JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

Intern. J. Comput.-Support. Collab. Learn.
DOI 10.1007/s11412-014-9190-y

PolyCAFe—automatic support for the polyphonic analysis of CSCL chats

Stefan Trausan-Matu · Mihai Dascalu · Traian Rebedea

Received: 29 April 2013 / Accepted: 21 March 2014

Abstract Chat conversations and other types of online communication environments are widely used within CSCL educational scenarios. However, there is a lack of theoretical and methodological background for the analysis of collaboration. Manual assessing of non-moderated chat discussions is difficult and time-consuming, having as a consequence that learning scenarios have not been widely adopted, neither in formal education nor in informal learning contexts. An analysis method of collaboration and individual participation is needed. Moreover, computer-support tools for the analysis and assessment of these conversations are required. In this paper, we start from the "polyphonic framework" as a theoretical foundation suitable for the analysis of textual and even gestural interactions within collaborative groups. This framework exploits the notions of dialogism, inter-animation and polyphony for assessing interactions between participants. The basics of the polyphonic framework are discussed and a systematic presentation of the polyphonic analysis method is included. Then, we present the *PolyCAFe* system, which provides tools that support the polyphonic analysis of chat conversations and online discussion forums of small groups of learners. Natural Language Processing (NLP) is used in order to identify topics, semantic similarities and links between utterances. The detected links are then used to build a graph of utterances, which forms the central element for the polyphonic analysis and for providing automatic feedback and support to both tutors and learners. Social Network Analysis is used for computing quantitative measures for the interactions between participants. Two evaluation experiments have been undertaken with *PolyCAFe*. Learners find the system useful and efficient. In addition to these advantages, tutors reflecting on the conversation can provide quicker manual feedback.

Keywords Chat conversations · Dialogism · Polyphony · Inter-animation · Learning analytics · Natural Language Processing · Automatic feedback · Collaboration assessment

S. Trausan-Matu (✉) · M. Dascalu · T. Rebedea
Department of Computer Science and Engineering, University "Politehnica" of Bucharest, 313 Splaiul Independetei, Bucharest, Romania
e-mail: stefan.trausan@cs.pub.ro

M. Dascalu
e-mail: mihai.dascalu@cs.pub.ro

T. Rebedea
e-mail: traian.rebedea@cs.pub.ro

Introduction

Instant messaging (text chat) is already used in many collaborative learning sessions (e.g., Stahl 2006, 2009a). One of its distinctive features is the support of online interaction in real time for small groups of students, offering a high potential of inter-animation that can facilitate learning. Any of the participants may enter utterances at the same time, allowing a higher degree of participation than in face-to-face settings. Moreover, the facility of explicitly referencing previous utterances provided by several chat environments (e.g., the VMT environment (Stahl 2009a)), allows and even encourages the existence of more than one discussion thread simultaneously. Simultaneity is not desirable in the case of face-to-face collaboration, where normally only one person should take the floor and speak at a given moment. However, the co-presence of multiple threads in chat sessions is desirable and it should be encouraged because this way a larger number of students may participate in discussions and, meanwhile, an inter-animation process may appear among the different threads of discussion. This is analogous, we shall argue, to what happens in classical polyphonic music or in improvisations during jazz jam sessions.

Although it was shown that the usage of chat sessions for CSCL can be effective for learning (Stahl 2006, 2009a; Rebedea et al. 2010; Dascalu et al. 2011), our experience with such assignments for our university courses showed some shortcomings as well. It is very difficult for a tutor or a professor to read, analyze and assess chat sessions, especially if they are not moderated and if there are a large number of teams. It is particularly difficult to track the threading of the arguments, of ideas, and of the contributions of each student. For example, Trausan-Matu (2010a) reports that the time needed for assessing chat sessions is at least equal to their duration and can extend, on some conversations, to even twice the initial debating time. Because a typical duration of the chat session assignments in this context was around 2 h (and, for example, the duration of chat sessions discussed by Stahl (2006) was even up to 3 h), it is obvious that the tutors' task is extremely time consuming, making almost impossible the detailed assessment of chat sessions in formal learning.

Understanding collaboration and successfully tackling the above-mentioned difficulties requires a model and an analysis method that encompasses all the specific phenomena, including the potentially complex threading and inter-animation of arguments and ideas. Such a model is provided by the analogy with polyphonic music, which considers that any language-mediated interaction is characterized by a weaving of different positions, similar to the counterpoint in a musical polyphony of inter-animating voices (Trausan-Matu et al. 2005, 2010; Trausan-Matu and Stahl 2007; Trausan-Matu and Rebedea 2009; Trausan-Matu 2010c). If possible, computer tools should be developed to support the analysis.

Inter-animation in CSCL chat sessions—as in musical polyphony—is generated by the combination of divergent (generating conflict) and convergent (towards harmony) interactions between participants and discussion threads (Trausan-Matu et al. 2007b). In musical polyphony, the equivalents of inter-animated discussion threads are the voices that inter-animate according to counterpoint rules. The musical metaphor, which is the basis for our polyphonic model, for the polyphonic analysis method, and for the design CSCL scenarios like that presented in the next section is justified also by data from discourse analysis (Tannen 2007), neurology and anthropology (Sacks 2007).

The polyphonic model has already been used for the design and implementation of several systems (Trausan-Matu et al. 2007a; Dascalu et al. 2008, 2010a, b; Trausan-Matu and Rebedea 2010). The *PolyCAFe* system (**Poly**phonic **C**onversation **A**nalysis

JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

and **F**eedback Generation) integrates the facilities offered by the previous solutions and is focused on the analysis of chat sessions. It provides visualization, abstraction and feedback services for supporting researchers and tutors in analyzing chats with or without human moderators, this latter case providing the greatest benefit because the analysis of this type of chats is more difficult. *PolyCAFe* has also been tried experimentally for discussion-forum analysis and also providing feedback to students, as will be presented later in the paper. All of the provided services are packed into web widgets that can be easily integrated into most learning-management systems, personal learning environments or other web applications (e.g., blogs that use Wordpress). *PolyCAFe* uses techniques from Natural Language Processing (Manning and Schütze 1999; Jurafsky and Martin 2009; Dascalu et al. 2010b; Trausan-Matu and Rebedea 2010), Social Network Analysis (SNA) (Dascalu et al. 2010b), and Information Retrieval (Adams and Martell 2008; Manning et al. 2008).

The paper continues with a presentation of the learning scenario and settings. The third section will present the polyphonic model and the associated analysis method, while the following one will present the *PolyCAFe* system. The paper is completed by presenting the results of two evaluation experiments of the system and by conclusions.

The learning scenario and settings

There are many advantages for using chats in contexts that involve collaborative problem solving (Eastman and Swift 2002; Stahl 2009b), engaging in debates, or stimulating the creativity of learners through brainstorming sessions (Trausan-Matu 2010b). However, taking into consideration the difficulty and the required time for providing feedback to students involved in such conversations, especially when they are not moderated (Trausan-Matu 2010a), this scenario may become less appealing to teachers and decision makers in universities and schools.

PolyCAFe has been designed starting from the experience of participating as tutors/professors while using instant messaging (chat) for CSCL in two different settings. The first one is the Virtual Math Teams (VMT) project (Stahl 2009a). In this case, chats were moderated and learners had to work collaboratively on mathematical problems.

The second setting is the usage of non-moderated CSCL chats using the VMT environment for debates related to the competing approaches for a subject presented during lecture hours. Such assignments were given at the Human-Computer Interaction course for undergraduate senior year students, as well as for MSc students studying Adaptive and Collaborative Systems, Natural Language Processing, and Symbolic and Statistical Learning at the Computer Science and Engineering Department of the University Politehnica of Bucharest (UPB). In these courses, students were given between one to three assignments that needed to be solved using a chat conversation in unmoderated small groups of about four participants. A typical assignment given to students was the following:

> You should group in teams of four. Consider that each of you is a director of a company selling a different collaborative technology presented at the course (chat, forum, blog and wiki). Before the chat, you are supposed to individually study collaborative technologies and after that, to have a 1–2 h chat using the VMT environment. In the first part of the chat conversation, each of you have to champion the technology you represent by presenting its features and advantages and criticize the others by invoking other technologies' flaws and drawbacks. In the second part of the chat, you should

discuss how you could integrate all these technologies in a single online collaboration platform.

An excerpt of such a chat is presented below (the second column "ref" contains the number of the referenced utterance using the facility of the VMT environment):

Nr.	Ref	User	Text
76	73	florin	why are blogs search-engine friendly? I think wikis are more search-engine friendly! :P
77	72	bogdan	chats work as long as you have an internet connection....
78	76	elena	because you can use keywords
79	74	florin	never say never—what about Wikipedia ?!
80	78	florin	can you detail on that feature a bit?
81	77	Raluca	in terms of 'coding' your application, I bet yours is the hardest
82	77	Raluca	and only that could take up a lot of time
83	81	bogdan	it may be, but you're only going to do it once....and then use it like that for a very long time
84	80	elena	the blogs can be grouped by interest and all the articles can be full of keywords that are search engine friendly
85		elena	do not forget that a blog is the best way to promote a site
86	84	Raluca	same with forums, and reading the posts looks easier
87	84	florin	ok, thanks for the info
88	85	bogdan	sorry, but the est way is a message to all your friends..using a CHAT
89	88	Raluca	but if you want to study, you have a lot of information on the forums, moreover, if you have questions people can take their time to answer you

We should mention that the polyphonic model, which will be presented in the next section, was also used for designing the way students interact: In the assignment, they are first engaged in a debate where differential positions (dissonant voices) are taken. Afterwards, the previous results were used for collaboratively building a solution (a consonant whole) for the given problem. The VMT environment was used due to its features of allowing explicit references to previous utterances. After the students finished a chat conversation, the tutors read the transcript and graded the students.

The polyphonic model and analysis method

The CSCL community considers that a paradigm shift occurred in the sense that learning can be achieved through social participation in dialogue that constructs discourse, rather than through a transfer of knowledge from teachers or textual documents to students (Bereiter 2002; Stahl 2006; Trausan-Matu et al. 2006). However, even if discourse building is considered essential in collaborative learning, there are very few theories and models of these processes, and even fewer computer applications for supporting its analysis.

Discourse may take different shapes. For example, for the two situations presented in the previous section, discourse is composed by the steps for solving a mathematical problem in the VMT project chats (Stahl 2009a) and, respectively, by the debates that identify attributes, differences and similarities of competing approaches to a given subject. Both cases include threads of discussion inter-animated through moments of conflict (divergences) and of convergence, as in polyphonic music, as will be shown below.

JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

The polyphonic music metaphor for discourse building in CSCL

Several researchers (Koschmann 1999; Trausan-Matu et al. 2005; Stahl 2006) consider that, in addition to the social-cultural ideas of Vygotsky (Vygotsky 1978; Cazden 1993), Bakhtin's dialogism and the musical metaphor of polyphony (Bakhtin 1981, 1984) are appropriate theoretical starting points for CSCL. However, only a few elaborations of a CSCL model based on dialogism and its related concepts in Bakhtin's work (e.g., multivocality, polyphony, chronotope, etc.) have been proposed (Dong 2006; Trausan-Matu et al. 2007a; Ligorio and Ritella 2010). One of them is *the polyphonic model of discourse building in human communication and inter-animation* (Trausan-Matu et al. 2005, 2006; Trausan-Matu and Stahl 2007) used for the design of *PolyCAFe*. This is one of the very few systems that offer learning-analytics tools based on Bakhtin's dialogism ideas, as presented in detail in subsequent sections.

There are at least two reasons for considering a musical metaphor for modeling and analyzing discourse building in CSCL. The first justification is based on the resemblance of the phenomena that appear in successful collaborative chat sessions, characterized by "collaborative moments" (Stahl 2006) and, respectively, in classical musical polyphony or in jazz improvisation. In all these cases a multiplicity of participants start from a given theme (the subject of the CSCL chat session or a musical melodic theme) and act at both individual and small-group levels (Stahl 2006). Meanwhile, they try to achieve *coherence*, a characteristic feature of discourse, both in human communication (Jurafsky and Martin 2009) and in music (Webern 1963) and to be *creative*, to include novelty (assured by diversity, by divergence). In this aim, participants should "inter-animate," they should be aware of the others' utterances and build a new but coherent discourse that integrates their utterances with the others' ones.

A second reason for considering music in analyzing collaborative learning goes beyond just a simple metaphor. The main features of music: repetition and rhythm also have central roles in discourse building within human language, as has been recognized and analyzed by researchers of different domains (Sacks 2007; Tannen 2007). In these cases, music is not only a metaphor, but rather a means that triggers language communication and discourse building. Sacks (2007) provided evidence for the fact that music has an important role in the recovery of human language abilities in the cases of people with brain injuries. Tannen (2007) emphasized that, in correspondence with music, repetition and rhythm are means for assuring involvement.

The musical metaphor is the basis for the polyphonic model and for the analysis method presented below, which proved useful for analyzing CSCL chats. However, we should mention that this approach has the limitations that derive from the differences between music and text. Human language has a higher semantic dimension than music. On the other hand, consonances and dissonances are more obvious in music that in text.

The polyphonic model of discourse building in collaborative learning

The basic idea of our theoretical framework is that discourse building in polyphonic music and in collaborative learning are two analogous cases of the more general phenomenon of human collaboration. As emphasized in the previous section, discourse should include both coherence (harmony) and divergence (conflict), basic characteristics of a polyphonic framework. The polyphonic model that we have introduced starting from Bakhtin's ideas and from the polyphonic music considers that *discourse* in any communicative situation (text, speech, music, gestures) *is structured in inter-animating voices that follow the principles of counterpoint rules from polyphonic music, which assure both coherence and diversity.*

Counterpoint, which rules musical polyphony is seen by Bakhtin (1984) as a special case of the more general concept of *dialogic relationships among utterances* that are: "a much broader

phenomenon than mere rejoinders in a dialogue, laid out compositionally in the text; they are an almost universal phenomenon, permeating all human speech and all relationships and manifestations of human life—in general, everything that has meaning and significance" (Bakhtin 1984). In this vision everything (we may say any discourse, textual or musical) is a dialog: Novels, essays or even words should be analyzed from a dialogical point of view, in which *utterances* should be the unit of analysis. *Utterances* may range from a simple word, a gesture, an idea, to a reply in a conversation or even a whole book.

Another main idea of Bakhtin's dialogism is that an indefinite number of *voices* are present as echoes in any utterance, even in each word, re-voicing previous utterances (seen as voices): "Utterances are not indifferent to one another, and are not self-sufficient; they are aware of and mutually reflect one another. These mutual reflections determine their character. Each utterance is filled with echoes and reverberations of other utterances to which it is related by the communality of the sphere of speech communication." (Bakhtin 1986).

Regarding the re-voicing process, Stahl emphasized:

> There is a sense of "voice" in Bakhtin's literary analyses where the voice of one character in a text is "re-voiced" by another character. Indirect speech, in which the speaker quotes another person's earlier utterance, is the clearest example. Less explicit forms involve the speaker adopting the tone, vocabulary or concerns of a previous speaker. If this paper referred here to Bakhtin's discussion of Dostoyevsky's lead character quoting his landlady telling a story of several people talking, this paper would be engaging in a multivocal polyphony of voices. The paper's text would be re-voicing the cacophony of voices of Bakhtin, Dostoyevsky, Raskolnikov, the landlady and the people re-voiced by the landlady. If one listens closely to language—whether spoken, written in a novel or typed in an online chat—one can hear potentially many voices interacting in a given utterance coming ostensibly from one person. (Gerry Stahl, personal communication, December 19, 2013).

Starting from the polyphonic music metaphor and Bakhtin's aforementioned ideas, in our model of knowledge building the concept of voice is considered similar to a melodic line initiated by one or more utterances and continued by re-voicings. Multiple voices coexist and they inter-animate as a result of entering in dialogues that imply dissonances and consonances, similarly to Johann Sebastian Bach's fugues in four voices (Trausan-Matu 2010c). In this way intersubjective meaning making is achieved (a process similar to small-group jazz improvisation), as we shall exemplify later. Utterances act like a gear between the personal and the social knowledge-building levels of Stahl's model (2006 p. 203). Therefore, our concept of "voice" should not be thought of as the particular features of the speech sounds produced by a given person, but rather, as in polyphonic music, where an organist in a Bach fugue can concomitantly play several voices or a group of instruments in an orchestra can play a single voice in parallel with other groups playing other voices. Similarly, a student in a CSCL chat may express more than one voice, which means that his/her utterances contain not only his/her voice, but also echoes of other voices, according to the re-voicing phenomenon mentioned above.

In our polyphonic model we define a voice as a distinct dialogical position (for example, a proposal, a hypothesis, an opinion, an idea, an approval, a rebuttal, etc.) manifested by emitting one or a recurrent series of utterances that influence the conversation by having (re-voiced) echoes of other utterances (Trausan-Matu 2010c). A voice is not always associated with a single participant. There may be group voices emitting collective utterances (several participants emitting the same utterance in the same time). Another case in which several participants contribute to a voice is that of successive re-voicings and debates, as described below.

Any utterance may contain multiple, alien voices (Bakhtin 1986) through the echoing/re-voicing process, in addition to the voice of the person that uttered it. The distinct position

JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

(voice) of the utterer is influenced by the inherent particular features like personal knowing, tacit pre-understanding (Stahl 2006), intentions, personality, etc.

A voice initially associated to an utterance may be echoed (re-voiced) by different participants in long threads of utterances. Eventually, after a number of repetitions and debates, the association of that voice to its initial emitter may be forgotten and even its initial utterer may now speak with another voice(s), which may enter in dissonance with the other(s). We may say that the initial voice, by re-voicing and successive debates may be transformed, it cannot be associated to a single person and can be considered as a new voice of its own, which is collectively built. This phenomenon of dialogical transformation is a manifestation of what Stahl calls "building collaborative knowing" (Stahl 2006). As he now puts it, a term or thought articulated by an individual under specific circumstances can become reified, sedimented and generalized through repetition into a persistent idea that is decontextualized and may even be institutionalized (see Stahl 2013, Chapter 8, esp. Figure 8.6).

An example of an utterance whose thread of echoes/re-voices become a voice of its own may be, for example, a proposal emitted by a participant in a CSCL chat session, debated by other participants' (voices), and eventually leading to a solution of a problem. Note that the emitter's voice later may conflict with her initial voice.

As exemplified elsewhere (Trausan-Matu 2012), a repeated sequence of a word or a phrase may transform it into an artifact used by learners in solving a problem in a chat CSCL session. In our vision, these artifacts may be seen as voices.

An example of a collective utterance behaving like a voice was identified in a face-to-face collaborative-learning session of a class of 6th grade Japanese students. Towards the end of the class, in response to a question from the teacher, a group of students moved their gaze down (even if they had been active before). This non-verbal group utterance, this collective gesture uttering that they will not answer (Trausan-Matu 2013), may be seen as a collective voice (like a group of instruments in a symphony) which may influence a teacher's future utterances.

In Fig. 1, for example, many voices may be identified containing a transcript of a chat session fragment from the assignments (at UPB, see the second section), where students debated what features to consider for integrating collaborative technologies. From our definition, any utterance may potentially be a voice. The determination of the voice is made by the existence of threads of re-voicings. Two types of such threadings need to be investigated. The first one consists of threads of links to previous utterances explicitly indicated by participants through the facility of the VMT chat environment (these links are represented with curved arrows in Fig. 1). The second type contains threads of implicit links among utterances. This type includes the repetitions of some words (represented with straight lines in Fig. 1: the voices 'topic', 'reply', 'presentation', etc.), adjacency pairs, justification links, co-references or other discourse links (Jurafsky and Martin 2009).

To see an example of how word repetition can be understood as re-voicing, consider the following response pair from lines 17 and 19 in Fig. 1:

Tim: You discussed about a topic separation

John: yes. because we did not like the way the topics were present in concert chat

Tim's utterance is in effect an indirect speech act, which could be restated:

Tim: John said, "topic separation"

In this alternative format, John's voice is explicitly being re-voiced by Tim. This shows how Tim's posted utterance implicitly references a previous utterance by John. It does so to keep that voicing alive or present and to elicit an elaboration by John. Tim's posting is not some kind of "externalization" of an "idea" in Tim's head so much as an interactive action to

Nr	Ref	Time	User	Text
17		10.26.25	tim	You discussed about a topic separation
18	15	10.26.37	adrian	First of all, the reply method is cumbersome
19	17	10.26.50	john	yes.. because we did not like the way the topics were presented in concert chat
20	18	10.26.56	john	yes !!
21	20	10.27.04	john	i hate double-clicking!
22	20	10.27.18	tim	and how can we find topics ?
23	18	10.27.26	adrian	What bothers me is the linear presentation of the discussin
24	23	10.27.43	john	Yep
25	18	10.27.46	adrian	and double-clicking too
26		10.27.54	tim	You mean u want somethig like a chat forum ? :D
27	24	10.27.58	john	and the reply-to facility is supposed to help you
28	18	10.28.15	adrian	i'd like a tree presentation more
29	18	10.28.38	adrian	or maybe multiple chat columns, for each chat sub-thread
30	27	10.28.58	john	but it is really difficult to use in real time, because there are so many topics discussed which intertwine each other
31	28	10.29.18	john	i subscribe to a tree-like presentation form
32	P 30	10.29.20	adrian	yes, that's why a clear separation of topics is needed
33	31	10.29.47	adrian	this is easy to implement, no problem here :)
34	30	10.29.49	tim	You need also a clever visual representation
35	30	10.30.05	tim	you'll need also a clever visual interface
36		10.30.22	tim	Who decides the topics ?

Fig. 1 Threads of re-voicings (Trausan-Matu et al. 2007b; Trausan-Matu and Rebedea 2009)

elicit a response from John by building on John's previous post and keeping it's shared meaning-making process active. In the second line of the pair, John responds to that elicitation by elaborating in his own voice the discussion of "topics." First, he says, "Yes" to mark linguistically that he is responding to Tim's utterance. Also, he uses a referencing feature of the chat technology (VMT, aka Concert Chat) to point graphically from his new posting to the posting by Tim to which he is responding in a chat thread. However, in addition, he repeats the term "topic" to re-voice Tim's re-voicing of his own earlier comment, and thereby to continue that voice as an on-going mini-discussion. Finally, semantically, the two utterances form an adjacency pair, through which intersubjective meaning making takes place. For humans, the thread of Tim's voice about topics is so over-determined that a human speaker can follow it implicitly in reading the chat log. For automated analysis, however, we needed to operationalize some of these indicators of paths followed by dialogic voices. Searching for repetitions of works like "topics" is one heuristic that proved useful.[1]

If we consider a voice, as we defined it, as a distinct position manifested in one or a series of utterances that have echoes, the set of utterances emitted by each participant might be used as a basis for detecting common features reflecting his/her particular features and positioning: personal knowing, tacit pre-understanding (Stahl 2006), intentions, personality, etc. This perspective may determine a kind of a "generic voice" of a participant, which may be compared with those of the other participants and even of collectively generated voices. It also allows determining the degree of participation in a discussion, as used in PolyCAFe and presented in a further section.

Voices weave in a polyphonic way in a discourse, keeping their individuality and meanwhile creating a more complex whole: "The essence of polyphony lies precisely in the fact that the voices remain independent and, as such, are combined in a unity of a higher order than in homophony" (Bakhtin 1984). An essential fact is that, like in the polyphonic musical case, several voices co-exist in any moment of time, dialogical relations appear among them, and they inter-animate through dissonances and consonances. The polyphonic texture assures the most important features that characterize discourse: coherence and diversity (novelty). The

[1] The analysis of this response pair is due to Gerry Stahl (Personal communication, December 19, 2013).

JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

transversal co-presence of multiple voices (threads of re-voicings) at the same time inherently gives birth to both consonances and dissonances, which nevertheless, in successful situations, *tend to weave along the longitudinal time dimension* towards coherence, while dissonances create novelty, which assure diversity or induce solutions in problem solving, eventually driving intersubjective meaning making. The phenomenon appears in CSCL and is similar to the polyphonic classical musical case or to jazz improvisation (Trausan-Matu et al. 2006): "The deconstructivist attack […]—according to which only the difference between difference and unity as an emphatic difference (and not as a return to unity) can act as the basis of a differential theory (which dialectic merely claims to be)—is the methodical point of departure for the distinction between polyphony and non-polyphony" (Mahnkopf 2002, p. 39).

The polyphonic model was used for the analysis of chat conversations (Trausan-Matu et al. 2007b), of face-to-face learning sessions including also non-verbal acts (Trausan-Matu 2013) and of socially built discourse. An example of three instances of inter-animation patterns for the excerpt from Fig. 1 is illustrated in Fig. 2: The first (between voices 'reply' and 'topics' at utterances 27 and 30) and third (voices 'topics' and 'presentation' at 30–34) inter-animation patterns are divergent (and the cue phrases are 'but' and respectively the comparative 'clever') and the second (voices 'presentation' and 'topics' at 37–38) is convergent (the cue phrase being 'also').

The polyphonic analysis method

Starting from the polyphonic model of collaborative learning, a qualitative polyphonic analysis method was developed (Trausan-Matu et al. 2005, 2007b) and used for several purposes: investigating collaborative knowledge construction and learners' participation in collaborative chat sessions (Trausan-Matu et al. 2007b; Trausan-Matu 2013); identification of the artifacts that enable problem solving (Trausan-Matu 2012); and identification of important (pivotal) moments in conversations (Trausan-Matu 2013). This method has been used in analyzing chat sessions performed at VMT and UPB, and also face-to-face collaboration in a Japanese class (Trausan-Matu 2013). However, a first systematic presentation of the method is described below.

In our analysis method, we follow several steps towards identifying the polyphonic structure of conversations. That means detecting voices along the longitudinal dimension and their transversal inter-animations. First of all, utterances are delimited. As mentioned above, utterances may range from a word or a phrase, to a reply in a conversation, a post in a forum, a sentence or a paragraph in a text, a whole essay or novel, or even gestures and other non-verbal acts. We may even have utterances that are included in other utterances. For the case of conversations (online or face-to-face) and discussion forums, obvious utterances are units of interaction marked as such by participants, for example the text between two carriage-returns in instant messaging.

In the second step, links are identified between an utterance and a previous one, which is considered as a precursor. Links may be those explicitly indicated by the participants or those implicit, which can be detected with Natural Language Processing (NLP) techniques:

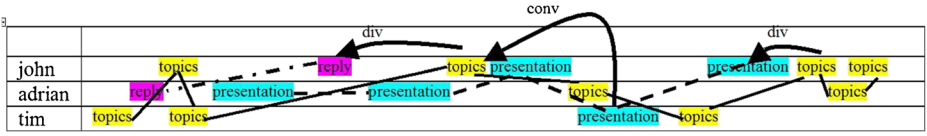


Fig. 2 Divergent and convergent inter-animation patterns (Trausan-Matu and Rebedea 2009)

repetitions of words and phrases, adjacency pairs, justification links, co-references, etc. (Jurafsky and Martin 2009). After this step a *graph of utterances* may be constructed, containing utterances as nodes and links as arcs, which may be labeled with their type (explicit, implicit, repetition, justification, adjacency pair, etc.). This graph may be seen as containing potential re-voicings. The utterance graph may be also used for computing quantitative values, as will be shown in the next sections.

In the third step of the analysis, we focus on voices. For this purpose, threads of re-voicings (echoes), which indicate voices (for example, 'topic', 'presentation', 'tree' in Fig. 1), should be detected starting from repeated important words (the most frequent words obtained after eliminating the so-called stop words that do not bring meaning, like "a", "and", etc.). Links detected at the previous step may also be indicators of threads. For example, a thread starting from an utterance introducing an idea and containing adjacency pairs, argumentation links, etc. may signal the presence of a voice influencing the conversation.

As mentioned in the previous section, the set of all utterances of each participant may be considered to detect generic features of the generated voices. Moreover, even if the teacher is not participating in the session, the presence of his/her implicit voice(s) should be taken into account. For example, in an assignment for a chat session (at UPB, this is the case of the collaborative technologies—forum, chat, wiki and blog—see the second section of this paper) the topics to be discussed are important voices, which were uttered by the teacher and debated by the students.

A fourth step is dedicated to the identification of inter-animation patterns among voices, starting from utterances or pairs of utterances where voices intersect. As discussed in detail elsewhere (Trausan-Matu et al. 2007b), inter-animation patterns may be classified as convergent and divergent, similarly with consonances and dissonances in polyphonic music. The identification of these patterns is facilitated by the presence of cue phrases like 'but', 'nevertheless', 'different', 'same', 'also', 'other', etc. (see the discussion at the end of the previous section).

The fifth step concludes the analysis using the graph of utterances, the detected voices and their inter-animation patterns for analyzing different aspects of discourse building: meaning making, identification of artifacts in problem solving, investigating pivotal moments, rhythm, collaboration regions, assessing learners' participation and the collaboration of the team as a whole.

As seen from the implementation hints inserted in the presentation of the steps above, the operationalization of the polyphonic method is done by using heuristics and Natural Language Processing techniques (Jurafsky and Martin 2009): utterances delimitation (step 1), identification of repeated words, speech acts, adjacency pairs, and co-references; Latent Semantic Analysis for detecting semantic similarities (or, the inverse, semantic distances) among utterances (steps 2 and 3), cue phrases identification (step 4). Consequently, automated tools may be used for assisting in several moments of analysis. However, the detection of all re-voicings and inter-animation is very difficult (if not impossible in general). What it is possible is to provide graphical facilities for a human to investigate re-voicings, potential voices and their inter-animation starting from different threads in the graph of utterances. Quantitative measures may also be computed based on the utterance graph and semantic distances among utterances, as will be detailed in the sections dedicated to the *PolyCAFe* system.

Nevertheless, these tools have a limited power, in fact the detection of voices and their inter-animations being a very challenging task even for a human expert as multiple discussion threads co-occur and overlap throughout the conversation, making it even more difficult to keep focus on each participant's points of view. These constitute the most difficult part of the polyphonic analysis method and an important limitation of automated tools.

JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

The polyphonic analysis method may be compared with Conversation Analysis, which is already used for analyzing CSCL, as described by Zemel et al. (2009). They consider that an important issue is that interleaved coherent long sequences may be detected in CSCL chats. However, this kind of sequence are rather different from our voices and the authors do not discuss of any details related to the interactions between them, as we do in the polyphonic framework.

Another term of comparison for our method is the approach of Suthers and Desiato (2012). They start from a model and method for analyzing interactions in CSCL logs based on sociograms, contingency and uptake graphs, which similarly to our case are the basis for developing analysis computer tools. Contingencies, the basis for uptakes, are similar to implicit links and echoes. The difference between the two theoretical frameworks is that, while Suthers' approach is centered on the uptakes, in the polyphonic framework the focus is on threads of utterances, voices and their inter-animations, which may not be uptakes, as is the case of collaborative utterances (Trausan-Matu et al. 2007b).

The PolyCAFe system

In recent years several CSCL applications were developed for analyzing interactions in conversations using transcriptions of spoken conversations, chat logs, forum discussion threads and wikis. Such examples are *CORDTRA* (Hmelo-Silver et al. 2006), *COALA* (Dowell and Gladisch 2007; Dowell et al. 2009), *DIGALO* and other tools used in the *Argunaut* system (Harrer et al. 2007), *ColAT* (Avouris et al. 2007), the Scaffold-Argument visualization (Law et al. 2008), *KSV* (Teplovs 2008), *VMT-Basilica* (Kumar et al. 2009) and the system of Suthers and Desiato (2012). However, no system really provides complex analysis and feedback facilities for chat and forum discussions in terms of discourse structuring, participant involvement and collaboration assessment.

There are at least two factors that provide insight into this situation. The first factor is that, even if dialogism (Bakhtin 1984) is considered a well suited theoretical model for tackling the complexity of CSCL (Koschmann 1999; Stahl 2006), extremely few software implementations started from it due to its complexity. The second factor is related to the fact that the majority of collaboration acts in conversation-based CSCL are based on the exchange of textual (spoken or written) messages. Thus, another problem arises because current NLP systems are far from providing reliable text understanding capabilities. Moreover, in CSCL chats and forums there are usually more than two interlocutors, a case generally ignored in most NLP theories for dialogues, most of them developed for conversation analysis (Trausan-Matu and Rebedea 2010). Nevertheless, even the two interlocutors' situation is far from being tackled satisfactory in non-trivial dialogs.

PolyCAFe and its precursor *Polyphony* (Trausan-Matu et al. 2007a) are probably the first systems that are designed and implemented starting from Bakhtin's ideas on dialogism, with emphasis on polyphony and inter-animation (considering the counterpoint analogy to music: longitudinal voices that interact transversally, as mentioned in a previous section). The *PolyCAFe* system was designed, implemented and validated within the LTfLL—Language Technologies for Lifelong Learning project (Trausan-Matu et al. 2008, 2009; Trausan-Matu and Rebedea 2010; Rebedea et al. 2010) (see http://www.ltfll-project.org/) funded by the European Commission under the 7th Framework Programme (Berlanga et al. 2009). The online version of the system can be accessed at http://ltfll-lin.code.ro/ltfll/wp5/index.php.

PolyCAFe was developed in order to follow two main aims. First, as its precursor, it was designed to offer computer support for a teacher or a researcher in CSCL when applying the

polyphonic method in chat analysis. There are several ways in which the polyphonic model and method of analysis are used in the design and implementation of *PolyCAFe*. Multiple voices, their echoes and their interactions are considered to be present in a chat, a principal goal of the system consisting in their identification. As mentioned earlier, the concept of "voice" has an extended range (Trausan-Matu et al. 2007b; Trausan-Matu and Rebedea 2009) and in this first aim, *PolyCAFe* helps to identify implicit links, voices, re-voicings, and inter-animation patterns using NLP techniques, as will be shown in the next sections.

Second, in addition to supporting the qualitative polyphonic analysis, quantitative measurements were designed and implemented derived from the polyphonic model. Both teachers and learners may use these integrated tools developed in the general framework of the LTfLL project.

The main beneficiaries of *PolyCAFe* are tutors and researchers. However, experiments were made for analyzing also how learners would accept this system, in addition to tutors. The results will be presented in a separate section of this paper.

Widgets overview

In order to empower researchers, tutors and learners with extensive control over the facilities of the system, *PolyCAFe* was implemented as an online platform that displays results in web widgets that can be used independently or together, in different combinations (Dascalu et al. 2010a, 2011). In this manner, the processing is decoupled from the interface and the widgets can be easily integrated into most online learning environments and other web platforms. Any number of widgets may be displayed together on the screen, even more instances of the same type of widget. This facility allows users to see in the same time different perspectives of the collaboration process.

PolyCAFe provides two management and five feedback widgets. The *management widgets* enable tutors to define, edit and delete assignments and, respectively, to create, read, update, and delete conversations (chats or discussion threads from online forums). In this section we will focus on the most representative feedback widgets, which support the underlying polyphonic model: the *conversation visualization widget*, the *conversation feedback widget*, the *utterance feedback widget*, and the *search widget*. In addition to these, there is also a *participant feedback widget* that is not presented here.

While the other feedback widgets use the polyphonic model and analysis method for quantitative analysis, the *conversation visualization widget* is the principal facility that may be used for the qualitative polyphonic analysis method of a conversation. It directly supports the first two steps of the polyphonic method of analysis through the visualization of utterances and links (explicit and implicit) between them. It also helps users in discovering re-voicings by identifying sequences of explicit or implicit links (using the zoom and "Conversation thread" facilities—see below) or of repeated words (using the "Special threads" tab—see below). Inter-animation patterns may also be discovered by combining the visualization of voices generated by important repeated words and discourse markers specific to inter-animation patterns like "different", "same", "but", "nevertheless", etc.

The conversation visualization widget displays a diagram of the utterances and the connecting links (the "graph of utterances"). Utterances are represented as small rectangles (whose length is proportional to their number of characters) aligned to the right of each participant's name, following the conversation timeline (the scale below indicate utterances' numbers; alternatively, the "scale to time" option allows a time-based scaling). The links between them are differently colored. For example, in Fig. 3 explicitly mentioned links through the referencing facility of the VMT or ConcertChat environments (Holmer et al. 2006) are marked as red, whereas the implicit links detected with NLP techniques are green.

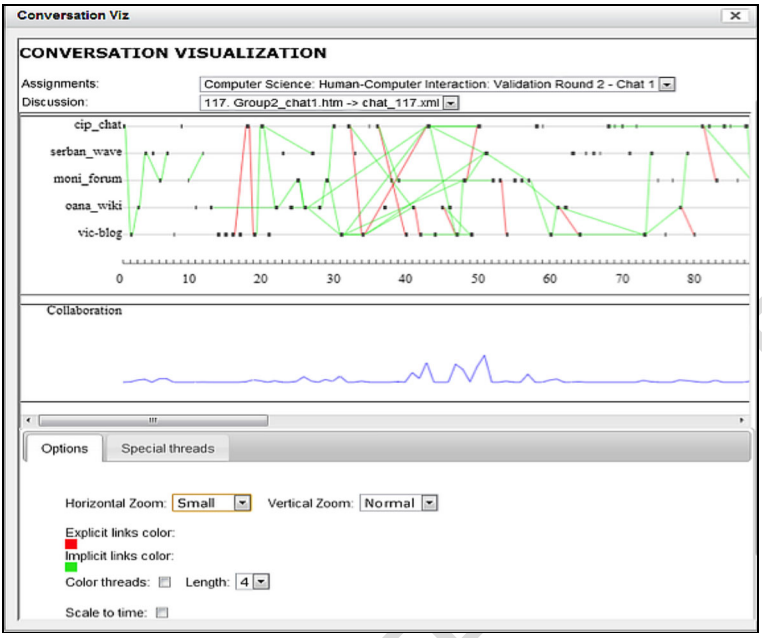


Fig. 3 Conversation visualization widget. The graph follows the conversation timeline (the Y axis consists of utterance IDs) split among chat participants

The "Special threads" tab offers the possibility of the identification of re-voicings generated by repeated words. For example, Fig. 4a displays the evolution of the concepts (topics, reply, presentation) from Figs. 1 and 2, analyzed in a previous section (it should be mentioned that the discussion about topic separation, analyzed after Fig. 1, started in a precedent session, therefore the thread associated to "topic" starts in Fig. 4a with Tim's utterance).

By considering discourse markers like "but", "nevertheless", "also", "different", "same", etc., inter-animation patterns may be discovered. As an example, in Fig. 4b the co-presence of

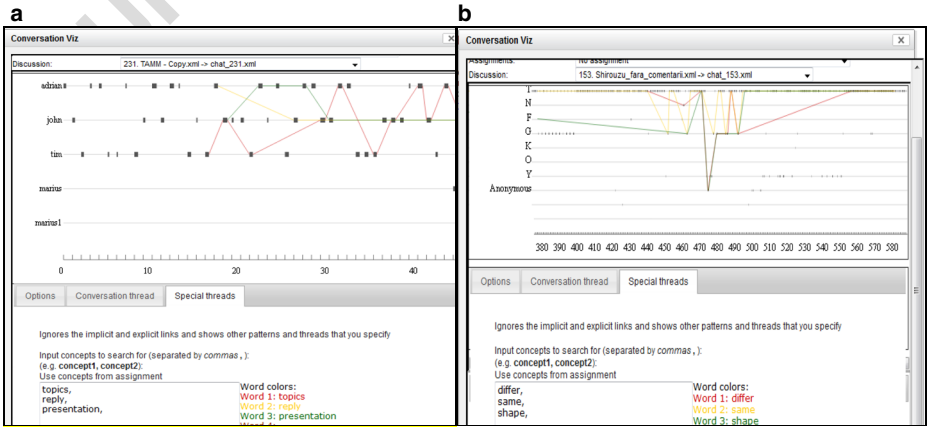


Fig. 4 **a** The three voices associated to the concepts (topics, reply, presentation); **b** Distribution of discourse markers supporting the identification of inter-animation patterns

the discourse markers "same" and "differ" with the voice related to the word "shape" indicate potential inter-animation patterns that marked the discussion between a teacher ("T") and several students (the other letters).

The Conversation visualization widget offers also quantitative data. A graphical representation indicating an estimation of the degree of collaborative discourse is presented in the same interface as a graphics below the graph of utterances (see Fig. 3), concomitantly following the conversation timeline. A detailed presentation of how the collaboration degree is actually computed is included in a subsequent section.

The *conversation feedback widget* presents general information and statistics about the entire conversation: the most relevant concepts from the conversation, a suggestion of concepts that are semantically similar (from the Latent Semantic Analysis—LSA (Landauer and Dumais 1997; Landauer et al. 1998) semantic space) to the ones discussed in the chat and statistics regarding the density of the graph of utterances or the percent of several types of dialog acts, such as personal opinions, request for information and arguments (see Fig. 5). These data are important for the polyphonic method of analysis because the most relevant concepts are usually those re-voiced and thus they provide candidates for voices, which may be investigated and visualized with the Conversation visualization widget.

The qualitative estimators (e.g., "GOOD") are determined through predefined thresholds that were imposed after analyzing a corpus of about 100 conversations and determining variance intervals computed using the distribution of scores for each assessment factor. This is done in a similar manner with the grading performed in some schools and countries that use percentage intervals for assigning grades (scores).

The *utterance feedback widget* (Fig. 6) gives indicators for each post in the conversation: speech acts and argumentation patterns that were detected in each utterance. A numerical value is included, computed according to the utterance evaluation process, later described in extent. Moreover, this widget also presents the users a summary of the conversation that includes only the most important utterances in the discussion (the ones marked with a star icon in Fig. 6)

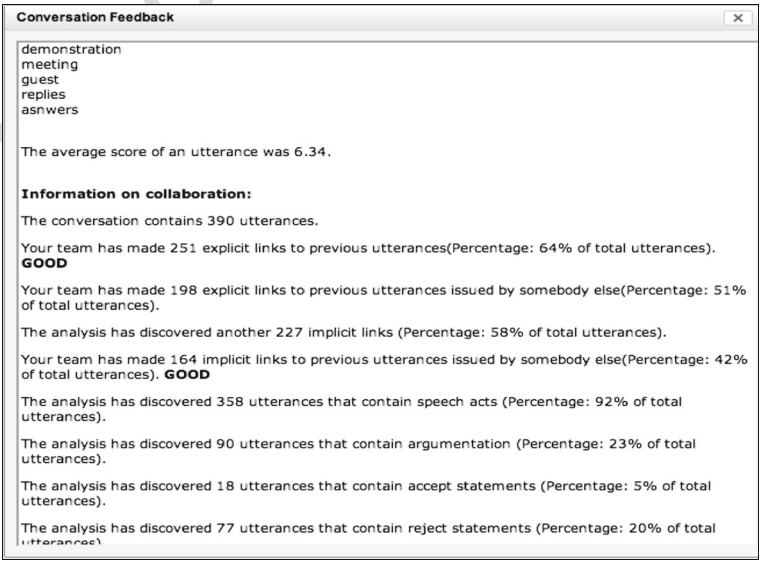


Fig. 5 Conversation feedback widget

JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

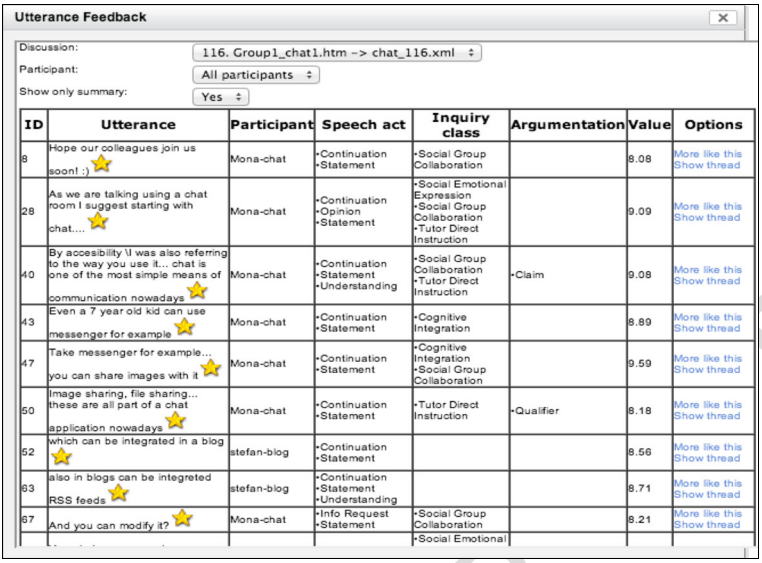


Fig. 6 Utterance feedback widget

with regards to the numerical values computed for both content and collaborative discourse (presented in a subsequent section).

The *search conversation widget* provides a mechanism for ranking utterances and participants with regards to a search query provided by the user. The search engine takes into consideration not just the lexical items, but also the semantic relatedness scores using WordNet (http://www.wordnet.edu) and LSA, and the importance of each utterance as considered by the utterance evaluation process, described in a subsequent section (see Fig. 7).

Different visualizations are useful for analyzing inter-dependencies that can be observed between data in different widgets. On one hand, we can easily observe from the *conversation visualization* widget the distribution per concept, or more specifically the voice's evolution, spanning throughout the discourse. On the other hand, the *search widget* provides the results of the search for "blog" ordered by participant and by the most important interventions related to this voice.

Architecture and core functionalities

PolyCAFe integrates a series of processing modules incorporating Natural Language Processing techniques, Social Networks Analysis (SNA) and polyphonic analysis (Trausan-Matu et al. 2009; Rebedea et al. 2010, 2011; Trausan-Matu and Rebedea 2010). The raw data is a chat conversation encoded as an XML file. According to the polyphonic analysis method, the first processing step is the detection of utterances. Although the borders of an utterance may vary greatly from delimiting a simple word or interjection to a set of intertwined utterances or even to an entire novel (Bakhtin 1986), the *PolyCAFe* system implementation separates utterances in chats based on the end-of-message (carriage-return). Therefore, there may be successive utterances of the same participant.

As seen in the previous section, the outputs take the form of graphical visualizations and feedback on several distinct levels: for each utterance in the conversation, individually for each participant and globally for the conversation as a whole.

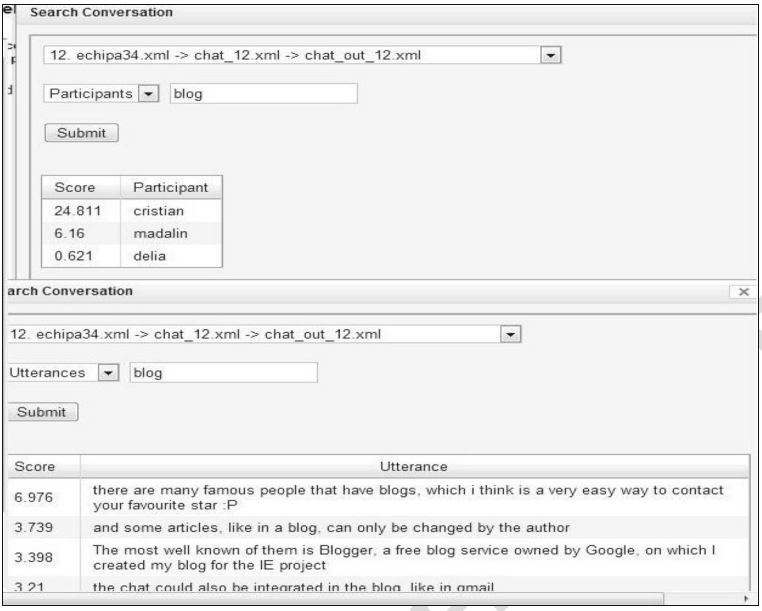


Fig. 7 Semantic search—relevance scoring and ordering of: **a** Participants and **b** Utterances, in contrast to an initial distribution between participants of the "blog" voice

The modules of the *PolyCAFe* system can be grouped corresponding to four major processing steps, as depicted in Fig. 8 with different colors. Only the SNA module is included in two groups, being applied on two types of graphs: the graph containing participants as nodes and exchanged utterances (the explicit and implicit links) as arcs, and the graph having utterances as nodes and the same links as arcs (the graph of utterances). The first is used to determine participants' involvements and the second for the polyphonic analysis and utterance evaluation.

The *first group* of modules contains underlying tools and resources for basic NLP processing. The first step consists in a typical series of basic language processing (a "NLP processing

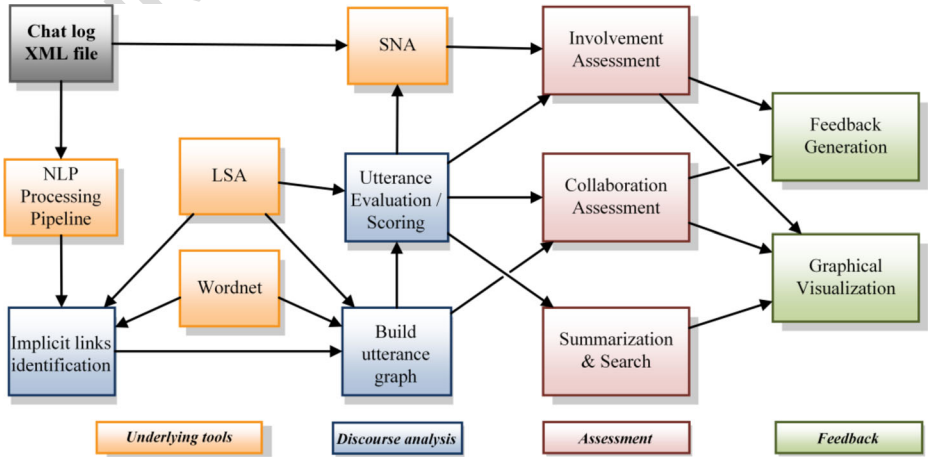


Fig. 8 *PolyCAFe* Technical architecture

JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

pipeline"): tokenization, spelling correction, stemming, part of speech tagging and parsing (Manning and Schütze 1999). The *WordNet* lexical database and LSA spaces compose the semantic resources modules used mainly for concept extraction, which may be candidates for voices. They form the basis for a semantic evaluation of the participants' involvement and evolution.

The *second group* contains advanced NLP and discourse analysis modules for the automatic identification of underlying interactions among participants (Dascalu et al. 2010b; Trausan-Matu and Rebedea 2010). To this aim, speech acts, lexical chains, adjacency pairs, co-references and semantic similarities (Manning and Schütze 1999) are identified. All these are the starting points for detecting candidates of implicit links (the second step of the polyphonic analysis method) that constitute the arcs in the graph of utterances, in addition to the explicit links indicated by participants as references to previous utterances in the chat environment room (Holmer et al. 2006; Stahl 2009a).

The graph of utterances is essential for the last three steps of the polyphonic analysis method. Meanwhile it plays a central role in the scoring process of each utterance and of each participant (Dascalu et al. 2010a). In addition, starting from it, conversation clusters of interlinked utterances are identified using specific graph algorithms. A simple approach consists of identifying connected components in the graph of utterances, while other methods employing the use of the flow graph could also be applied (Cormen et al. 2009).

The modules in the *third group* process the graph of utterances with the associated scores for inferring metrics for collaboration and individual involvement. In addition, extractive summarization and semantic search make also use of previous components and are provided as additional features of *PolyCAFe*

An estimation of the degree of collaboration (or collaborative discourse) starts from the analysis of the graph of utterances with Social Network Analysis and LSA techniques in connection with the polyphonic analysis method. SNA specific metrics are computed on the graph of utterances for identifying also the most central utterances within each discussion cluster (Dascalu et al. 2010a, 2011). Participant involvement is evaluated through the *interaction graph* in which participants are the nodes, edges are the inter-changed utterances and the weights of the edges are determined as the sum of scores of the utterance multiplied by their similarities. LSA is employed to measure the semantic similarity between interventions or the strength of explicit or implicit links, as co-occurring thematic concepts induce a high textual cohesion between utterances.

Moreover, the individual involvement of participants derived from SNA applied on the previous interaction graph is tightly connected to collaboration assessment that considers the information transfer between different interlocutors, enabling in the end a deeper representation of the conversation's social-cognitive dimension. In terms of the proposed dialogic model, an analogy was proposed between the computation model of assessing collaboration and the intertwining of voices from Bakhtin's theory (Trausan-Matu et al. 2005). The estimation of the involvement of participants may be viewed as the cumulated impact of the set of all his/her utterances, as mentioned in a previous section.

The final step in the analysis, consisting of the modules from the *fourth group*, aggregates all the factors obtained as outputs from the previous modules and displays them in an intuitive manner within the user interface, in order to offer textual and graphical feedback overall or for each participant, on several levels.

Utterance evaluation

From the perspective of the polyphonic model, each utterance contains alien voices (echoes of other utterances) as mentioned in a previous section. In this idea, if we would like to assign a

score to an utterance, a starting point is the number of echoes (re-voicings) it contains and it generates, determined from the graph of utterances whose arcs get weights corresponding to a LSA-based semantic similarity function between the utterances (Dascalu et al. 2010a, 2011).

The actual scoring process of each utterance has three distinctive components: a surface one, a semantic and a social one (Dascalu et al. 2010a, 2011). In order to provide a clearer image of the previous metrics, Fig. 9 presents a slice of a conversation that could also represent a partial discussion thread centered on the utterance that is under analysis, with the demarcation of possible links (explicit or implicit) from the graph of utterances and with the presentation of the considered analysis factors.

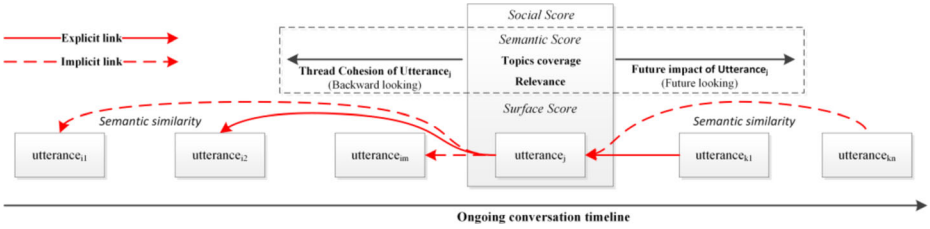
The first, "surface" score is inspired from basic textual complexity measures (Nelson et al. 2012). First of all, a typical NLP preprocessing is done by the elimination of the so-called stop words (that do not carry content, for example, "a", "the", "to", etc.), spellchecking and stemming (extracting the root of the words, by eliminating suffixes like "ed", "ing", "ly", etc.). In order to keep the inputs of the system as clean as possible, only dictionary words are considered. Moreover, as it is a common practice to use abbreviations in CSCL conversations, a list of translations is used to expand the shortened versions encountered in the discussion. The first score considers the length in characters of the remaining words.

The *semantic dimension* is the most important component in the utterance evaluation process and it is a combination of four different components: *thread cohesion*, *future impact*, *relevance* and *topics coverage*. It involves the usage of LSA applied on the content of the utterances and taking into account the graph of utterances.

Thread cohesion of a given utterance is the percentage of links (explicit and implicit) to previous utterances that share a semantic similarity above a given threshold with that specific utterance. Thus, thread cohesion is a backward-looking mechanism used for assessing the importance of an utterance within the ongoing discussion threads it is part of. Thread cohesion is an important assessment factor as any utterance should build on previous ones in the same discussion threads.

Future impact enriches thread cohesion by quantifying the actual impact of the current utterance within future inter-linked utterances from all discussion threads that include the specified utterance. In terms of the polyphonic model based on Bakhtin's dialogism (Bakhtin 1981, 1984), future impact resembles echo as it measures the information transfer from the current utterance to all future ones (explicitly or implicitly linked) by summing up all similarities above the previously defined threshold. From this point of view, an utterance is more important if it has strong echoes in all the future ones from the discussion threads it is part of. The summation takes into account the strength of all these echoes.

The *relevance* of an utterance about the overall discussion can be approximated by computing the LSA semantic similarity between the current utterance and the entire conversation. To this extent, relevance is the simplest supplement for the surface score as it adds more



Q3 **Fig. 9** Slice of the graph of utterances emphasizing the utterance analysis factors

JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

importance to utterances that are similar with what has been discussed most and decreases the importance of lengthy utterances that are not in the scope of the overall conversation.

Because each discussion has a predefined set of topics that had to be followed and which should represent the focus concepts of the chat, *topics coverage* measures the degree in which the participants used these keywords in their interventions. In our implementation, topics coverage is obtained by evaluating the similarity between each utterance and the specific set of keywords specified by the tutor or teacher as important topics of the discussions. Semantic distances and cosine similarity within the LSA vector space are used for this task. In other scenarios or for other tasks, the initial topics can be computed automatically from a given corpus of documents that should be read by the students, before participating in the discussion.

The social dimension implies an evaluation from the perspective of social network analysis performed on the graph of utterances. In the current implementation only two measures from graph theory (Cormen et al. 2009) are used (in-degree and out-degree), but other metrics specific to SNA (Freeman 1977; Brandes 2001; Newman 2010) and minimal cuts (Cormen et al. 2009) will be considered.

Collaboration assessment

Knowledge may be built in two different manners, each effecting the individual: *personal knowledge building* (building personal knowing, see Stahl (2006)) when new information is derived through self-study and self-experience and *collaborative learning*, through *social knowledge building* by interacting with other people (Scardamalia 2002). The concept of gain (Dascalu et al. 2010a) may be used for evaluating the contribution of each utterance to the overall discourse. It is derived from information theory (Shannon 1948; Kent 1983) and starting from the two types of knowledge-building processes, the following types of gain can be defined: *personal gain* when the interlinked utterances have the same speaker and *collaborative gain* when further information in the discussion thread is given by a different participant (Dascalu et al. 2010a).

As mentioned above, each utterance is evaluated and gets an importance score. The personal gain is obtained by summing up the utterance importance score and the gain of the previous inter-linked utterances of the same participant multiplied by the similarity between the previous interventions and the current one. Collaborative gain is obtained similarly, but considering utterances from different speakers. The process of computing the two types of gains is a recurrent process as each utterance gain is computed starting from the gain of the previous inter-linked ones.

Combining the utterance importance score with the gain gives us an estimation of the actual importance of an utterance in a given context, while cosine similarity (Manning and Schütze 1999) measures the strength, the impact, and the echoes between the two explicitly or implicitly inter-linked utterances. By summing up all previous influences we obtain a clear estimation of the retrospective effect for each utterance.

Eventually, collaboration for the entire discussion is evaluated by comparing the overall collaborative gain of all utterances to the sum of all individual utterance scores or to the sum of overall gains. These two measures provide the means to determine the percentage of actual collaboration within the discussion: *score based collaboration* expresses the percentage of information that is built/transferred in a collaborative manner, whereas *gain based collaboration* weights the collaborative gain relative to the overall gain (Dascalu et al. 2010a).

To conclude, gain measures the strength of the echo, score expresses the individual importance of each unit of analysis and, by combining them, the proposed method evaluates collaboration concerning voice intertwining and inter-animation. Figure 10 depicts an example

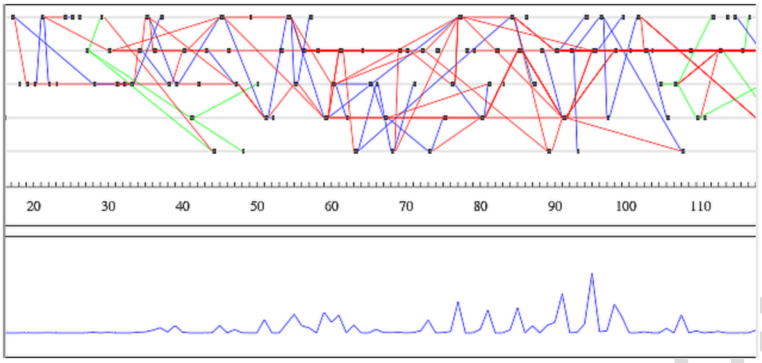


Fig. 10 Collaboration evolution within a chat conversation following the conversation timeline expressed in terms of utterance IDs

of collaboration assessment for a chat conversation from which intense collaboration zones can be identified, meaningful to the tutor as these areas contain a dense inter-exchange of semantically related utterances between different chat participants. In the upper part of the image there is the graph of utterances with the explicit links in red and the implicit ones in green, while in the lower part there is the graphics of the collaboration score for the conversation. It can be seen that our perspective on collaboration correlates with a high distribution of links between utterances of different participants in a short timeframe.

It should be emphasized that the actual formulas for computing the collaboration degree consider only the graph of utterances and similarity metrics. We are working now for considering also differential inter-animation patterns in utterance and collaboration evaluation. Even if *PolyCAFe* did not include this class of patterns in the automatic analysis, the visualization facilities support their detection, as discussed in a previous section

Transferability

Concerning transferability of *PolyCAFe* in different educational scenarios, the following dimensions must be taken into consideration: *domain*, *language* and the *learning task*. As the system was developed for English only, in order to ensure *language transferability* new linguistic tools must be integrated for each new language (e.g., the entire NLP pipe, lexicalized ontology, adjacency pairs and other linguistic patterns).

Domain transferability is mostly concerned with the existence of a large corpus of text documents, relevant to the task, that are required in order to build the LSA vector space. In addition, all domains, where textual descriptions of descriptive knowledge are used, are well suited (e.g., *PolyCAFe* was successfully used on medical discussion forums at the University of Manchester, in the LTfLL project). On the contrary, there are domains where *PolyCAFe* is not well suited due to the need of graphical elements or images or general discussions, without a clear focus and for which is difficult to build a relevant LSA space.

Moreover, from a *pedagogical point of view*, *PolyCAFe* can be used in a wide variety of collaborative contexts: role-based discussions and debates, open argumentations, problem solving (in mathematics and design or any domain specific task) or creative discussions (brainstorming) (Trausan-Matu 2010b). More specifically, we envision the following contexts: revising exams and discussions on given topics, finding collaborative solutions to problems that can be described without the importance of a sequence of steps (PBL) or further investigation of a given topic of interest to the learner (Self-Regulated Learning). On the other

hand, *PolyCAFe* is not suitable for learning scenarios in which collaboration is not required, nor encouraged, or settings that involve scripted collaboration.

Similar approaches with PolyCAFe

Very few systems can be used for similar tasks or for solving similar problems. We consider *KSV* (Teplovs 2008) and the system of Suthers and Desiato (2012) to be the most similar to *PolyCAFe*. Our approach is also in a way similar to the analysis performed by Fuks and Pimentel (2009). In addition, the detection of speech and justification acts from *PolyCAFe* may be used to support the detection of the multidimensional codes proposed by Strijbos (2009).

Both *PolyCAfe* and *KSV* offer visualizations of participation and interactions between users through SNA and semantic similarities between concepts or analysis elements (using LSA). However, there are several important differences between the two systems, the most important one is that *PolyCAFe* uses the notion of graph of utterances and provides feedback by implementing some elements from the polyphonic analysis model. In order to provide more insight between these differences and similarities, a detailed comparison between *PolyCAFe* and *KSV* is provided in Table 1.

The graph of utterances of *PolyCAFe* resembles the uptake network proposed by Suthers and Desiato (2012). Both approaches focus on an underlying discourse structure of the conversation highlighting dependencies between interventions, but the actual mechanism and factors are completely different. Whereas the uptake network focuses on lower level

t1.1 **Table 1** *PolyCAFe* versus KSV (Teplovs 2008)

t1.2	Benefits of *PolyCAFe*	Benefits of *KSV*
t1.3	Educational perspective	
t1.4	Dialogical perspective induced by voice inter-animation	
t1.5	Emphasis on collaboration in addition to a qualitative assessment of participant's involvement	A more shallow perspective of individuals and links between them
t1.6	Conversation topics extraction relevant for highlighting the focus of the discussion	
t1.7	The analysis is strictly based on textual information	Integration of addition relationships between 'notes' (e.g., annotation, authorial)
t1.8	Technical perspective	
t1.9	Explicit or implicit links between interventions are taken into consideration	Multiple types of relations between the nodes are considered: structural (e.g., reply-to, build-on, reference, annotation, contains), authorial, or semantic (based on LSA)
t1.10	Multiple NLP techniques applied on the initial interventions	
t1.11	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
t1.12		Clustering of nodes
t1.13	Comprehensive mechanism of utterance importance scoring	
t1.14		A multitude of parameters, configurations and visualizations available from the interface

functions of discourse and uses rules for inferring non-accidental relationships between participants' contributions, the focus in terms of the graph of utterances consists of highlighting a local cohesive context in which voices inter-animate.

The *PolyCAFe* approach of analysis may be considered similar with that of Fuks and Pimentel (2009) because both make a *recency analysis* (a windows of maximum 20 utterances is regarded while measuring semantic similarity between interventions, in the end enabling the evaluation of collaboration), a *cohesion analysis* (that is expressed through repetitions, synsets from WordNet and LSA) and a *coherence analysis*.

Based on the feedback already collected on *PolyCAFe*, a new system was developed—*ReaderBench* (Trausan-Matu et al. 2012; Dascalu et al. 2013). The graph of utterances was generalized towards a multi-layered cohesion graph (Trausan-Matu et al. 2012; Dascalu et al. 2013) in which cohesive links are determined through an aggregated similarity measure integrating semantic distances in ontologies (Budanitsky and Hirst 2006), cosine similarity in latent semantic vector spaces (Landauer and Dumais 1997) and similarity through topic models from Latent Dirichlet Allocation (Blei et al. 2003).

Evaluation

PolyCAFe's evaluation with emphasis in tutor and learner feedback consisted of two rounds of experiments. The first one was a pilot usage of the system with a limited number of participants and it showed promising results. Therefore, a second evaluation round was conducted with more participants that lasted for a longer period. The conversations resulted from this evaluation experiment were also manually annotated by tutors in order to verify the accuracy of *PolyCAFe*'s results in terms of utterance and participant assessment.

Pilot evaluation

A first pilot study (Rebedea et al. 2010; Dascalu et al. 2011) has been performed during the Human-Computer Interaction (HCI) course, in the academic year 2009–2010, at the Computer Science Department of the University Politehnica of Bucharest involving nine senior (4th year) students and five tutors that used *PolyCAFe* for analyzing the conversations and providing feedback to the students. The experiment was structured in the following manner: students had to read online and printed materials on a given topic (web technologies used for collaborative tasks—chat, log, forum and wiki) and then they had a debate using *ConcertChat* in two small groups of 4–5 students. After the debate, they used *PolyCAFe*'s feedback widgets to understand their involvement in the conversation and what could have been improved. Two tutors monitored this activity, provided help to the students and took notes regarding the asked questions, comments, and the actual behavior when interacting with the widgets. As the students were encouraged to think aloud when using the system, this data was useful for identifying the main problems of the software. Besides thinking aloud, the students were also asked to use a document where they registered what they considered misleading in the feedback returned by the system. This activity lasted 90–120 min and was followed by a questionnaire with 32 evaluation statements with answers on a 5-level Likert scale (1-strongly disagree—5-strongly agree), grouped in five categories: Pedagogic effectiveness, Efficiency, Cognitive load, Usability, and Satisfaction. Afterwards, a focus group with all students was conducted in order to find the most important advantages and disadvantages, plus suggestions for improvements.

On the other hand, tutors were asked to provide feedback to a chat conversation using *PolyCAFe* and to another one without the system. After this step, they were invited to answer a

JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

questionnaire with 35 evaluation statements using the same scale and categories as the one for the students. Then all tutors took part in a focus group where they were invited to share their points of view about the each feature's utility, about the reliability of the feedback, and the improvements they envisioned.

Overall, all students and tutors considered the feedback provided by *PolyCAFe* to be useful and relevant for their task (Rebedea et al. 2010), but the opinion of students was divided as just 63 % of them considered that the feedback was helpful in improving their learning experience. One explanation for this result might be the fact that the students have never used *PolyCAFe* prior to the evaluation and it might have been difficult for them to understand how to use all the provided facilities and to envision how they may use them for improving future participations in similar tasks. This explanation might be also suggested by the low score provided by students for cognitive load items, where the average agreement is as low as 56 %. On the other hand, the evaluation session highlighted that the usability of the widgets could also be improved and thus, their relevance and user acceptance might increase in the future.

Table 2 (Dascalu et al. 2011) presents the aggregated evaluation results on all the five categories for both tutors and students. It is clear that all the tutors found *PolyCAFe* efficient for their task, as it helps them reduce the time needed for providing feedback to students and it improves the quantity and consistency of this feedback among tutors. Moreover, it is easily noticeable that the student results are worse for all categories than the ones for the tutors. The lowest score was obtained for cognitive load, showing that the users had some problems accommodating to *PolyCAFe* on their first use. In addition, the results show that more than a quarter of the learners are not satisfied by the system and the main presented reason was that the students did not trust the statistical results displayed as they considered some indicators are not always accurate. In the case of tutors, pedagogical effectiveness items had the lowest average agreement percentage (83 %) as not all the tutors considered all the widgets effective for assessing the conversations: the search widget was considered the least effective, while the conversation visualization received the highest scores. Moreover, one of the items in this category was the following "The support provided by *PolyCAFe* is complementary to my expertise", and two tutors did not consider the system as providing a complementary function, but rather as a tool for enhancing the productivity when providing feedback while only two responded with "not applicable" and one with neutral.

While taking a closer look at the questionnaires, the tutors had agreed with all but one statement with average scores between 3.50 and 5.00, while the students had agreed with 27 out of the 32 statements, with average scores between 3.56 and 5.00. As it can be noted, there are considerable differences between the students' results and those of the tutors. However,

t2.1 **Table 2** *PolyCAFe* First evaluation results per category using the 5-level Likert scale (1-strongly disagree—5-strongly agree)

t2.2 Evaluation statement category	Tutors		Students	
t2.3	Average score	Agreement	Average score	Agreement
t2.4 Pedagogic effectiveness	4.11	83 %	3.94	77 %
t2.5 Efficiency	5.00	100 %	4.22	78 %
t2.6 Cognitive load	4.60	100 %	3.56	56 %
t2.7 Usability	4.36	93 %	4.11	81 %
t2.8 Satisfaction	4.57	91 %	3.89	72 %
t2.9 Total	4.53	93 %	3.94	73 %

possible explanations for these discrepancies might be: first, the tutors were more familiarized with the system as they had used it prior to analyze other conversations, second, the tutors overrated the system as it helps them provide feedback more quickly, as PolyCAFe improved the time required for the analysis by up to 50 %, and in a reliable manner, and third, the students did not perceive the utility of the tool as they are not involved in this kind of activity very often and therefore are not educated on how to use the feedback for future tasks.

All the identified aspects within this preliminary evaluation were used to increase the reliability and the usability of *PolyCAFe* and were treated in detail in a second, more elaborated, evaluation study.

Second evaluation experiment

After the first pilot showed that the system was efficient and effective for both learners and tutors, a new evaluation experiment (Rebedea et al. 2011) was undertaken to further study the effects of using an improved version of *PolyCAFe*, with a larger group of students. The experiment was integrated as a learning task and assignment for a group of senior year undergraduate students studying HCI during the academic year 2010–2011. A total of 35 students have been engaged in the study for several weeks: 25 students were part of the experimental group and 10 students were assigned to the control group. The only difference between the experimental and control group is that the latter did not receive any feedback from *PolyCAFe*, but only from the tutors. The learners were divided into groups of five students, thus having five experimental and two control groups, and were given two successive chat assignments related to web collaboration technologies (chat, blog, wiki, forums and *Google Wave*) to debate using *ConcertChat*.

In the first assignment, the experimental group was asked to use *PolyCAFe* to get feedback, while the control group did not use the system. The use of *PolyCAFe* for the second assignment was not mandatory, so the learners had an option to use the system only if they considered it would be useful for them. The tutors had to provide manual feedback to each of the students involved in the chat conversations for the first assignment. Each tutor assessed at least one conversation without using *PolyCAFe* and one conversation using the system. With regard to the second assignment, no manual feedback was provided, only the outputs of *PolyCAFe*.

At the end of the evaluation session, all the students and tutors were required to answer a questionnaire and to participate in focus groups. The results of the evaluation have been devised into several topics similar to the pilot study: tutor efficiency, quality and consistency of the automatic feedback, making the educational process transparent, quality of educational output, motivation for learning, etc. All these topics have been evaluated conditionally or with minor qualifications, but only three are presented extensively as they are considered central to the educational scenario and to the usage of the system for similar CSCL tasks:

- VT1: Tutors/facilitators spend less time preparing feedback for learners compared with traditional means (tutor efficiency);
- VT2: Learners perceive that the feedback received from the system contributes to informing their study activities (quality and consistency of automatic feedback);
- VT3: Learner performance in online discussions is improved in the areas of content coverage and collaboration, when using PolyCAFe (quality of educational output).

In order to evaluate *tutor efficiency*, several methods have been used: measurements, questionnaires and the answers to the interviews. Overall, all show a good consensus of the

six tutors with regard to the efficiency of using *PolyCAFe*, with averages over 4.5 and agreement factors of over 83 % for all the evaluation statements. In addition to the first evaluation, time measurements for preparing the feedback by tutors were also used. Thus, four tutors analyzed each chat conversation, two using *PolyCAFe* and the other without using the system. This data has been compared for all the seven chats resulted for the first assignment. The average time needed to prepare feedback without *PolyCAFe* was of 84 min, with a standard deviation of 15 min, while the average time required for providing feedback with *PolyCAFe* was of 55 min with a standard deviation of 20 min. These results show a significant average time reduction for a single chat conversation: (84−55)/84=35 %. However, as the standard deviation has increased, it also demonstrates that not all tutors managed to use the software efficiently.

Quality and consistency of the automatic feedback has been evaluated using questionnaires for the group of 25 students, plus system logging. The statements focused on the accuracy, the relevance, the usefulness and consistency of the provided feedback—all were evaluated with agreement factors between 60 and 80 % and means between 3.70 and 4.00. The system logging utilities monitoring student access to *PolyCAFe* have shown that for the whole period of the evaluation there have been 285 visits and 1,447 page-views, resulting in more than 40 page views on average per student. Therefore, the students have been actively using the system in order to reflect on their activity in specific chat conversations.

The last topic, *the quality of the educational output*, was evaluated through measurements computed by *PolyCAFe* for the second chat assignment, as a comparison between the experimental groups versus the control groups: the most important concepts in the conversation and their score, the average grade for utterances throughout the entire conversation, the number of interventions and the density of implicit and explicit links between utterances. However, only the average scores of utterances and the quantitative estimation of collaboration through the density of links (average number of links/utterance) showed a noticeable increase between the two groups: 6.8 % for the number of utterances and 29 % for the estimation of collaboration, both in favor of the experimental group.

Participant ranking verification

For all the chats of the first assignment, the tutors had to rank the participants according to the importance they had throughout the conversation, taking into account the content of their interventions, their involvement, and the degree of collaboration with the other participants. Therefore, each tutor assigned a rank from 1 to 5 for each participant; additionally, each participant was ranked by all other participants pertaining to the same conversation These results were then compared with the automatic ranking of participants provided by *PolyCAFe* (see Table 3).

Table 3 *PolyCAFe* Sample of participant rankings for a single chat conversation

Rank	S1	S2	S3	S4	S5	Stud. avg.	Tutor 1	Tutor 2	Tutor avg.	PolyCAFe
Student 1 (S1)	–	2	2	1	1	2	4	4	4	4
Student 2 (S2)	2	–	3	2	2	3	1	2	1–2	2
Student 3 (S3)	3	3	–	3	4	4	5	5	5	5
Student 4 (S4)	1	1	1	–	3	1	2	1	1–2	1
Student 5 (S5)	4	4	4	4	–	5	3	3	3	3

t4.1 t4.2 **Table 4** *PolyCAFe* Comparison of average participant rankings

Comparison	Correlation	Precision	Average error
Tutors—System	0.94	77 %	0.23
Students—System	0.84	66 %	0.43
Tutors—Students	0.84	71 %	0.40

By analyzing all seven conversations, *PolyCAFe* achieved an excellent precision and correlation with the average tutor scores (r=0.94 and P=77 %) and good results with the average student scores (r=0.84 and P=66 %) (see Table 4) (Rebedea et al. 2011). Moreover, these results show that students are also able to judge correctly who the most important peers in the conversation are. However, they achieved a lower correlation with the tutors (r=0.84 and P=71 %) than *PolyCAFe* most probably because they are not used to assessing their participation in online conversations.

Moreover, although the average rankings of the students, the tutors and *PolyCAFe* are quite well correlated one with the other, individual basis correlations drop dramatically due to the fact that a simple inversion in a series of 5 elements (the number of participants per conversion) changes the trend and therefore drastically diminishes inter-rater correlations. We have also observed that individual rankings provided by students have a more complex variation than those provided by tutors tend to agree more often. Nevertheless, the results encourage us to conclude that, although the grading or ranking criteria of tutors and students are not the same, the system is well correlated with the average values of these rankings, thus being more objective.

Conclusions

Chat-based CSCL brings new phenomena, which need theoretical foundations, methods, and computer support. Collaborative discourse building is one of the essential aspects of CSCL. The most remarkable case of discourse is encountered in natural language, but the musical case, especially classic polyphonic music and jazz improvisation, is another important example. Both cases have many features in common, as linguists (Tannen 2007), philosophers (Bakhtin 1984; Confucius 2003), philologists (Bakhtin 1984) and musicologists (Webern 1963) remarked.

The paper presents in more detail the polyphonic model we previously proposed (Trausan-Matu et al. 2005, 2007b; Trausan-Matu 2010c) and provides novel insights about the associated analysis method and the computer support provided by *PolyCAFe*. Justifications are provided for considering a musical metaphor and for developing a model and method of analysis for collaborative conversations. Moreover, in addition to the support for manual qualitative analysis, *PolyCAFe* provides automated quantitative measurements that estimate collaboration and personal involvement.

In other papers, we presented successful applications of the polyphonic model and the qualitative analysis method for chat session and face-to-face settings (Trausan-Matu et al. 2007b; Trausan-Matu and Rebedea 2009; Trausan-Matu 2013). This is not a surprise if we take into account that Bakhtin's dialogism, our substrate, is considered a philosophical system that may be viewed as including Hegel's dialectics (Marková 2003) and may provide an adequate basis for CSCL.

However, we should discuss some limitations we have identified. In the analysis, we have access only to utterances (and a problem is also the granularity level at which we consider

JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

them) and explicit links. Detection of implicit links, of voices and inter-animation patterns may be very difficult sometimes. Moreover, potentially each noun or verb or even some other parts of speech (e.g. the adjective *perpendicular* in a chat on geometry) may become a voice. For a comprehensive analysis, the whole context of an utterance should be considered and this context might include an extremely high number of previous utterances. In addition, it is extremely difficult, if not even impossible to implement efficient algorithms for their detection and for the evaluation of involvement and collaboration. We tried to design and implement automatic analysis tools that support human analysts (and teachers). This is a complex task and we did not even complete the implementation of all the ideas of our analysis method, for example, the detection of differential inter-animations.

PolyCAFe, in addition to the support for qualitative analysis of collaboration for research purposes, was designed for helping tutors in providing feedback to learners. To further this aim, quantitative values are computed starting from the graph of utterances built in the first steps of the polyphonic analysis method. In contrast to previously implemented systems (Trausan-Matu et al. 2007a; Dascalu et al. 2008), *PolyCAFe* experimentally implemented facilities for providing feedback to learners as well.

An evaluation of the effectiveness of the provided facilities was performed. Tutors considered that reducing the time needed to provide manual feedback to their students was the greatest advantage of using *PolyCAFe*. From the learner perspective, the displayed indicators and the provided textual feedback were the main benefits for improving the student's future collaborative learning activities.

Overall, from a technical viewpoint, *PolyCAFe* can be considered a learning-analytics tool for online conversations that underpins the dialogic and polyphonic theories and that employs NLP and SNA techniques in order to discover implicit relations between utterances. These links are used to build a graph of utterances that is exploited later on for evaluating utterances, participants' involvement and collaboration. Moreover, the evaluations performed in a formal, academic environment have shown the acceptance and usefulness of *PolyCAFe*, both for tutors and learners. However, *PolyCAFe* is a starting point; it implements only some of the first steps of the polyphonic analysis method. It is not considering differential inter-animation and the cohesion evaluation is based on LSA, which is rather limited in capturing complex semantic relations. We also started to investigate new computer support facilities, which take into account the differential patterns, detection of voices and of discourse structures using other NLP methods (Chiru and Trausan-Matu 2012).

Acknowledgments The authors wish to express their thanks to the anonymous reviewers for their extensive and very useful comments. We would like to mention the thoughtful advice of Gerry Stahl. We would also like to thank Alexandru Gartner, Dan Banica and the students and tutors who participated in the validation and verification experiments. The research presented here has been partially performed under a Fulbright Scholar post-doc grant (awarded to Stefan Trausan-Matu) and was also supported by the FP7 ICT STREP project LTfLL (http://www.ltfll-project.org/) and by project FP7-REGPOT-2010-1, nr. 264207, ERRIC.

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: IEEE.

Avouris, N., Fiotakis, G., Kahrimanis, G., Margaritis, M., & Komis, V. (2007). Beyond logging of fingertip actions: Analysis of collaborative learning using multiple sources of data. *Journal of Interactive Learning Research, 18*(2), 231–250.

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

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

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

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

Berlanga, A.J., Van Rosmalen, P., Trausan-Matu, S., Monachesi, P., & Burek, G. (2009). The language technologies for lifelong learning project. In *9th IEEE Int. Conf. on Advanced Learning Technologies (ICALT 2009)* (pp. 624–625). Riga: IEEE.

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

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology, 25*(2), 163–177.

Budanitsky, A., & Hirst, G. (2006). Evaluating WordNet-based measures of lexical semantic relatedness. *Computational Linguistics, 32*(1), 13–47.

Cazden, C. B. (1993). Vygotsky, Hymes, and Bakhtin: From word to utterance and voice. In E. A. Forman, N. Minick, & C. A. Stone (Eds.), *Contexts for learning: Sociocultural dynamics in children's development* (pp. 197–212). Oxford: Oxford University Press.

Chiru, C. G., & Trausan-Matu, S. (2012). Identification and classification of the most important moments from students' collaborative discourses. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *11th Int. Conf. on Intelligent Tutoring Systems (ITS 2012)* (pp. 330–339). Chania: Springer.

Confucius. (2003). *Analects*. Indianapolis: Hackett Publishing Company, Inc.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (Eds.). (2009). *Introduction to algorithms* (3rd ed.). Cambridge: MIT Press.

Dascalu, M., Chioasca, E. V., & Trausan-Matu, S. (2008). ASAP – an advanced system for assessing chat participants. In D. Dochev, M. Pistore, & P. Traverso (Eds.), *13th Int. Conf. on Artificial Intelligence: Methodology, Systems, and Applications (AIMSA 2008)* (pp. 58–68). Varna: Springer.

Dascalu, M., Rebedea, T., & Trausan-Matu, S. (2010a). 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: Springer.

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

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: ISLS.

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: Springer.

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.

Dowell, J., & Gladisch, T. (2007). Design of argument diagramming for case-based group learning. *ACM International Conference Proceeding Series, 250*, 99–105.

Dowell, J., Tscholl, M., Gladisch, T., & Asgari-Targhi, M. (2009). Argumentation scheme and shared online diagramming in case-based collaborative learning. In *9th Int. Conf. on Computer supported collaborative learning (CSCL'09)* (pp. 567–575). Rhodes: ISLS.

Eastman, J. K., & Swift, C. O. (2002). Enhancing collaborative learning: Discussion boards and chat rooms as project communication tools. *Business Communication Quarterly, 65*(3), 29–41.

Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*(1), 35–41.

Fuks, H., & Pimentel, M. (2009). Studying response-structure confusion in VMT. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 373–397). Boston: Springer.

Harrer, A., Hever, R., & Ziebarth, S. (2007). Empowering researchers to detect interaction patterns in e-collaboration. In R. Luckin, K. R. Koedinger, & J. E. Greer (Eds.), *13th Int. Conf. on Artificial Intelligence in Education (AIED 2007)* (pp. 503–510). Los Angeles: Frontiers in Artificial Intelligence and Applications.

Hmelo-Silver, C.E., Chernobilsky, E., & Masto, O. (2006). Analyzing collaborative learning: Multiple approaches to understanding processes and outcomes. In *7th Int. Conf. of the Learning Sciences (ICLS '06)* (pp. 1061–1062). Bloomington, IN, USA.

Holmer, T., Kienle, A., & Wessner, M. (2006). Explicit referencing in learning chats: Needs and acceptance. In W. Nejdl & K. Tochtermann (Eds.), *First European Conference on Technology Enhanced Learning, EC-TEL 2006* (pp. 170–184). Crete: Springer.

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

JrnlID 11412_ArtID 9190_Proof# 1 - 31/03/2014

Kent, J. T. (1983). Information gain and a general measure of correlation. *Biometrika, 70*(1), 163–173.

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.

Kumar, R., Chaudhuri, S., Howley, I., & Rosé, C.P. (2009). VMT-Basilica: An environment for rapid prototyping of collaborative learning environments with dynamic support. In *9th Int. Conf. on Computer supported collaborative learning (CSCL'09)* (pp. 192–194). Rhodes: 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.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes, 25*(2/3), 259–284.

Law, N., Lu, J., Leng, J., Yuen, J., & Lai, M. (2008). Understanding knowledge building from multiple perspectives. Utrecht, Netherland.

Ligorio, M. B., & Ritella, G. (2010). The collaborative construction of chronotopes during computer-supported collaborative professional tasks. *International Journal of Computer-Supported Collaborative Learning, 5*(4), 433–452.

Mahnkopf, C. S. (2002). Theory of polyphony. In C. S. Mahnkopf, F. Cox, & W. Schurig (Eds.), *Polyphony and complexity* (Vol. 1, p. 328). Hofheim: Wolke Verlags Gmbh.

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

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

Marková, I. (2003). *Dialogicality and social representations*. Cambridge: Cambridge University Press.

Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of text difficulty: Testing their predictive value for grade levels and student performance*. Washington, DC: Council of Chief State School Officers.

Newman, M. E. J. (2010). *Networks: An introduction* (1st ed.). Oxford: Oxford University Press.

Rebedea, T., Dascalu, M., Trausan-Matu, S., Banica, D., Gartner, A., Chiru, C. G., et al. (2010). Overview and preliminary results of using PolyCAFe for collaboration analysis and feedback generation. In M. Wolpers, P. Kirschner, M. Scheffel, S. Lindstaedt, & V. Dimitrova (Eds.), *5th European Conference on Technology Enhanced Learning (EC-TEL 2010)* (pp. 420–425). Barcelona: Springer.

Rebedea, T., Dascalu, M., Trausan-Matu, S., Armitt, G., & Chiru, C. G. (2011). Automatic assessment of collaborative chat conversations with PolyCAFe. In C. D. Kloos, D. Gillet, R. M. Crespo García, F. Wild, & M. Wolpers (Eds.), *6th European Conference of Technology Enhanced Learning (EC-TEL 2011)* (pp. 299–312). Palermo: Springer.

Sacks, O. (2007). *Musicophilia: Tales of music and the brain*. New York: Vintage Books.

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.

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: MIT Press.

Stahl, G. (2009a). *Studying virtual math teams*. New York: Springer.

Stahl, G. (2009b). The VMT vision. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 17–29). New York: Springer.

Stahl, G. (2013). *Translating Euclid: Designing a human-centered mathematics*. San Rafael, CA: Morgan & Claypool Publishers. http://gerrystahl.net/elibrary/euclid.

Strijbos, J. W. (2009). A multidimensional coding scheme for VMT. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 399–419). Boston: Springer.

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.

Tannen, D. (2007). *Talking voices: Repetition, dialogue, and imagery in conversational discourse* (2nd ed.). Cambridge: Cambridge University Press.

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: 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. (2010c). The polyphonic model of hybrid and collaborative learning. In F. L. Wang, J. Fong, & R. C. Kwan (Eds.), *Handbook of research on hybrid learning models: Advanced tools, technologies, and applications* (pp. 466–486). Hershey: Information Science Publishing.

Trausan-Matu, S. (2012). Repetition as artifact generation in polyphonic CSCL chats. In *Third Int. Conf. on Emerging Intelligent Data and Web Technologies* (pp. 194–198). IEEE.

Trausan-Matu, S. (2013). Collaborative and differential utterances, pivotal moments, and polyphony. In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions* (Computer-supported collaborative learning series, Vol. 15, pp. 123–139). New York: Springer.

Trausan-Matu, S., & Rebedea, T. (2009). Polyphonic inter-animation of voices in VMT. In G. Stahl (Ed.), *Studying virtual math teams* (pp. 451–473). Boston: Springer.

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). Iasi: Springer.

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: ISLS.

Trausan-Matu, S., Stahl, G., & Zemel, A. (2005). Polyphonic inter-animation in collaborative problem solving chats. Philadelphia: Drexel University, http://mathforum.org/wikis/uploads/Stefan_Interanimation.doc.

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, (CRIWG 2006)* (pp. 132–139). Medina del Campo: Springer.

Trausan-Matu, S., Rebedea, T., Dragan, A., & Alexandru, C. (2007a). 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., & Sarmiento, J. (2007b). Supporting polyphonic collaborative learning. *Indiana University Press, E-service Journal, 6*(1), 58–74.

Trausan-Matu, S., Dessus, P., Lemaire, B., Mandin, S., Villiot-Leclercq, E., Rebedea, T., et al. (2008). Deliverable D5.1 LTfLL – Support and feedback design. Heerlen: OUNL, Research report of the LTfLL Project.

Trausan-Matu, S., Dessus, P., Rebedea, T., Mandin, S., Villiot-Leclercq, E., Dascalu, M., et al. (2009). Deliverable D5.2 LTfLL – Learning support and feedback. http://dspace.ou.nl/handle/1820/2251.

Trausan-Matu, S., Rebedea, T., & Dascalu, M. (2010). Analysis of discourse in collaborative learning chat conversations with multiple participants. In D. Tufis & C. Forascu (Eds.), *Multilinguality and interoperability in language processing with emphasis on Romanian* (pp. 313–330). Bucharest: Editura Academiei.

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: Springer.

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

Webern, A. (1963). *The path to the new music*. Pennsylvania: Theodore Presser Co.