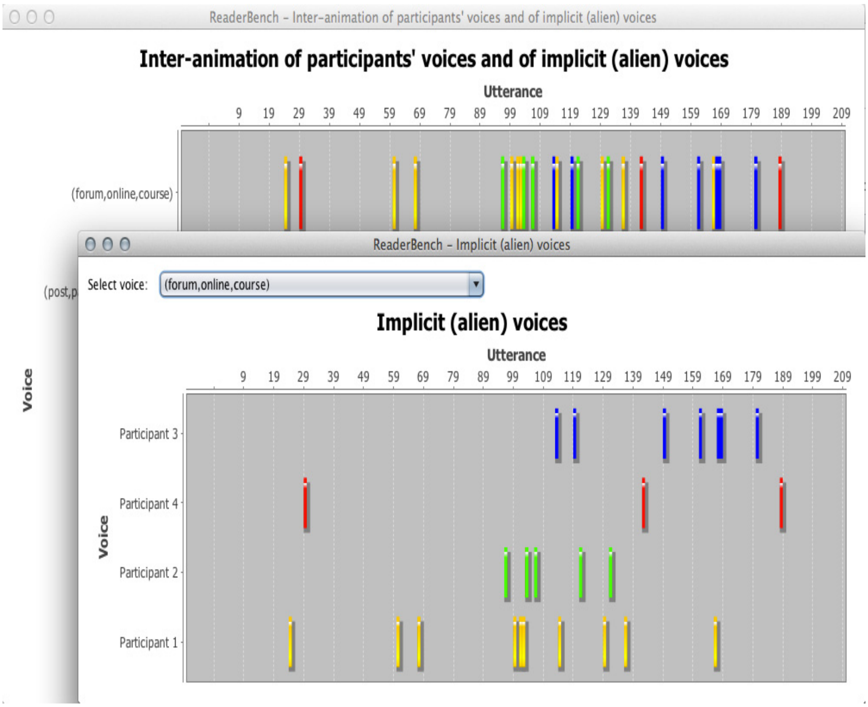

Fig. 55 *ReaderBench* (3) Implicit (alien) voices split per participant and spread throughout the conversation. The window frame from the background depicts the *(forum, online, course)* voice that was split per participant in order to highlight personal coverage of the conveyed concepts. The initial distribution of the voice can be obtained by overlapping the individual implicit (alien) voices for all participants.

<Turn nickname="Participant 2"><Utterance genid="134" time="03.47.38" ref="130">wiki wiki means rapidly in Hawaiian language</Utterance></Turn>

<Turn nickname="Participant 3"><Utterance genid="135" time="03.48.31" ref="0">the forum was the place where in roman times people used to come and talk business</Utterance></Turn>

<Turn nickname="Participant 1"><Utterance genid="136" time="03.49.01" ref="135">and now the next best thing could be the blog – where someone shares its knowledge</Utterance></Turn>

<Turn nickname="Participant 2"><Utterance genid="137" time="03.49.22" ref="134">so it is a very quick way of letting others know what you have discovered</Utterance></Turn>

<Turn nickname="Participant 4"><Utterance genid="138" time="03.50.31" ref="136">yes, but knowledge is stored in books</Utterance>.

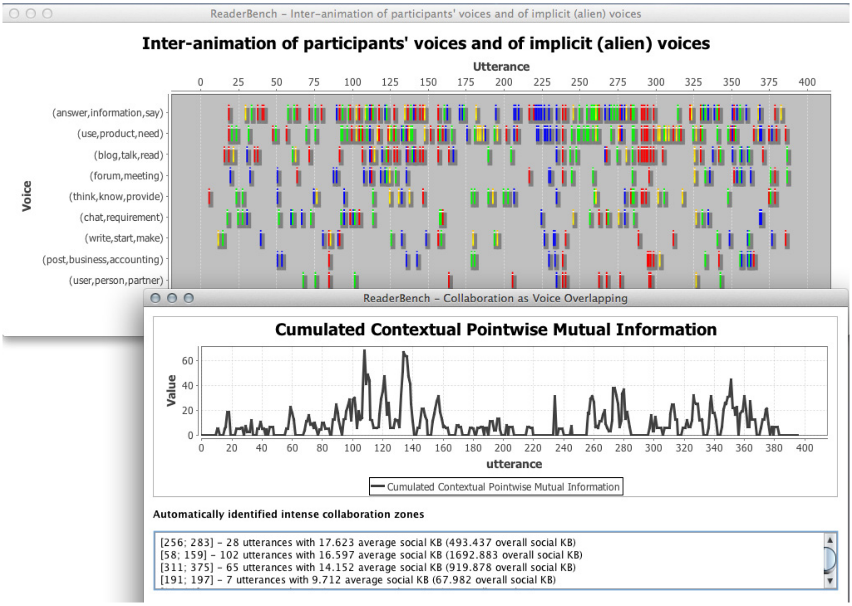

Fig. 56 *ReaderBench* (3) Collaboration evolution viewed as voice overlaps between different participants (intertwining of different viewpoints), including the automatic identification of intense collaboration zones. The presented participant meaningful interventions denote a peak value in the collaboration evolution graph in the [134; 138] range where multiple voices pertaining to all conversation participants (e.g., knowledge, wiki, forum, blog, chat, people) co-occur.

The inter-animation frame from Figure 56 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 mentioned: 1/ as the algorithm uses the moving averages and applies PMI on sliding windows, the user must also consider a frame of 5 utterances in which each individual occurrence is equally dispersed (if not the case of a split horizon due to a pause in the conversation) and 2/ all the voices from the conversation are considered (even those that have as low as 3 constituent words); this explains greater cumulative values encountered in the graph (e.g., the excerpt centered on utterance 136 in which all conversation participants are engaged and in which multiple concepts, pertaining to different voices, are encountered).

As an analogy, from an individual point of view, participant's overall collaboration can be seen as the cumulated mutual information between his viewpoints and all other participant viewpoints. In other words, for a given participant, we compare through mutual information his viewpoint or individual voice distribution to all other speakers' viewpoints, for all voices identified in the conversation. Therefore, by comparing individual voice distributions that span throughout the discussion, collaboration emerges from the overlap of viewpoints pertaining to different participants.

9.2.3 Validation of Collaboration Assessment

Preliminary experiments (Dascalu et al. 2013b) were conducted in order to validate the dialogic models used for evaluating chat conversations, with emphasis on participant involvement and collaboration assessment. Three chat conversations conducted in an academic environment, with students from the 4th year undergoing the Human-Computer Interaction course and debating on CSCL technologies, were manually assessed by 4 tutors. More specifically, each student had to focus on a CSCL technology (chat, wiki, blog or forum), to present and debate on its benefits in specific use case scenarios generated throughout the conversation. These three conversations (Team 4, Team 34 and Team 36) were selected for detailed analysis after an overview of approximately 50 discussions engaging more than 200 students. Although high discrepancies were noticed in terms of the quality of the content, the involvement and the collaboration of its participants, these conversations were considered representative for the entire sample and the preliminary evaluations were conducted only on these conversations due to the high amount of time it takes to manually assess a single chat conversation (2 to 4 hours for a deep understanding of involvement and of collaboration).

Additionally, the time evolution interface depicted in Figure 57.a was developed in order to facilitate the manual evaluation of chats in terms of intense collaboration zones. In this context, the presentation of the conversation follows the timeline and models the intertwining of utterances, based on the cohesion graph. This component is useful for manually identifying: 1/ breaks within the conversation, zones with limited or no collaboration, due to the fact than within a specific time-frame we have a monologue of a participant, without any interventions from other users, and 2/ zones with high collaboration due to the dense inter-animation of utterances between different participants. In the particular case presented in Figure 57, all utterances with identifiers between 27 and 50 belong a single user, within a limited timeframe, therefore making the social knowledge-building effect zero. Afterwards, as multiple participants get involved in the ongoing discussion, collaboration increases.

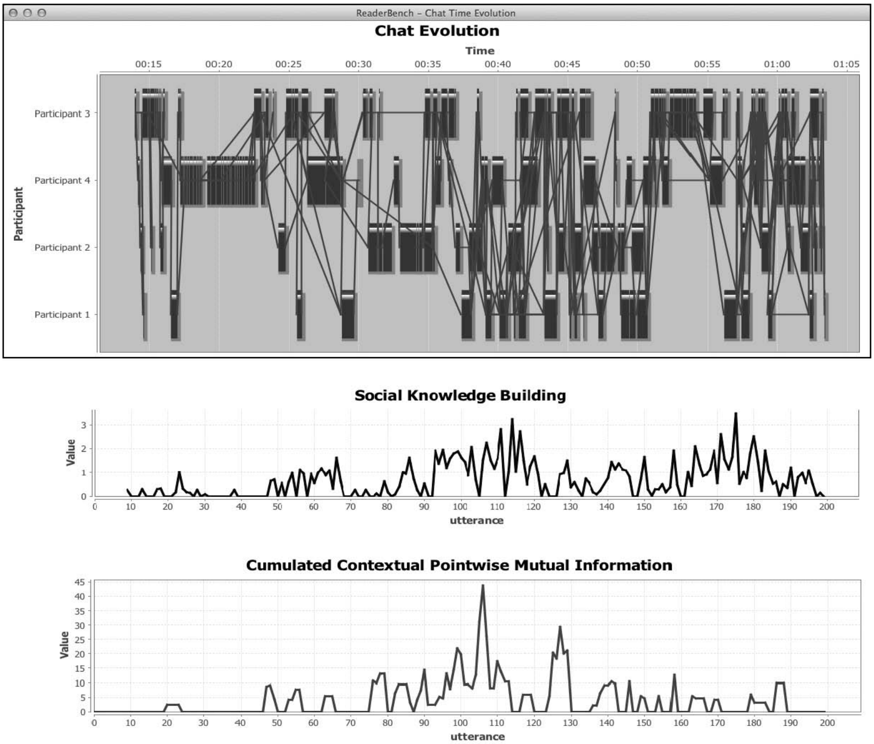

Fig. 57 *ReaderBench* (3) Time slice of a conversation highlighting cohesion links and a monologue. Matching graphs of: a) Evolution in time of the chat conversation; b) Collaboration evolution seen as social knowledge-building; c) Collaboration evolution derived from voice overlapping

Table 31 presents the correlation between different evaluation factors extracted from *ReaderBench* and the final grades assigned by the experts. Although the participant's identifiers coincide, each conversation had different students attending it. Moreover, in order to ensure the equitability of our analysis, the correlations between the factors automatically determined by the system and the average values of the grades manually assigned by the experts were computed after combining the participants' scores from all conversations.

As an interpretation of the results presented in Table 31, we can observe in Team 4 conversation a discrepancy, as the involvement of the participants from a personal point of view was good, while the actual collaboration throughout the conversation is highly unbalanced. Team 34 conversation has the lowest scores in all the factors, whereas Team 36 conversation, that was considered the best by the tutors in terms of both quality and involvement, has the highest scores assigned by the system.

Table 31 *ReaderBench* (3) Correlation between manual and automatic participants' evaluations

Participant Name	No. Utter.	Overall Utter. Score	Overall Personal KB	Overall Social KB	MI Viewpoint Overlap	Expert Grade 1	2	3	4	Avg.
				Team 4						
Participant 1	90	90.53	138.09	99.77	223.92	9.0	10.0	8.0	9.5	*9.13*
Participant 2	61	47.22	60.80	75.04	199.55	8.0	9.0	8.5	10.0	*8.88*
Participant 3	120	95.18	185.70	86.44	232.39	8.0	6.5	9.0	8.0	*7.88*
Participant 4	118	92.24	136.80	111.05	240.61	9.0	10.0	8.0	9.5	*9.13*
				Team 34						
Participant 1	23	21.43	35.34	22.77	82.20	6.0	6.0	7.0	6.0	*6.25*
Participant 2	34	25.02	32.69	25.39	105.30	5.0	7.0	7.0	6.0	*6.25*
Participant 3	73	44.83	74.80	44.72	100.03	8.0	7.0	8.0	7.0	*7.50*
Participant 4	60	45.38	85.41	33.68	110.57	7.0	4.0	7.0	5.5	*5.88*
				Team 36						
Participant 1	54	55.53	71.11	99.61	223.53	9.0		9.5	8.0	*8.83*
Participant 2	67	69.91	95.20	111.83	313.56	10.0		8.0	10.0	*9.33*
Participant 3	119	134.45	236.91	145.82	288.53	9.0		8.0	8.5	*8.50*
Participant 4	57	60.91	81.77	98.66	271.21	9.0		9.5	8.0	*8.83*
				Overall – all conversations						
Correlation	.64	.79	.69	.89	.84					

Moreover, by analyzing each factor's correlation, it becomes quite clear that the tutors emphasized on the quality of interventions, not on the mere number of utterances or their interdependencies. Additionally, the social knowledge-building dimension and the collaboration extracted from the mutual information of participant viewpoints are better correlated with the expert's grades; this sustains that collaboration was more important in the expert's evaluation than the personal effect of each participant, reflecting his/her involvement. As the Intraclass Correlation Coefficient (ICC) was *.61* on single measures and *.86* on average, results in terms of intervention scores (qualitative involvement evaluation) and social knowledge-building and mutual information between viewpoints (collaboration assessment at individual level) correlate extremely well with the average expert grades.

Table 32 *ReaderBench* (3) Overlap between manual and automatic identification of intense collaboration zones

Conversation	Number of utterances	Manually annotated collaboration zones	Automatically identified intense collaboration zones	
			Social KB	Voice Overlap
Team 4	389	[90; 160]	[15; 35]	[33; 39]
		[320; 360]	[71; 169]	[58; 159]
			[197; 208]	[191; 197]
			[238; 245]	[256; 283]
			[260; 376]	[311; 375]
Team 34	190	[90; 120]	[48; 66]	[47; 56]
		[170; 178]	[93; 121]	[76; 129]
			[127; 134]	[138; 170]
			[140; 150]	
			[158; 184]	
			[190; 195]	
Team 36	297	Relatively uniform distribution	[21; 126]	[18; 124]
			[136; 182]	[139; 149]
			[199; 241]	[188; 196]
			[270; 288]	[205; 217]
				[249; 257]
				[271; 287]

In terms of intense collaboration zones, manual annotations and automatically identified zones are presented in Table 32, whereas the comparison between the zones identified through the two automatic collaboration assessment methods is covered in Table 33. The manual annotations were not covered in the later table as for Team 36 the tutors agreed that collaboration was uniformly distributed, thus making an automatic comparison inapplicable. Moreover, by analyzing the results from Table 32 and Table 33 we observe a good overlap in terms of accuracy measured as precision, recall and F-score (Manning et al. 2008) between the two computational models. This proves that one model is consistent with the other, but also a good match with the tutor annotations, therefore demonstrating the feasibility of our two approaches. The rather low correlation scores in Table 33 are completely justifiable as the two models are built from orthogonal dimensions of conversation analysis and consider completely different mechanics of evaluation, but in the end both properly address the purpose of identifying intense collaboration zones.

Table 33 *ReaderBench* (3) Overlap measurements between automatic models used to identify intense collaboration zones

Conversation	Precision	Recall	F-score	Correlation
Team 4	.87	.71	.78	.41
Team 34	.68	.65	.67	.27
Team 36	.88	.67	.76	.48

Based on the previous analyses, the following indicators of bad collaboration (mostly in Team 34 conversation) were observed: 1/ the high number of automatically identified zones containing 1 to 3 utterances, which were not considered intense collaboration zones in the end, 2/ the low average value of social knowledge-building effect and 3/ no automatically identified collaboration zone with a wide spread (over 50 utterances, although it was the shortest conversation of the three). In contrast, conversations with good collaboration (namely Team 36 which had the best overall collaboration) have: 1/ higher average values of social knowledge-building and 2/ a more balanced distribution and higher coverage of the entire conversation in terms of the automatically identified intense collaboration zones.

Additionally, we have performed an evaluation for proving the inter-dependencies between the two collaboration assessment models: starting from all explicit links added by users in the chat environment (Holmer et al. 2006), we have measured the correlations between the cohesion scores and the similarities between utterances in terms of voice distributions; the later similarity is computed as a Pearson correlation between the utterances' voice occurrences. Linked with the nature of the evaluations (overlap of semantic chains versus an aggregated cohesion function), results were though medium: *average* $r = .46$, with r(Team 4 with 106 explicit links) = .54, r(Team 34 with 76 explicit links) = .48 and r(Team 36 with 226 explicit links) = .34.

Although the perspectives of the two collaboration assessment models are orthogonal while observing the unfolding of a conversation, there are multiple resemblances between the two proposed computational models. Firstly, the evaluation process of collaboration is based, in some extent, on the exchange of information between different participants; whereas in the first case, cohesion expresses the strength of the link in terms of the social knowledge-building effect between interventions of different chat participants, in the second voice overlaps are considered only while comparing different viewpoints and the exchange is expressed through mutual information. Secondly, cohesion, seen as a link between analysis elements and an equivalent to a voice's echo, is caught in some degree through the process of overlapping occurrences of semantic chains, smoothed in predefined conversational contexts. Thirdly and most importantly, although one method is based on the effect of social knowledge-building and the other on the intertwining or overlap of voices belonging to different speakers, both computational models support dialogism and emphasize the dialogical perspective of collaboration in CSCL environments.

9.3 Long-Term Discussion Groups Evaluation

Starting from the analysis of a single conversation, our aim in terms of assessing discussion groups consists of providing an automatic aggregation facility of multiple conversations, of building a global social network with all the involved participants and of verifying the validity of the automatic analysis proposed in *ReaderBench*, applied on a larger scale. Long-term discussion groups depict a set a participants or members of that group involved in subsequent conversations, over a longer timespan. This discrepancy between a local view, initially introduced in *A.S.A.P.* and *Ch.A.M.P.*, and continued by *PolyCAFe*, and a global one has multiple implications as specific technical aspects needed to be taken into consideration when merging a multitude of discussion threads. Therefore, from a technical perspective, the shift required a normalization of individual conversation scores, performing a distributed analysis due to the size of the corpus of discussion threads (Dascalu et al. 2011a) (see *6.4.5* Distributed Computing Framework) and building a global interaction graph between all the participants. In the end, in order to perform the validation of the automatic importance scores, a critical thinking assessment framework was used to annotate the relevance of members' messages (Weltzer-Ward et al. 2009).

Moreover, we specifically limited the perspective to long-term discussion groups, without clearly pinpointing at this moment the particularities of communities of practice (CoP) (Wenger 1999; Lave and Wenger 1991), as further refinements of our automatic assessment procedure are required to best fit the specificity of such communities. Nevertheless, the overall conducted study (Nistor and Fischer 2012; Nistor et al. 2013a; Nistor et al. 2013b) was positioned at the intersection of development of the expert status in CoP (Nistor 2010; Nistor and Fischer 2012) and technology acceptance (Bagozzi 2007; Nistor et al. 2012). In other words, in terms of educational practice, the conducted study represented an extended application of *ReaderBench* towards monitoring and assessing participation and collaboration (Strijbos 2011) in communities.

The study included $N = 179$ participants (20 full-time faculty employees and 159 part-time faculty members), all of them holding a doctoral degree. The automatic analysis was focused on 3 variables extracted from all the messages of the asynchronous forum discussions (7370 interventions) available between August 2010 and June 2012: *participation*, *expertise* and *expert status* (Nistor et al. 2013a) (see Table 34). The intensity of participation was operationalized as the number of interventions of each group member; the quality of these interventions (utterance scores determined by *ReaderBench*) was considered an indicator of expertise; expert status was measured as in-degree and betweenness centrality within the group interaction graph (see Figure 58).

Table 34 *ReaderBench* (3) General long-term discussion group statistics ($N = 179$)

Factor	Mean	Standard deviation
1. *Participation*(number of interventions)	41.17	95.07
2. *Expertise* (quality of interventions or cumulative utterance importance scores)	160.53	418.94
3. *Expert status* (In-degree from interaction graph)	289.34	761.43
4. *Expert status* (Betweenness from interaction graph)	266.08	703.20

More specifically, through the dynamics of the interaction and through the quality of the interventions, a member of the discussion group obtains in time the status of expert. In addition, as most discussions followed a simple pattern – an inquiry, administrative or related to educational sciences, was initially formulated as a new thread and other members of the group responded subsequently -, participation can be considered mediated by the quality of the interventions, as only members with valuable insight contribute in each academic discussion thread (Nistor and Fischer 2012). Moreover, participation influences the expert status determined after a longer timespan and reflected through specific SNA factors in a central position within the group. These dimensions of the analysis, with corresponding interdependencies, were later on studied in extent in (Nistor et al. 2013b; Nistor et al. 2013a).

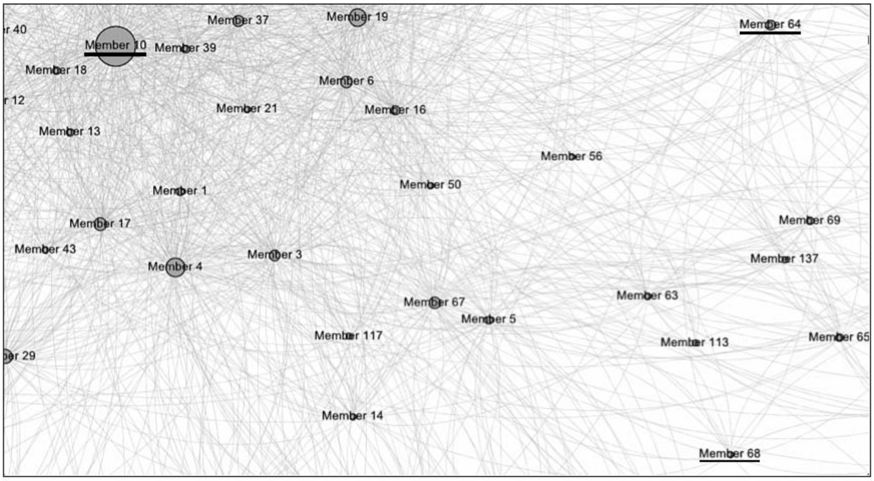

Fig. 58 *ReaderBench* (3) Partial view of a group interaction graph. A clear demarcation can be observed between different types of users: e.g., *Member 10*, by far the most actively involved member (830 interventions, in contrast to the second and third most active members: *Member 26* – 510 and *Member 19* – 458) and *Member 64* (68 interventions) or *Member 68* (12 interventions).

Although Table 34 depicts a high discrepancy across the group members in terms of involvement and of participation as the standard deviation values are approximately three times greater than the averages, these values are consistent for all analysis variables. Moreover, as expected because we are dealing with a large group equitable between members, the cross correlations between the variables were high ($r > .70$). This can be also explained from the perspective that the analysis was focused on topics with a broad diversity and that the involvement of members, with similar backgrounds, within an academic environment, was balanced in terms of the impact of each intervention measured in the utterance's importance score.

In contrast, if we analyze the average quality of each intervention per member (expertise divided by participation), there are rather high fluctuations for all members of the group ($M = 6.62$, $SD = 4.67$) and for the 10 most actively involved members in terms of participation ($M = 9.20$, $SD = 4.03$), with no correlation to any other variable. This allows us to consider that, although participation and expertise are highly correlated, this is a cumulative effect induced at group level, without any direct dependency between quantitative and qualitative evaluations of interventions. Nevertheless, the high correlation also resides in the intrinsic dependency that more interventions increase the overall importance score, as the scoring function (see *7.4* Cohesion-based Scoring Mechanism) always returns a positive result for each intervention.

Regarding the validation of the cohesion-based scoring mechanism, the bivariate correlation between the average relevance of messages determined manually (Weltzer-Ward et al. 2009) and the cumulative intervention scores per participant was of $r = 0.72$, $p < .001$ (for 414 messages sent by $N = 15$ discussion participants), which clearly demonstrates the adequacy of the scoring mechanism proposed in *ReaderBench*.

The visualization of the entire long-term discussion group was also of particular interest. Although the aggregated interaction graph uses the same measures described in section *9.1* Participant Involvement, the visualization became more relevant when applied on a larger scale, as the *expert status* now is visibly reflected in the dimension of each node (directly proportional to its betweenness score) and in a more central position within the social network graph (see Figure 58). All group member names have been anonymized to avoid privacy issues and the indexes are attributed in the order of first occurrence within the discussion group.

Additional experiments were conducted for splitting the discussions on two topics (research centered and administrative), therefore addressing the specificity of each intervention and focusing on the extraction of two sub-groups, hopefully as disjunctive as possible. List of concepts were manually built through questionnaires administered to 3 domain experts and included in the end 268 words for academic administration, respectively 857 words specific to educational sciences. Based on these lists, a new score of specificity was assigned to each analysis element, equal to its initial cohesion-based importance score multiplied by a normalized coverage of a given topic, seen as lemma overlapping between the predefined list and the words within each intervention. Therefore, besides the overall score that was initially assigned, each group member had a set of cumulative scores based on his/her interventions' specificity with regards to selected topics.

Table 35 *ReaderBench* (3) Statistics on the long-term discussion group specificity analysis ($N = 179$)

Evaluation scenario	$M_{\text{specificity}}$	$SD_{\text{specificity}}$
1. Administration	40.60	116.14
2. Educational sciences	33.53	90.15
3. Overall (equivalent to Expertise)	160.53	418.94

Starting from Table 35 and corroborated with the construction assumptions, we can conclude that: 1/ the group discussions were mostly administratively oriented as we obtained a greater average specificity by using a much shorter list of words; 2/ similar to the general scenarios, there was a high variability in terms of expertise between the members, observable in the standard deviation approximately three times greater than the average value; 3/ the topics had a good cumulated coverage (46.18%) by using only 21.38% of the vocabulary/word lemmas mentioned throughout the group discussions; 4/ although the values suggest a rather clear categorization per topic, the final statistics for the sub-groups, each centered on a topic, have not induced a split of the initial long-term discussion group, but an overlap, as the majority of members addressed both topics throughout their interventions. This suggests a merge of the two topics by observing the entire interventions exchange during the long timespan, without a clear demarcation of membership to a sub-group.

9.4 Comparison of *ReaderBench* to *KSV*

Starting from the presentation of specific systems in *3.1.3* CSCL Computational Approaches, we considered the *Knowledge Space Visualizer – KSV* to be the most similar one to the current *ReaderBench* facilities (see Table 36). Both applications

Table 36 *ReaderBench* versus *KSV* (Teplovs 2008)

Benefits of *ReaderBench*	Benefits of *KSV*
Educational perspective	
Dialogical perspective induced by voice inter-animation	
Emphasis on collaboration in addition to a qualitative participation evaluation	A more shallow perspective of individuals and links between them
Conversation topics extraction relevant for highlighting the focus of the discussion	
The analysis is strictly based on textual information	Integration of addition relationships between ‘notes’ (e.g., annotation, authorial)

Table 36 (*continued*)

Benefits of *ReaderBench*	Benefits of *KSV*
Technical perspective	
Explicit or implicit/cohesive links between interventions are taken into consideration. Cohesion by itself integrates multiple perspective: semantic dimension, Latent Semantic Analysis and Latent Dirichlet Allocation	Multiple types of relations between the nodes are considered: structural (e.g., reply-to, build-on, reference, annotation, contains), authorial, or semantic (based on Latent Semantic Analysis)
Multiple NLP techniques applied on the initial interventions	
Post analysis centered on logs or excepts of conversations (chats and forum discussion threads)	Integration within the *Knowledge Forum* and the encouragement of continual analytic improvement
	Clustering of nodes
Aggregation of multiple discussion threads or conversations	Integration of multiple data sources
Comprehensive mechanism of utterance importance scoring	
	A multitude of parameters, configurations and visualizations available from the interface

envision the visualization of participation and interaction between users through Social Network Analysis and semantic similarities between concepts or analysis elements, but the overall aims differ: 1/ *ReaderBench* is focused mostly on a deep analysis of each conversation/discussion thread with emphasis on involvement and collaboration, with the possibility of automatically aggregating them, whereas 2/ *KSV* was designed especially for obtaining an overview of interactions, with accent on visualization.

References

Bagozzi, R.P.: The legacy of the Technology Acceptance Model and a proposal for a paradigm shift. Journal of the Association for Information Systems 8(4), 244–254 (2007)

Bastian, M., Heymann, S., Jacomy, M.: Gephi: an open source software for exploring and manipulating networks. In: International AAAI Conference on Weblogs and Social Media, San Jose, CA, pp. 361–362. AAAI Press (2009)

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

Brandes, U.: A Faster Algorithm for Betweenness Centrality. Journal of Mathematical Sociology 25(2), 163–177 (2001)

Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C. (eds.): Introduction to Algorithms, 3rd edn. MIT Press, Cambridge (2009)

Dascalu, M., Chioasca, E.-V., Trausan-Matu, S.: ASAP – An Advanced System for Assessing Chat Participants. In: Dochev, D., Pistore, M., Traverso, P. (eds.) AIMSA 2008. LNCS (LNAI), vol. 5253, pp. 58–68. Springer, Heidelberg (2008)

Dascalu, M., Dobre, C., Trausan-Matu, S., Cristea, V., Beyond Traditional, N.L.P.: A Distributed Solution for Optimizing Chat Processing. In: 10th International Symposium on Parallel and Distributed Computing (ISPDC 2011), Cluj-Napoca, Romania, pp. 133–138. IEEE (2011a)

Dascalu, M., Rebedea, T., Trausan-Matu, S.: A deep insight in chat analysis: Collaboration, evolution and evaluation, summarization and search. In: Dicheva, D., Dochev, D. (eds.) AIMSA 2010. LNCS (LNAI), vol. 6304, pp. 191–200. Springer, Heidelberg (2010)

Dascalu, M., Trausan-Matu, S., Dessus, P.: Cohesion-based Analysis of CSCL Conversations: Holistic and Individual Perspectives. In: Rummel, N., Kapur, M., Nathan, M., Puntambekar, S. (eds.) 10th Int. Conf. on Computer-Supported Collaborative Learning (CSCL 2013), Madison, USA, pp. 145–152. ISLS (2013b)

Holmer, T., Kienle, A., Wessner, M.: Explicit Referencing in Learning Chats: Needs and Acceptance. In: Nejdl, W., Tochtermann, K. (eds.) EC-TEL 2006. LNCS, vol. 4227, pp. 170–184. Springer, Heidelberg (2006)

Lave, J., Wenger, E.: Situated learning: Legitimate peripheral participation. Cambridge University Press, Cambridge (1991)

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

Nistor, N.: Knowledge Communities in the Classroom of the Future. In: Mäkitalo-Siegl, K., Zottmann, J., Kaplan, F., Fischer, F. (eds.) Classroom of the Future -Orchestrating Collaborative Spaces, pp. 163–180. Sense Publishers, Rotterdam (2010)

Nistor, N., Baltes, B., Mihaila, D., Dascalu, M., Smeaton, G., Trausan-Matu, S.: Virtual communities of practice in academia: An automated analysis of expertise and expert status. Paper presented at the 15th Biennial EARLI Conference for Research on Learning and Instruction, Munich, Germany, August 27-31 (2013a)

Nistor, N., Baltes, B., Smeaton, G., Dascalu, M., Mihaila, D., Trausan-Matu, S.: Virtual Communities of Practice in Academia: An Automated Discourse Analysis. In: Paper presented at the 1st International Workshop on Discourse-Centric Learning Analytics (DCLA13), a pre-Conference Workshop at Learning Analytics and Knowledge (LAK 2013), Leuven, Belgium, April 8-12 (2013b)

Nistor, N., Dascalu, M., Trausan-Matu, S., Mihaila, D., Baltes, B., Smeaton, G.: Virtual Communities of Practice in Academia: Participation under the Influence of Technology Acceptance and Community Roles. In: Hernández-Leo, D., Ley, T., Klamma, R., Harrer, A. (eds.) EC-TEL 2013. LNCS, vol. 8095, pp. 623–624. Springer, Heidelberg (2013)

Nistor, N., Dascalu, M., Trausan-Matu, S., Mihaila, D., Baltes, B., Smeaton, G.: Virtual Communities of Practice in Academia: Automated Analysis of Collaboration Based on the Social Knowledge-Building Model. In: Hernández-Leo, D., Ley, T., Klamma, R., Harrer, A. (eds.) EC-TEL 2013. LNCS, vol. 8095, pp. 623–624. Springer, Heidelberg (2013)

Nistor, N., Fischer, F.: Communities of practice in academia: Testing a quantitative model. Learning, Culture and Social Interaction 1(2), 114–126 (2012)

Nistor, N., Lerche, T., Weinberger, A., Ceobanu, C., Heymann, O.: Towards the integration of culture into the Unified Theory of Acceptance and Use of Technology. British Journal of Educational Technology (2012), doi:10.1111/j.1467-8535.2012.01383.x

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

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

Strijbos, J.W.: Assessment of (Computer-Supported) Collaborative Learning. IEEE Transactions on Learning Technologies 4(1), 59–73 (2011)

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

Trausan-Matu, S.: 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 (2010b)

Trausan-Matu, S.: Collaborative and Differential Utterances, Pivotal Moments, and Polyphony. In: Suthers, D., Lund, K., Rosé, C.P., Law, N. (eds.) Productive Multivocality in the Analysis of Group Interactions, vol. 16, p. 15. Springer, New York (in press)

Trausan-Matu, S., Dascalu, M., Dessus, P.: Textual Complexity and Discourse Structure in Computer-Supported Collaborative Learning. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) ITS 2012. LNCS, vol. 7315, pp. 352–357. Springer, Heidelberg (2012)

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

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

Trausan-Matu, S., Rebedea, T., Dragan, A., Alexandru, C.: Visualisation of learners' contributions in chat conversations. In: Fong, J., Wang, F.L. (eds.) Blended Learning, pp. 217–226. Pearson/Prentice Hall, Singapour (2007a)

Trausan-Matu, S., Stahl, G.: Polyphonic inter-animation of voices in chats. In: CSCL2007 Workshop on Chat Analysis in Virtual Math Teams, New Brunwick, NJ, p. 12 (2007)

Weltzer-Ward, L., Baltes, B., Lynn, L.K.: Assessing quality of critical thought in online discussion. Campus-Wide Information Systems 26(3), 168–177 (2009)

Wenger, E.: Communities of practice. Learning, meaning, and identity (Learning in Doing: Social, Cognitive and Computational Perspectives). Cambridge University Press, Cambridge (1999)