Implementing Bakhtin's Dialogism Theory with NLP Techniques in Distance Learning Environments

Philippe Dessus, Stefan Trausan-Matu

Implementing Bakhtin's Dialogism Theory with NLP Techniques in Distance Learning Environments

Philippe Dessus[1] and Stefan Trausan-Matu[2,3]

[1] Université Pierre-Mendès-France, Grenoble, France
[2] University "Politehnica" of Bucharest, Department of Computer Science and Engineering, Bucharest, Romania
[3] Research Institute for Artificial Intelligence of the Romanian Academy, Bucharest, Romania
Philippe.Dessus@upmf-grenoble.fr, stefan.trausan@cs.pub.ro

Abstract. The purpose of this paper is to present and discuss a Natural Language Processing-based operationalization of Bakhtin's dialogism ideas, and to insert it in the context of distance learning. Three core concepts from Bakhin are introduced and their application in the domain of distance learning discussed. Then, two computer-based systems based on these ideas, *PolyCAFe* and *Pensum* are presented.
Keywords: Dialogism, CSCL, Natural Language Processing, Latent Semantic Analysis

1 Introduction

Contemporary research on education often considers education-based events as processes developed through dialogue, embedded in a community of practice [1]. This view has been taken in many research fields, like teacher education [2], pedagogy [3] and Computer-Supported Collaborative Learning (CSCL) [4]. The dialogic approach considers learning as taking a perspective of another in a dialogue [5].

The above vision is rooted in the ideas of Mikhail Bakhtin [6, 7, 8], who extended the scope of dialogs emphasizing the role of considering the otherness, considering that any linguistic act, either spoken or written is a dialog. Moreover, the universality of dialog gives birth to dialogism, which "has been characterized as a meta-theory for the human mind" [9, p. 30]. Koschmann [4] was one of the first scholars to present a dialog-based model of CSCL, starting from Bakhtin's theories, which emphasizes the role of the language and especially of utterances, either spoken or not, language being a social mediator of knowledge building [10]. In many domains learning may be seen as mastering a speech genre, as becoming able to enter into a dialog or even to participate to a polyphony of voices [11]. These voices should be understood in a larger extend than the acoustic sense, as it will be discussed later. They may be actually present, like in a voiced conversation, but they may also be implicitly present in any text. From this perspective, Bakthin's dialogism suits not only collaborative knowledge building but also individual knowledge building —see also Stahl's two cycles model [12]—which implies explicit or implicit inner dialogs among learner's voice and the other voices (e.g., author's, professor's, commenters' voices).

The purpose of this paper is to present and discuss an attempt to operationalize Bakhtin's ideas on dialogism with Natural Language Processing (NLP) techniques, and to insert them in the context of distance learning. We have extended this dialogical view by using computer-based natural language technologies as a means to "expand the space of learning" by emphasizing the presence of the 'other voices' through generating automatic feedback. This feedback is given to different types of text written by students: essays, instant messenger (chats) and forums.

The remainder of this paper is as follows. We first detail the core concepts in Bakhtin's dialogism that could be at hand in every distance learning situation, namely utterances, voices and echoes. Then we present a comprehensive framework for instructional situations at a distance that take into account Bakhtin's ideas. Eventually, we present two computer-based systems (*PolyCAFe* and *Pensum*), also based on dialogical concepts. We assume that these systems can foster student's knowledge building, either individually or collaboratively.

2 Core Concepts in Bakhtin's Dialogism

We consider in this paper three core concepts from Bakhtin's work: utterances, voices and echoes. This section introduces each of them and explains how they can be used to describe usual instructional interactions. We further try to illustrate the application of Bakhtin's theory in the context of distance-learning environments. We claim here that this theory is particularly well suited to instructional computer-mediated situations.

Let us consider students and a teacher/tutor engaged in a distance learning situation. They can either read texts, write out notes, essays, or discuss such or such topic. All the stakeholders performing these activities write

texts or say something in natural language, producing "utterances" that can become "voices" populating the distance learning platform, responding to each other. The way the students can, upon a given question, gather information from multiple text sources in order to compose their own piece of text (mainly, summaries or syntheses) might be viewed as "contexts" in which they try to handle the polyphony of voices. Let us define these three core concepts and contextualize them in distance learning.

2.1 Utterances

Natural language traditional units of analysis are words and sentences. Signs or words are not the mere representation of something, but a mediation between people [13]. As Wegerif [5, p. 144] puts it: "Any sign taken to be a mediation between self and other, a word or a facial expression, must pre-suppose the prior opening of a space of dialogue (an opening of a difference between voices) within which such a sign can be taken to mean something." However, even if Bakhtin considers words as very complex because "they are filled with overtones of previous voices and contexts" [6, p. 293], he states that *utterances* should be the central unit of analysis [7].

Each utterance may be considered as a potential "*voice*" which enters in relations with other ones, either of the same emitter of the voice or of other persons, therefore appearing a multitude of voices which would be a desideratum to become a polyphony [11]. The notion of utterance differs from that of sentence, whose boundary is more precisely defined. The boundary of an utterance is the change of subject at stake in a discussion. Wegerif [5, p. 144] expressed that "the meaning of an utterance is given by its location within a dialogue", and adds further that its meaning is neither reducible to the speaker's intentions, nor the response, but emerges from these two (Bakhtin's notion of inter-animation).

An *utterance* is a piece of text whose boundary is the change of speech subject [7, p. 91], and is not a purely individual act. Additionally, one can answer to an utterance: "When hearing or reading, we clearly sense the end of the utterance, as if we hear the speaker's concluding dixi." [7, p. 76]. Thus the boundary between utterances is not necessarily the turn-taking, as Dong [14, p. 524] expresses it: "An utterance boundary is defined by the turn-taking between speakers, that is, when one speaker stops talking and the next speaker begins. This orderly mapping can be revealed by looking at the coherence between any two utterances (communicative acts) as a function of the 'distance' between the utterance boundaries."

According to Bakhtin ([7, p. 91], see also [15]), an utterance is "[...] a link in the chain of speech communication of a particular sphere. […] Utterances are not indifferent to one another, and are not self-sufficient; they are aware of and mutually reflect one another… Every utterance must be regarded as primarily a response to preceding utterances of the given sphere (we understand the word 'response' here in the broadest sense). Each utterance refutes, affirms, supplements, and relies upon the others, presupposes them to be known, and somehow takes them into account… Therefore, each kind of utterance is filled with various kinds of responsive reactions to other utterances of the given sphere of speech communication". A major consequence is that having an utterance-based vision is forcing us to respond, to have an active, creative attitude, to inter-animate.

In our vision, an utterance may be a simple word or interjection or it may be even a a whole novel, as also Bakhtin mentioned [7]. An utterance may contain other utterances, at a different degree of complexity, the process of delimiting an utterance being a difficult task.

Since many instructional activities are dialog-based, utterances formed by questions or answers can be analyzed from a Bakhtinian perspective: they are both acts of communication and pieces of discourse [9] and represent the way teachers and students individually negotiate the construction of knowledge. The following notion of *voice* shows how utterances are embodied, having effects on the subsequent development of the conversations.

2.2 Voices and ventriloquism

An utterance may become a *voice* and may contain *echoes* and overtones of other voices [7]. Even if it is a central concept in Bakhtin's work, it is very hard to define exactly what a voice is and this difficulty is augmented by the phenomenon of *ventriloquism* [8], which is the (re)emitting of a voice by another one. As defined elsewhere [11] in the context of linguistic communication, we consider a voice "not the acoustical, physical, vocal expression of a given participant in a dialog but, rather, a distinct position, an utterance, an event or a recurrent series of events of emitting utterances that are heard, remembered, discussed and have influence on the utterances emitted by the other voices. In music, for example, a voice is not fixed to an instrument; the same instrument may play several voices, and different instruments may take the position of a given voice, simultaneously or sequentially. A voice may be seen as a distinctive position in a group, a person or a group of people who have uttered something, with effects on the subsequent utterances […] Moreover, a voice has some

particularities; it may have a personality, goals, beliefs, desires and emotions. Consequently, a dialog among several voices is not a dialog among impersonal entities. From another point of view, a voice may become a theme or may contribute to a theme of the discussion. " [11]

Cazden [16] restricts the notion of voice as follows, but to a larger extent, a voice can also be from another speaker: "Voice is Bakhtin's term for the 'speaking consciousness': the person acting—that is speaking or writing in a particular time and place to known or unknown others. Voice and its utterances always express a point of view, always enact particular values. They are also social in still a third meaning: taking account of the voices being addressed, whether in speech or writing. This dialogic quality of utterances Bakhtin calls 'responsivity' or 'addressivity'" [16, p. 198]. Linell [9] remarks that there is an extension of the concept of voice in which it is viewed as perspectives or topics. There are generalized voices, "tied to a group of sense-makers, rather than a single individual", reflecting the "perspective on a topical domain" [9, p. 116]. Moreover, "the same person may appropriate, internalize or express several different voices, whether these voices are taken from other individuals or they are generalized voices" (ibid.).

Voices express ideas and opinions and "polyvocality" (co-presence of multiple voices) is sought in instructional distant situations. Teachers and students can hold ideas or opinions from multiple perspectives, which are encouraged in a collaborative knowledge building. The next notion of *echo* describes how different voices are replicated among contexts.

2.3 Contexts and echoes

From Bakhtin's perspective [7, p. 147] : "A context is potentially unfinalized; a code must be finalized. A code is only a technical means of transmitting information; it does not have cognitive, creative significance. A code is a deliberately established, killed context." From our point of view, a *context* may be considered as a slice of a discussion thread whose utterances have a strong cohesion. The participants of a context are either all the persons who utter at least once, or the authors of the texts under discussion. The voice may be seen as a self-coherent line of contexts, uttered by one or several participants.

The *echo of a voice* is a replication of it with enough strength to influence other voices in one or more contexts. Two kinds of echoes are taken into consideration: individual echoes, i.e., the way a given student internalize a voice, and collective echoes, i.e., the way several students react to a voice and enrich it.

A support is a channel dedicated to the conveying of voices (not necessarily those of the participants). Supports we consider are essays, chat discussions, texts read. The timeline is the time-to-time track of all the voices uttered during a learning period. It can be represented and visualized in a textual or a graphical form. Some information can be added to this timeline.

2.4 An example

For understanding the ideas of utterance, voice and echoes, let us consider the following problem, which was solved only collaboratively in a chat session by a group of students after they had the opportunity to solve it also individually but they didn't (see also [11]):

> "Three years ago, men made up two out of every three Internet users in America. Today the ratio of male to female users is about 1 to 1. In that time the number of American females using the Internet has grown by 30,000,000, while the number of males who use the Internet has grown by 100%. By how much has the total internet-user population increased in America in the past three years?
> (A) 50,000,000 (B) 60,000,000 (C) 80,000,000 (D) 100,000,000 (E) 200,000,000"

In Figure 1 is a chat excerpt that includes the main utterances that contributed to the finding of the solution. The lines 360 to 364 may be seen as five utterances but they may be also considered (especially from the perspective of the discussion in section 2.1 about the boundaries of utterances) as a single utterance of Mic, which is becoming a voice which, even if it is a joke, a buffoonery, has echoes on the following voices: on Hal's utterance in line 371, which also becomes a voice, which in her turn has echo on Cosi's utterances 387, 392-396, 399-400 and 403-405, which inter-animate with Mic's and Hal's previous voices eventually becoming the voice that solves the problem. The text stating the problem and the chat excerpt constitute a context.

350 Mic how do we do this..
351 Mic: Without knowing the total number
352 Mic: of internet users?
….
357 Dan: it all comes from the 30000000
358 Mic: did u get something for 10?

359 Dan: we already know
360 Mic: 30000000 is the number of increase in american females
361 Mic: and since the ratio of male to female
362 Mic: is 1 to 1
363 Mic: thats all i got to give. Someone finish it
364 Mic: Haha
365 Cosi: *haha you jackass*
366 Mic: Haha
367 Dan: Hahaha
368 Mic: u all thought i was gonna figure it out didn't
369 Mic: U
370 Mic: huh?
371 Hal: it would be 60,000,000
372 Mic: Hal
373 Mic: its all u
374 Mic: See
375 Mic: i helped
376 Cosi: *ok, so what's 11 – just guess on 10*
….
386 Mic: lets get back to 5
387 Cosi: *i think it's more than 60,00000*
388 Mic: way to complicate things
389 Cosi: *haha sorry*
390 Mic: life was good until you said that
391 Mic: :(
392 Cosi: *they cant get higher equally and even out to a 1 to 1 ratio*
393 Cosi: *oh, no wait, less than that*
394 Cosi: *50000000*
395 Cosi: *yeah, it's that*
396 Cosi: *im pretty sure*
397 Mic: Haha
398 Mic: how?
399 Cosi: *because the women pop had to grow more than the men in order to even out*
400 Cosi: *so the men cant be equal (30)*
401 Mic: oh wow...
402 Mic: i totally skipped the first sentencwe
403 Cosi: *therefore, the 50,000,000 is the only workable answer*
404 Dan: very smart
405 Cosi: *Damn im good*

Fig. 1. An excerpt illustrating a collaborative solution construction

2.5 Utterances, Voices and Echoes All Together in Distance Learning Contexts

Detecting utterances during instructional interactions, voices expressing a given content and their echoes upon contexts can be of importance in distance learning, at least because a lot of these interactions are not multimodal (mostly text-based) and not synchronous (i.e., through chat or forums). Because of these features, the whole instructional process is opaque and the intentions of the stakeholders difficult to guess: teachers encounter difficulties to see what content is actually understood by students and the latter have difficulties to get just-in-time feedback on their productions.

We present now a framework that both serves to represent the different levels of an instructional event and has to be compatible with Bakhtin's theory. Every instructional event is embedded within lesson episodes [17], which have a single pedagogical intent. Each sequence can be split in segments. Figure 2 depicts the framework. Each instructional sequence can be viewed at three levels: a *structural* one (above the arrow), which comprises the main structural features of the episode (its goals, its taught knowledge, the planned pedagogical activities that have been designed by the teacher); a *behavioral* one (below the arrow), which encompasses the behaviors (either) of the stakeholders and a *social* one (all below), which pertains to the social scripts engaged in the instructional situation [18]. The classical learning approaches focus on the structural level, the dialogistic perspective of Bakhtin giving light on the importance of social and behavioral levels. We can rephrase the activity of learning and understanding a new topic with Bakhtin's own words: a word is learnt when "it becomes

one's own", when "the speaker populates it with his own intention, its own accent, when he appropriates the word, adapting it to his own semantic and expressive intention" [6, pp. 293–394]. First of all, from our perspective, that means that the learner uses the needed words in the chat and in summaries. Moreover, she should use them in inter-animation patterns [19] for example, in repetitions, difference making, etc.

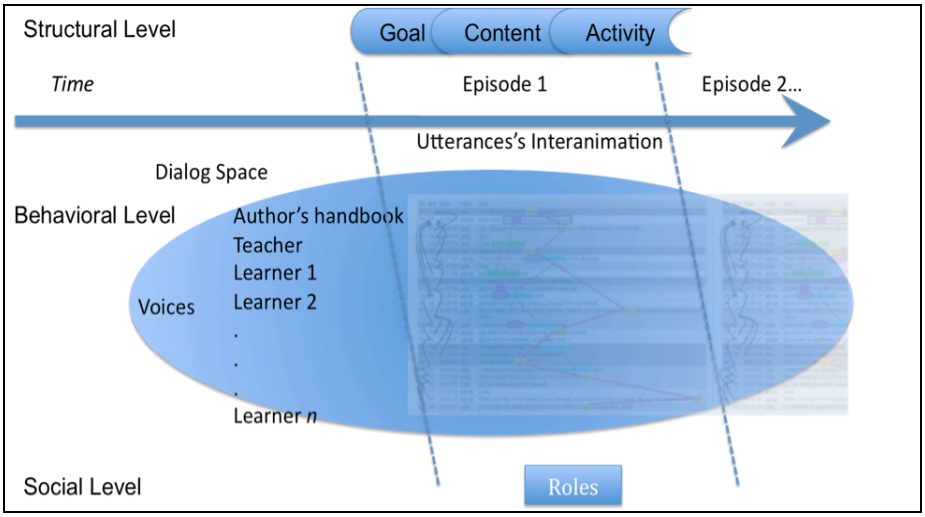

Fig. 2. Graphical representation of an instructional situation.

In every instructional situation there are several interacting voices, more or less explicit, uttered by e.g., the author, the teacher, the learner that reads texts, the echoes of the voices present in each word. Moreover, from this multivocal perspective, texts become meaning generation mechanisms, facilitating understanding and creative thought [20, 21]. A consequence is that in education, "the interaction of oral and written discourse increased dialogicality and multivoicedness and therefore provided more chances for students to learn than did talking or writing alone" [21, p. 385]; the dialogic and multivoicedness features of any utterance, even written, may be unifying factors for the integration of several learning environments. This claim has two main consequences:

- One could use the "dialogicity" (or "inter-animation") of the interactions in an environment to assess the degree of quality of learning/teaching. The more dialogic the situation is, the more the students can learn from it.
- One can use tools for analyzing the way these voices interact to each other.

Natural Language Processing techniques can be used to automatically analyze some features of the texts at hand during an instructional process at distance and to uncover some of their underlying utterances, voices and echoes. The next section is devoted to introduce two NLP-based systems (*PolyCAFe* and *Pensum*) pursuing this aim.

3 Dialogical NLP-based Systems

PolyCAFe and *Pensum* are respectively positioned on the collective and individual level of utterances. They use several NLP techniques that have been detailed elsewhere [22, 23].

PolyCAFe [24] analyses inter-animation and collaboration between participants to a distance learning situation. An utterance in the current version is considered an individual line input by the chat users, completed when the user enters a carriage-return. The echoes are determined starting from implicit or explicit links. From these links a graph is constructed connecting utterances and, in some cases, words. In this graph, contexts may be identified as groups of threads of implicit or explicit links. Simple examples of threads are repetitions of words or lexical chains. The same utterance may, of course, be included in several threads. Multivocality is detected by the presence of interwoven threads.

For example, in Figure 3, a high degree of collaboration (showed in the graph from the middle of the image) between utterances 150 to 200 (see the numbered scale) is generated due probably to the high degree of inter-animation (debate) among the usage of the words 'chat' and 'forum' (whose usage is shown in the top part of the figure). The debates on these words will echo what the learners will further read and the summaries they will write.

To that end, a second system, *Pensum* [25] (see Figure 4), has been devised to analyze individual threads of voices created when students read course materials and write out syntheses or summaries. In *Pensum* utterances are short pieces of syntheses, and not dialog-based turn-taking as in *PolyCAFe*. *Pensum* focuses on semantic

relations between the authors of courses and students. The voice of the teacher should be present in the reading-summarizing loop if the learner attended and understood the lectures. Similarly, the voices of the authors of teaching documents should be present in the texts written by the students, if they read and understood the important concepts. The voice of the learner should also be present in the creative writing, in which the learned concepts are deeply internalized. All these voices are embedded in cohesive contexts formed by students' syntheses, which represent the current state of understanding of a given course content. *Pensum* uses Latent Semantic Analysis (LSA) for detecting the presence of teacher's and teaching document authors' voices in students' summaries.

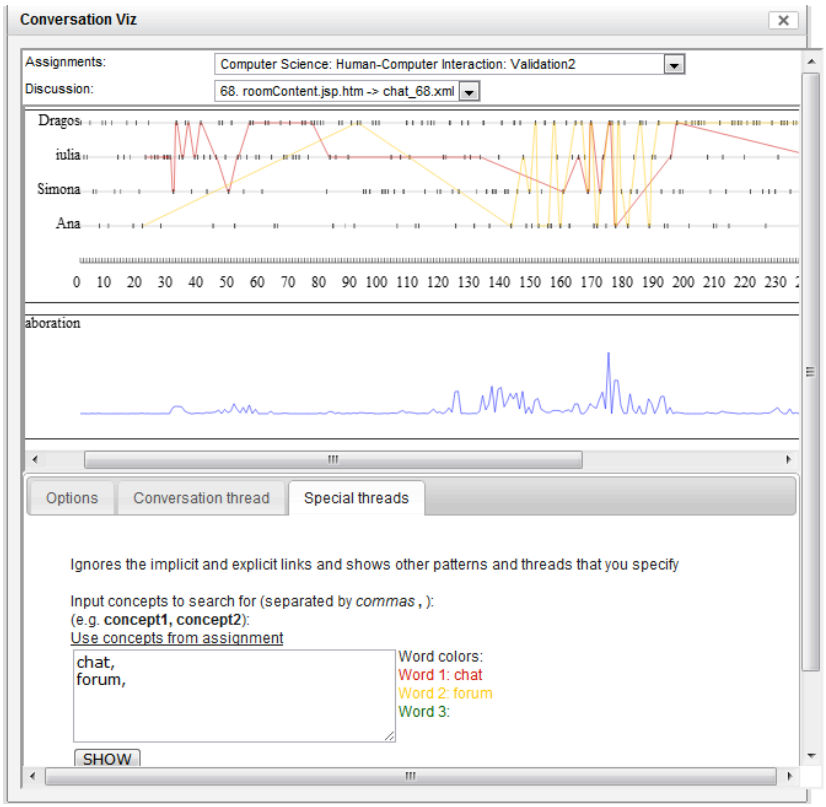

Fig. 3. Inter-animation visualization in *PolyCAFe*.

4 Conclusions and Further Developments

This study is the first step of our modeling of Bakhtin's views with NLP techniques. Despite some difficulties due to the vagueness and the multiple meanings carried by these three notions, we plan to improve their operationalization with NLP techniques in trying to improve the facilities offered by the implemented tools and their effectiveness in instructional settings. A very important point is to recognize whether or not a discussion is monologic or dialogic, polyphonic. Multivocality or polyphony means that there are permanently several voices entering in competition. Each utterance is filled with "overtones" of other utterances. A first problem is to detect these overtones. The first experiments with *PolyCAFe* are encouraging in this sense.

Pensum allows students and teachers to instantly recognize (and be aware of) the sources ("voices") of what they are learning and teaching. This awareness is very important, first to prevent from plagiarism and also for students' self-regulated learning.

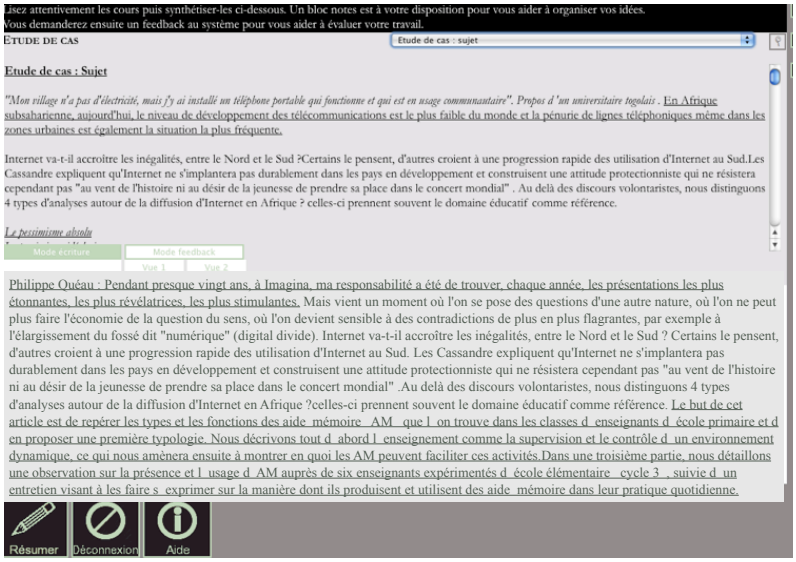

Fig. 4. *Pensum* and its two main windows. The course window (top) and the synthesis window (bottom). The underlining warns the student if the corresponding sentences are not in the synthesis (top) or are not in the course (bottom).

A further implementation would allow us to use these two systems in an integrated way. Through the framework depicted in Figure 2, we plan to model with NLP techniques a large range of the text-based learning activities such as searching documents, reading, writing summaries or forum posts, chatting, and even "moving" in a distance learning platform.

Since chat utterances (and even short course notes) can be very short pieces of texts, their analysis with LSA is difficult. According to Bakhtin's view of polyphony, our aim is to detect the "contexts" of a chat discussion. The method may be as follows:

1. Split the discussion into cohesive pieces of text. Each of them is called a context;
2. Enrich the context. In each of the contexts, the texts each student who participated to the context read or wrote are added, with a specific weight.

Several methods have to be tested for step 2. There are many ways to give an adequate weight to the texts. For instance, the texts read are added once while those written out are added twice in the database. We alternately plan some more elaborated ways of weighting, e.g., add the enrichment proportionally to the semantic relation of the utterance to the texts read/written. For instance, when an utterance has a weak relation to the texts previously read/written, its weight will be also weak (added once); it will be added many times as its semantic relation with texts read/written increases.

Acknowledgements

The research presented in this paper was partially performed under the FP 7 EU STREP project LTfLL (Language Technologies for Lifelong Learning). The authors wish to thank Benoît Lemaire, Sonia Mandin, Emmanuelle Villiot-Leclercq and Virginie Zampa who collaborated on the implementation of *Pensum*; as well as Traian Rebedea, Mihai Dascalu, Alexandru Gartner, Costin Chiru, Dan Banica and Dan Mihaila who collaborated on the implementation of *PolyCAFe*. Thanks to the MSH-Alpes which provided material help.

References

1. Wenger, E.: Communities of practice: Learning, meaning and identity. Cambridge University Press, Cambridge (1998)
2. Burbules, N.C. and Bruce, B.C.: Theory and research on teaching as dialogue. In: Richardson, V. (ed.): Handbook of research on teaching, pp. 1102-1121. AERA, Washington (2001)
3. Egan, K.: The Educated Mind, how cognitive tools shape our understanding. University of Chicago Press, Chicago (1997)
4. Koschmann, T.: Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration. In: Hoadley, C. and Roschelle, J. (eds.): In Proc. Computer Support for Collaborative Learning (CSCL'99). Mahwah, Palo Alto (1999)
5. Wegerif, R.: A dialogic understanding of the relationship between CSCL and teaching thinking skills. Computer-Supported Collaborative Learning, 1(1):143–157, (2006)

6. Bakhtin, M.M.: The dialogic imagination: Four essays. The University of Texas Press, Austin (1981)
7. Bakhtin, M.M.: Speech genres and other late essays. University of Texas, Austin (1986)
8. Bakhtin, M.M.: Problems of Dostoevsky's poetics. University of Minnesota Press, Minneapolis (1993)
9. Linell, P.: Rethinking language, mind, and world dialogically. Information Age Publications, Greenwich (2009)
10. Hicks, D.: Learning as a prosaic act. Mind, Culture, and Activity, 3(2):102-118, (1996)
11. Trausan-Matu, S. and Rebedea, T.: Polyphonic inter-animation of voices in VMT. In: Stahl, G. (ed.): Studying virtual math teams, pp. 451–473. Springer, New York (2009)
12. Stahl, G.: Group cognition. Computer support for building collaborative knowledge. MIT Press, Cambridge (2006)
13. Lorino, P.: Concevoir l'activité collective conjointe : l'enquête dialogique. Étude de cas sur la sécurité dans l'industrie du bâtiment. Activités, 6(1):87-110, (2009)
14. Dong, A.: The latent semantic approach to studying design team communication. Design Studies, 26(5):445-461, (2005)
15. Shotter, J.: Conversational realities. Constructing life through language. SAGE, Thousand Oaks (1993)
16. Cazden, C.B.: Vygotsky, Hymes, and Bakhtin: From word to utterance and voice. In: Forman, E.A., Minick, N. and Stone, C.A. (eds.): Contexts for learning: Sociocultural dynamics in children's development, pp. 197–212. Oxford University Press, Oxford (1993)
17. Leinhardt, G. and Greeno, J.G.: The cognitive skill of teaching. J. Educ. Psychol., 78(2):75-95, (1986)
18. Weinberger, A., Ertl, B., Fischer, F. and Mandl, H.: Epistemic and social scripts in computer-supported collaborative learning. Instructional Science, 33(1):1–30, (2005)
19. Trausan-Matu, S., Stahl, G. and Sarmiento, J.: Supporting polyphonic collaborative learning. E-service Journal, 6(1):58–74, (2007)
20. Wertsch, J.V.: Voices of the mind: A sociocultural approach to mediated action. Harvard University Press, Cambridge (1991)
21. Dysthe, O.: The multivoiced classroom: Interactions of writing and classroom discourse. Written Communication, 13(3):385–425, (1996)
22. Trausan-Matu, S., Dessus, P., Rebedea, T., Mandin, S., Villiot-Leclercq, E., Dascalu, M., Gartner, A., Chiru, C., Banica, D., Mihaila, D., Lemaire, B., Zampa, V. and Graziani, E.: Deliverable 5.2 – Learning support and feedback. OUNL, Heerlen (2010)
23. Trausan-Matu, S., Dessus, P., Lemaire, B., Mandin, S., Villiot-Leclercq, E., Rebedea, T., Chiru, C., Mihaila, D. and Zampa, V.: Deliverable D5.1 – Support and feedback design. OUNL, Research report of the LTfLL Project, Heerlen (2008)
24. Trausan-Matu, S. and Rebedea, T.: A polyphonic model and system for inter-animation analysis in chat conversations with multiple participants. In: Gelbukh, A. (ed.): Computational Linguistics and Intelligent Text Processing (CICLing 2010), pp. 354–363. Springer, New York (2010)
25. Villiot-Leclercq, E., Mandin, S., Dessus, P. and Zampa, V.: Helping students understand courses through written syntheses. An LSA-based online advisor In Proc. Tenth IEEE Int. Conf. on Advanced Learning Technologies (ICALT 2010), Sousse (Tunisia) (2010)