Deliverable 5.3 LTfLL – Learning support and feedback

Stefan Trausan-Matu, Philippe Dessus, Traian Rebedea, Mathieu Loiseau, Mihai Dascalu, Dan Mihaila, Isobel Braidman, Gillian Armitt, Alisdair Smithies, Maria Regan, et al.

Language Technologies for Lifelong Learning

LTfLL -2008-212578

Project Deliverable Report

Deliverable nr D 5.3

Work Package	5
Task	Learning Support and Feedback
Date of delivery	**Contractual**: 01-09-2010 **Actual**: 11-10-2010
Code name	**Version**: 4.6 Draft ☐ Final ☒
Type of deliverable	Report
Security (distribution level)	Public
Authors (Partner)	Stefan Trausan-Matu, Philippe Dessus, Traian Rebedea, Mathieu Loiseau, Mihai Dascalu, Dan Mihaila, Isobel Braidman, Gillian Armitt, Alisdair Smithies, Maria Regan, Benoît Lemaire, Jonathan Stahl, Emmanuelle Villiot-Leclercq, Virginie Zampa, Costin Chiru, Iulia Pasov, Andrei Dulceanu.
Contact Person	Stefan Trausan-Matu
WP/Task responsible	Stefan Trausan-Matu
EC Project Officer	Ms. M. Csap
Abstract (for dissemination)	This report presents Version 1.5 of the Learning support and feedback services (delivering recommendations based on interaction analysis and on students' textual production) that can be integrated within an e-learning environment.
Keywords List	Individual & Collaborative Knowledge Building, Social Network Analysis, Feedback, Free Texts, Latent Semantic Analysis, Bakhtin
Acknowledgements	The authors wish to thank W. Greller and J. Hensgens for providing useful comments on an earlier version of this Deliverable.

LTfLL Project Coordination at: Open University of the Netherlands

Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands

Tel: +31 45 5762624 – Fax: +31 45 5762800

Table of Contents

Executive Summary

This report presents a new version of the support and feedback services (delivering recommendations based on interaction analysis and on students' textual production) that can be integrated within an e-learning environment. The two services of WP5 (*PolyCAFe* and *Pensum*) propose to the learner and the teacher automated tools to extract relevant information from learning material (e.g., course texts, discussion sessions) in order to provide pieces of feedback that help them understand course content (learners) and take better instructional decisions (teachers).

Compared to the Version 1.0, the Version 1.5 of the services has been improved on many points, according to the lessons learnt from the last validation round and from reviewers' recommendations. It includes the importation of forum discussions and their enhanced analysis, the refinement of the evaluation and search modes (for *PolyCAFe*), feedback severity management and learners' control over feedback (for *Pensum*).

Version 1.5 is the final version of the "Learning Support & Feedback" services as a stand-alone application. A new version of the threaded services will be released together with the submission of LTfLL D2.4. Version 2.0 of the service will include additional thread functionalities in addition to the Version 1.5 functionalities.

The report contains the following parts: It starts with an introduction which sets up what is the goal of the two services, what is the state of the art, who will use the service and what value does it have for them. The second part sheds some light on the verification and validation process of the previous versions of the services. The third part describes the final model of the WP5 services, according to the validation results. The fourth part describes the software documentation and the fifth presents the next steps: verification/validation, some points about the integration of WP5.1 and WP5.2 in a thread with other LTfLL services, and the roadmap.

1. Introduction

One aim of the WP 5 of the LTfLL project is to design and develop services that facilitate life-long learners' and tutors' activity in delivering relevant feedback to learners' written production, either for free texts (e.g., essays, syntheses, course notes) or chat and forum conversations. In the first phase of the project, we presented in deliverable D5.1 an overview and a selection of existing models, methods and resources for: 1) the automated analysis of learner interactions using language technologies or social network analysis (Task 5.1) and 2) the automated analysis of learner text (Task 5.2). A first rough proposition of possible tools to be developed was eventually introduced. In the second phase of the project, we presented in D5.2 v. 1.0 of the support and feedback services to be integrated within an e-learning environment. Further steps toward the implementation of the next version of these services and their future integration with all the LTfLL services were also suggested. We described the educational use and the global design of both services with the use of Scenario-Based Design from D3.1 and D3.2. The purpose of this D5.3 deliverable is to present v. 1.5 of the same services, as well as the main validation results that led to the implementation of this version (fully described in D7.3).

Writing and reading play a crucial role in e-learning environments as a means to understand subject matter content and communicate with each other (i.e., peers, teachers, tutors). Our main assumption is that learners and teachers at a distance are immersed in a dialogical written environment (Linell, 2009), the latter itself inserted in a more general one, the e-learning platform. Every activity of the stakeholders of this twofold environment (i.e., learners, tutors, teachers, and even handbooks' authors) is intertwined with those of the others as textual utterances, or, as Bakhtin (Bakhtin, 1986) put it, 'inter-animating voices', yielding to interwoven "threads" of utterances, either read or written. These threads, by themselves, and their inter-animation carry their own level of affordances for learning, that is, lead to a given level of understanding and learning. Inner and outer dialogue (respectively through cognition and communication) is the way every learner builds meaning and therefore learns (Stahl, 2006). Moreover, we assume that written affordances can be strengthened, for the learners to cope with their understanding of the content taught (Dessus et al., 2010). In a nutshell, we claim that cognition is the cooperative appropriation of affordances (Reed, 1991). But the "meaning potential" of utterances (as Linell, 2009 renames utterances) is not merely understandable without tools: the first level of affordance they offer is complex to analyze. For instance, when summarizing a course to understand it, no student is able to thoroughly analyze inter-sentence coherence gaps in his or her production, or possible off-topic sentences; and no teacher is able to analyze the complex threads of utterances of a one-hour chat from a content point of view. Tools supporting these activities are thus needed.

The two tools we introduce below are both aimed at providing high-level feedback on learners' and teachers' productions, and thus at "expanding the space of learning" (Wegerif, 2007). They both use intensive Natural Language Processing (NLP) techniques to analyze some important textual features, all related to Bakhtin's theoretical view

(Dessus & Trausan-Matu, 2010). The reason why these tools can be useful in distance learning contexts (or more informal ones) is related to the fact that extracting relevant information from instructional material (e.g., course texts, learners textual productions) is a cognitively demanding and time-consuming task, either for learners or teachers. The two tools provide through feedback quick and efficient ways to analyze, extract and visualize this information, which becomes more understandable for teachers and learners.

1.1. Chat and Discussion Forums Analysis and Feedback-giving

Recent years showed an explosion of the usage of collaborative and social computer applications and platforms on the web. A normal consequence is the use of such tools, not only for general communication but also for informal and even formal learning. Moreover, Computer-Supported Collaborative Learning (CSCL) is not only a fashionable approach, it is also based on a totally different learning paradigm (Koschmann, 1999; Stahl, 2006), based on dialogism and the social-cultural ideas of Bakhtin (1984) and Vygotsky (1978), which appeared in the third and fourth decades of the 20th century, before the invention of the electronic computer.

Online discussion forums have been used as asynchronous collaboration tool in either formal or informal learning for some years. For example, at University of Manchester, medicine students have to discuss in a forum about some issues related to their profession. Professors manually analyze these discussions starting from the model of community of inquiry of Garrison et al. (2000). Because it is very difficult and time consuming to read and follow the discussion threads in detail, automated support is extremely beneficial.

Instant messenger (chat) started to be used recently for collaborative problem solving and other assignments (e.g., Stahl, 2006; Stahl, 2009; Trausan-Matu, 2010). CSCL chat analysis was performed until now manually by tutors, activity extremely time consuming and difficult, for each chat session the time needed for the analysis lasting more (even two times) than the actual duration of the online chat. In the analysis performed in the first year showcase of LTfLL, all tutors mentioned that it would be useful to have automated tools and indicated that the possibility of following and analyzing discussion threads is very important (Trausan-Matu, 2010).

An analysis of the state of the art shows that in the last years, several CSCL applications appeared for analyzing interactions in chats and forums. Some examples are CORDTRA (Hmelo-Silver, Chernobilsky, & Masto, 2006), COALA (Dowell & Gladisch, 2007; Dowell, Tscholl, Gladisch, & Asgari-Targhi, 2009), DIGALO and other tools used in the *Argunaut* system (Harrer, Hever, & Ziebarth, 2007), ColAT (Avouris, Fiotakis, Kahrimanis, & Margaritis, 2007), TATIANA, (Dyke, Lund, & Girardot, 2009), the Scaffold-Argument visualization (Law, Lu, Leng, Yuen, & Lai, 2008), KSV (Teplovs, 2008), VMT-Basilica (Kumar, Chaudhuri, Howley, & Rosé, 2009), Polyphony (Trausan-Matu, Rebedea, Dragan, & Alexandru, 2007) and *PolyCAFe* (Trausan-Matu & Rebedea, 2009, 2010). Some of them are using different kinds of argumentation graphs (Toulmin, 1958), more elaborated structures like the contingency graphs (Suthers, Dwyer, Medina, & Vatrapu, 2007) or polyphonic threads visualization in *PolyCAFe*. However, almost all the systems provide only facilities for manual annotation and input of links and of visualization. No system, except *PolyCAFe*, provides complex facilities for chat and forum discussions analysis and for providing useful feedback for learners and tutors.

In our opinion there are several causes that explain this situation. The first one is that even if Bakhtin's dialogism is considered a theoretical model of CSCL (Koschmann, 1999; Stahl, 2006), except our own system we don't know of any actual implementation based upon it. The second cause is related to the fact that the majority of collaborations in CSCL are based on the exchange of text messages. Thus, another problem arises from the fact that current Natural Language Processing (NLP) systems are far from providing reliable text understanding systems. Moreover, in CSCL chats and forums there are usually more than two participants, a case which is generally not considered in most NLP theories developed for conversation analysis (Trausan-Matu & Rebedea, 2010).

The *PolyCAFe* system (**P**olyphonic **C**onversation **A**nalysis and **F**eedback generation), was developed under WP5.1 considering the above mentioned problems (see also deliverable D5.2). It provides abstraction and feedback-giving services for supporting both learners and tutors that are involved in assignments that make use of chat or forum conversations. The services are packed into web widgets that can be easily integrated into any LMS, PLE, VLE or other web application.

1.2. Synthesis Writing and Feedback

In a lot of educational contexts, the so-called writing-to-learn pedagogical approach is used (Klein, 1999). Learners have to produce textual reports (e.g., summaries, essays, syntheses) about the lessons they have learnt, and some pieces of feedback can be offered about their results (see Appendix 3 for more details). The time-consuming nature of feedback on synthesis produces long waiting times and the limited time resources for support imply that most feedback does not exceed the shallow levels of writing (e.g., spelling) and higher levels are seldom addressed automatically. Since delivering high-level feedback requires lot of time for teachers, learners encounter long waiting time for feedback and the limited feedback opportunities force them to hand in mainly completed versions. On the other side, teachers have only a limited overview of the learners' progress and may find out specific problems too late to use them during the current course.

We developed *Pensum*, a web service that supports learners in the automated assessment of their essays in order to let teachers focus on higher-level activities. *Pensum* analyses how well learners understand course texts as shown by their textual productions, and provides just-in-time feedback on the ongoing writing activities. In every course attended by learners, two main pieces of texts are considered: the source text, which is composed of texts proposed by the teacher (e.g., course notes, readings), and the notes taken by the learner in order to highlight the main ideas to understand (i.e., a synthesis). *Pensum*'s architecture is organized around these two textual levels and provides three main pieces of feedback automatically. Two pieces of feedback are focused on the synthesis (whether two contiguous sentences are coherent with each other and whether the semantic content of a given sentence is close to some sentences of the source text), while one other piece is related to the source text (whether a given sentence is semantically related to some sentences of the learner's synthesis, see § 2.3 of D5.2 for more details). This semantic triangulation allows the learner to visually grasp what is at hand in the content of the course texts and what remains to be understood.

This web service offers some functionalities seldom encountered in the market (see D8.5 section 13 for more details). To our knowledge, there are very few written-based feedback advisors in the market and almost none of them are available open source—spell and grammar checkers, as well as databases for courses and revision management websites put aside. We can list them in the following four categories (see D8.5 § 13): *shallow assessment*, with software like *PaperRater* or *Turnitin* (see Lukashenko, Graudina, & Grundspenkis, 2007 for a review), which provide various lexical and style-oriented advice but are mainly at lexical level (word usage, grammar, mechanics, readability formula, plagiarism detection); *organization strategies prompters*, like *WriteThis* or *QuickStory*, which do not check students' essays from a content basis, but prompt them advice to organize or revise their writing (Rowley & Meyer, 2003); *collective course* revision: which use a special blending of web 2.0 collaborative functionalities (wiki or Twitter accounts); *semantic-based assessment*, which provide feedback on students' free texts—like *Intelligent Essay Assessor* from Pearson or *Criterion* from ETS—(Burstein, Chodorow, & Leacock, 2004; Landauer, Laham, & Foltz, 2000; Warschauer & Ware, 2006) on a semantic basis, using NLP techniques like Latent Semantic Analysis to compare students' essays against pre-assessed ones. *Pensum* pertains to the latter category, but also lets students be aware of their writing strategies (second category), so we can claim the functionalities *Pensum* provides are unique.

The strategies a learner using *Pensum* can use are, firstly, to understand better the *sources* of what he or she has understood at a semantic level. Secondly, the teacher can reread the learner's synthesis through *Pensum*, validating or not the delivered pieces of feedback, giving more structured or higher-level pieces of advice. The flexibility of *Pensum* lies in the fact that it prompts no specific strategy (related to writing or reading) and is thus adapted to a large range of learning strategies. Eventually, the new functionalities added in Version 1.5 of *Pensum* enable the learners to act upon the delivered feedback and question it explicitly if it is judged as inadequate.

The two WP5 web services are complementary: they both provide visual feedback on pieces of textual artifacts witnessing learning. On one side *PolyCAFe* provides feedback from a collective and social viewpoint based on the written utterances of the latter; while *Pensum* is focused on individual pieces of writing, based on the learners' written notes. This complementarity will be exploited in an integrated thread involving several services provided in the WP4, WP5 and WP6 of the LTfLL project.

The remainder of this report is as follows. The second part sheds some light on the verification and validation process of the previous versions of the services (Version 1.0). The third part describes the final model of the WP5 services, according to the validation results. The fourth part describes the software documentation since the fifth opens up new likely research directions (next validation processes, long thread, and roadmap).

2. Verification and Validation of the Services in Version 1.0

This section presents the main issues identified in the second round validation process (involving the versions 1.0 of our services), fully detailed in the D 7.3. These issues were, together to reviewers' feedback, main points considered for the improvement of the Version 1.5 of *PolyCAFe* and *Pensum*.

2.1. *PolyCAFe* v. 1.0 Verification and Validation Processes

PolyCAFe 1.0 has been validated using 9 senior (4th year) students, 5 tutors and the teacher of the Human-Computer Interaction course at the Computer Science Department of Politehnica University of Bucharest, where students had to debate using chat on a given topic. The participation to the study was an optional activity, after the course had been finished, without any influences on students' grade. All the students have been at the first usage of *PolyCAFe*, and only two of the tutors had used it prior to the validation experiment.

The results of verification and validation of *PolyCAFe* were very encouraging (see deliverable D7.3 for more details). They highlighted that *PolyCAFe* has an important potential for tutors and students as well. However, several issues were identified and they needed to be solved in order to increase the relevance for students and fix the misleads in version 1.5:

- Update the process of grading the utterances and participants in order to fix the misleads identified by the students. Thus, they shall have more trust in the results of the system.
- Improve the usability of the widgets by using tool-tips and provide help sections that explain how the results should be interpreted. This would solve one of the greatest problems for many students: the difficulty to interpret results because of their lack of experience in analyzing chat conversations.
- Provide extra feedback that should be more relevant for the students: the list of concepts that are similar to the ones in the discussion, but that did not appear in it (zone of proximal development for each student and for the whole conversation with regard to the semantic space).

2.2. *Pensum* v. 1.0 Verification and Validation Processes

The previous verification/validation round was threefold (see D5.2 and D7.3 for more information):

- *A pre-validation* of two computational methods for selecting important sentences (see D5.2 § 2.3): a human (involving 30 participants) vs. computer study of key-words selection and the use of CI-LSA model (Lemaire, Denhière, Bellissens, & Jhean-Larose, 2006). None of these methods has led to sufficiently valid results and we rejected their possible use in *Pensum*. Thus a simple LSA-based processing is used to diagnose the semantic relation between a sentence of the synthesis and one of the course.

- *A verification round* (see Appendix B.5 of D7.3), involving 5 users, which led to revise some interface issues (saving problems, hypertext links replaced by buttons, more readable presentation format) as well as pointed to important processing time.
- *A validation round* (see D7.3 § 3.3.4.4) in which 11 students in educational sciences and linguistics were involved (Master 1st and 2nd years), from an initial population of 23 students. They used *Pensum* v. 1.0 in authentic settings. They had a mixed opinion on *Pensum*'s use. They think that *Pensum* gives feedback and guiding different from humans but also think that *Pensum* is rather easy to use and that errors are easy to recover. Their cognitive load engaged in using *Pensum* has been measured as normal, except for the frustration level. Overall, students have a better opinion on the usability of *Pensum* than on its effects on learning or its pedagogical capabilities. Additionally, a teacher in the domain of distance learning was shown the widgetized version of *Pensum* 1.0.

Additionally to these rounds, we also took advantage from the recommendations of the second year review report, which stated "there are challenges in relation to the nature of the writing activity involved and cultural assumptions in relation what constitutes quality in a synthesis" (§ 1.3). As stated in the D7.3 § 3.3.4.4, the main lines of improvement concerning *Pensum* are mostly related to a better *feedback control* from learners, enabling them to give their own opinion on the feedback delivered by *Pensum* as well as to finely tune the feedback severity. Another line of improvement concerns feedback format, in allowing learners to add their own course documents. The first line has been added to the current stand-alone version of *Pensum* (v. 1.5), while the second improvement remains to be added in the widget version still to come (long thread integration).

3. Final Model: The Working Principles

Our services are developed for students and teachers/tutors engaged in a distance learning situation. They can read texts, write out notes, essays, or discuss such or such topic. All the stakeholders performing these activities write or speak in natural language, producing "utterances" that can become "voices" influencing the following utterances (Trausan-Matu & Rebedea, 2009), 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. We assume in this report that these contexts can be analyzed with some Natural Language Processing techniques uncovering the semantic relations between these "voices".

The overall working principle we used in the two tools is rooted on Bakhtin's view (Dessus & Trausan-Matu, 2010). 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 (Trausan-Matu & Rebedea, 2009). These voices should be understood in a larger extent than the acoustic sense. 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 (Stahl, 2006) — which implies explicit or implicit inner dialogs among learner's voice and the other voices (e.g., author's, professor's, commenters' voices).

3.1. The Polyphonic Model of *PolyCAFe*

PolyCAFe is based on the polyphonic dialogical model of Mikhail Bakhtin (Bakhtin, 1981, 1993) applied to CSCL using chats (Trausan-Matu & Rebedea, 2009) and it is implemented using NLP tools for discourse and content analysis, as described in detail in deliverables D5.1 and D5.2. From a pedagogical perspective we share the idea that learning is not a transfer of knowledge and it rather may be seen as "becoming a participant in a certain discourse" (Sfard, 2000).

From the NLP point of view, the *PolyCAFe* polyphonic model is materialized in an analysis of multithreaded discourse with multiple participants, each of them having an assigned role (Trausan-Matu & Rebedea, 2010). This contrasts with the typical discourse analysis (Jurafsky & Martin, 2009), which has a rather local scope and is based on a two-participant model, like in phone and face-to-face dialogs where only one person usually speaks at a given moment in time. Even if other systems analyze conversations with multiple participants, considering a more global, collaboration-based perspective, like transacts (Joshi & Rose, 2007), they are still hooked on a two-participant model.

In *PolyCAFe* the discourse analysis focuses on identifying voices in the analysis of discourse in chats and their inter-animation (voices are, in our view, important utterances with an important impact on the conversation as a whole, utterances giving birth to threads). For this aim, we analyze the links (adjacency pairs, repetitions, lexical chains and argumentation links) between utterances, we construct a graph, we identify threads and we compute for each utterance a value reflecting its importance, given not only by its content but also based on the 'echoes' of other previous utterances present in it and on the echoes generated by it and detected in other further utterances.

PolyCAFe is using similar techniques with other systems that analyze chats, trying to identify speech acts (Jurafsky & Martin, 2009), dialog acts (Kontostathis et al., 2009), adjacency pairs (Jurafsky & Martin, 2009), topics and links (Adams & Martell, 2008), or other complex relations (Rosé et al., 2008). This is also the case of techniques as TF-IDF (Adams & Martell, 2008; Schmidt & Stone, 1993), Latent Semantic Analysis (Dong, 2006; Manning & Schütze, 1999; Schmidt & Stone, 1993), Social Network Analysis (Dong, 2006). The lexical ontology *WordNet* (wordnet.princeton.edu) (Adams & Martell, 2008; Dong, 2006) is often used.

3.2. The Cognitive Model of *Pensum*

People either learn collaboratively or individually, and *Pensum* is dedicated to the latter. One learns to write (and thus learns a content) only when practicing writing. The more learners practice writing activities, the more efficient their writing, and the deeper their understanding of what they wrote about become. Even though learners, to some extent, can assess by themselves what they learned, feedback remains necessary to help them develop their writing and learning skills, as additional information. Computers and teachers can both give such feedback.

Pensum supports a cognitive model already described in Section 2.3 of D5.2: this model has not been subject to important modifications so far. Briefly put, it relies on the important capability of Latent Semantic Analysis (Landauer & Dumais, 1997) to account for semantic relations between pieces of texts. The cognitive model behind LSA has been extensively tested in the literature, and solid evidences have been published: LSA is able to detect the likely coherence gap between two sentences and to account for the semantic relations between two distinct paragraphs (i.e., relevance and topic coverage). We assume these pieces of feedback are at the core of the understanding process: writing summaries or syntheses can improve the process of understanding (Thiede & Anderson, 2003), and the latter is evidenced in highly coherent writing material (Tapiero, 2007). Moreover, the lack of coherence of a given summary or synthesis is better detected when the learner is given cues on the extent to which the text is pertinent (the summary or synthesis do not contain semantically unrelated content to the source text) and covers the intended topic well (the main ideas of the source text are semantically related to the summary or synthesis). The learner can go back and forth among these pieces of feedback and revise the summary or synthesis accordingly. *Pensum*'s functionalities are strongly organized around the notion of *self-regulated learning* (Boekaerts, 1997; Lajoie & Azevedo, 2006; Winne & Hadwin, 1998). Since lifelong learning entails long periods of work without any direct interventions of teachers, the self-regulated learners have to rely on other sources of feedback (e.g., computer-based, peer-based) to understand the content read or taught, to monitor or control their own cognitive processes and thus to make progresses.

4. Final Software Documentation: Widgets, Services, and Data

4.1. Pedagogical Orientation

In distance learning contexts, lesson and lecture content, often "evaporates" as soon as the learners have left the classroom or the distance learning course (Hübner, Nückles, & Renkl, 2006), because they lack of opportunities to monitor their understanding of the course, and to detect possible misunderstandings. Chat conversation and free text writing may be two occasions to give learners complex activities that let them build knowledge instead of rote information.

Chat and forum conversations, due to their popularity, are now used also for teaching purposes, in a new approach, Computer Supported Collaborative Learning—CSCL, see Stahl (2006), becoming an alternative or supplement to classical learning, like for small virtual groups using chat systems for learning together (Koschmann, 1999; Stahl, 2006). CSCL is a change of vision on learning replacing the idea of the transfer of knowledge from a human or a written source to the student.

A teacher giving a course and assigning the learners the task to write out notes on what they have understood from it is a very frequent setting of a free text writing activity. This task is too rarely followed by precise feedback from the teacher on what has been understood. The underlying idea of our service is that learners proceed with this task with the help of *Pensum*, which can deliver just-in-time feedback on some semantic relations at stake between given parts of the course and the learners' writing.

4.2. Overview of the Systems

The *PolyCAFe* system for the automated analysis of CSCL chat and forum conversations considers three categories of data in chats and forums: the expected content of students' utterances, the degree of collaboration (including involvement, acceptance of others, leader/lurker role, interaction or inter-animation) and the existence of discourse and debate threads. Flexible interfacing is provided in the form of a set of seven widgets (Rebedea et al., 2010) for managing the assignments, for uploading chat logs and for providing interactive graphical and textual feedback for each of the students and for the group as a whole.

For detecting if chats and forums include the correct content, NLP techniques are used. The NLP pipe (see Figure 1) preprocesses the interaction logs for both content and inter-animation analysis (Trausan-Matu & Rebedea, 2010). In order to detect inter-animation, threads are identified starting from the explicit links indicated by the student and from implicit links, detected using also NLP techniques: speech acts, adjacency pairs, co-references, cue-phrases or starting from repeated words and LSA (Trausan-Matu & Rebedea, 2010). Several criteria are used for detecting inter-animation patterns, such as the presence of questions, agreement, disagreement or explicit and implicit referencing.

In addition, the strength of a voice (of an utterance) is computed starting from the strength of the utterances that refer to it. If an utterance is referenced by other utterances that are considered important, obviously that utterance also becomes important (Trausan-Matu & Rebedea, 2010). More details about the design and implementation of *PolyCAFe* are provided in deliverable D5.2.

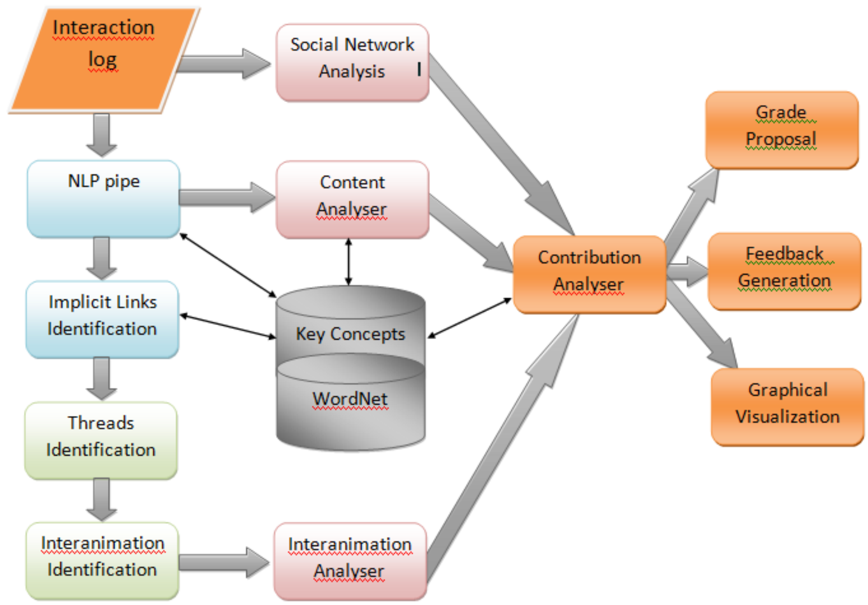

Figure 1 – The architecture of *PolyCAFe* analysis system.

Pensum v. 1.5 revolves around a classical repository architecture (Garlan & Shaw, 1993), the blackboard role being taken by a database, which stores all information relevant to the user: the courses to synthesize, the syntheses produced, user information, notes, feedback and user action on feedback. Figure 2 represents the described architecture through a data flow diagram.

A display module gathers from the database the teaching material and synthesis work associated to the user (learner for v. 1.5), in order to present it to them. This module also allows to retrieve the feedback returned by the system and to present the user with the sum of his/her past interactions with *Pensum*. The database now stores most of the successive states of the system and is able to take into account older interactions to display their influence in the current state of the synthesis.

To interact with the system, the user has two main means: editing the synthesis and exerting control over feedback. *Edition of the synthesis* is done through a text field and the data sent through very basic NLP routines to separate the text into sentences, to store them in the database. It is then retrieved by an R-LSA based web service provided by WP 2, which performs cosine measurement of the distance between the sentences of the teaching material and of the synthesis. The data is processed with previous feedback to generate the current state feedback (see § 4.3 Changes in Version 1.5 for mechanisms).

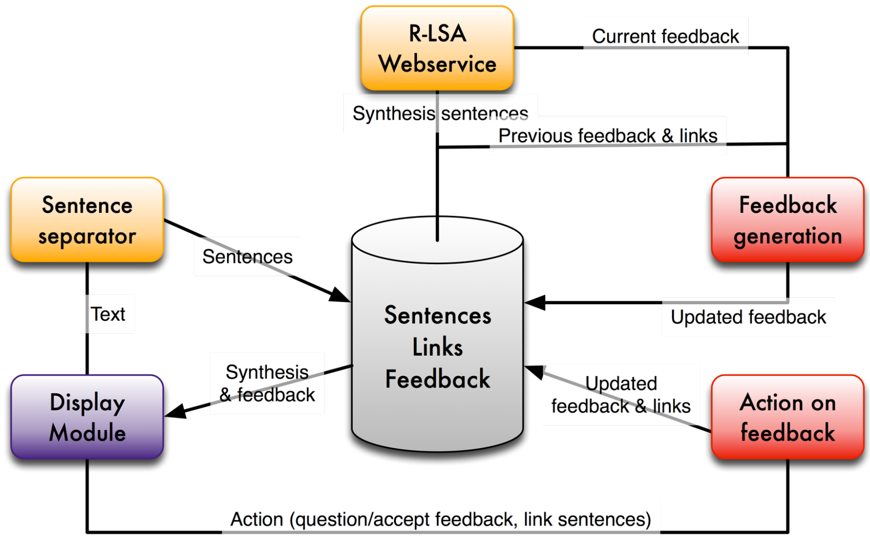

Figure 2 - *Pensum* repository architecture.

As suggested by the above, the other mean of interaction with the data at the user's hand is *to act upon feedback*. He can choose to toggle the severity of the feedback should the space used for LSA computation not be evaluated strictly or loosely enough by the feedback procedure. Or he/she can choose to directly question a feedback item, which he can justify by linking a sentence from the teaching material to a sentence from the text. Beyond the current version, the development of a full tutor/administrator will give access to tutors to the detail of the decisions and actions performed by the learner, and to judge them from a pedagogical viewpoint, and then propose more appropriate system settings (severity, semantic links).

4.3. Changes in Version 1.5

Changes in PolyCAFe

The new version of *PolyCAFe* contains improvements on several directions, as a result of the validation activities (see D7.3) and of the reviewers' feedback. The analysis of discussion forums was improved, including also the Garrison model of the community of inquiry (Garrison et al., 2000), used by the Medicine Department of the Manchester University. Medicine was introduced as a new domain of analysis in addition to computer science (in order to analyze the discussion forums at Manchester University). The analysis of the utterances and the content and form of the feedback were improved and some steps towards including Romanian as a second language were done (a POS Tagger for Romanian was developed). Meanwhile, an evaluation of the performance of the speech act annotation was performed studying both the machine learning approach and the linguistic rules approach using the special PatternSearch module described in D5.2. Table 1 summarizes the changes in T5.1 services. In the following sections details will be provided.

Table 1 – Overview of the Changes of T 5.1-related Services.

Service	Deliverable Description	Activity Loops	Main Features	Main changes compared to the previous version
Polyphony + ChAMP	D5.1	Chat analysis and preliminary feedback	Desktop-based applications for providing analysis tools for tutors grading chat conversations	N/A
PolyCAFe 1.0	D5.2	Chat analysis and feedback; preliminary discussion forum analysis	Web-based widgets for providing feedback to tutors and students participating in chat conversations and, partially, in discussion forums	Technology: - web widgets for feedback and visualization Feedback on 5 levels: - conversation feedback - conversation visualization - utterance feedback - participant feedback - improved semantic search
PolyCAFe 1.5	D5.3	Chat and discussion forum analysis and feedback	Web-based widgets for providing feedback to tutors and students participating in chat conversations and discussion forums	Technology: - adapted the LSA input to accept semantic spaces trained using the R-LSA service described in D2.3 - discussion forums import module - full analysis of medical discussion forums Feedback: - still on 5 levels - new feedback for discussion forums related to the classes of the model of inquiry - improvements of the evaluation and search

Discussion forums

A major part of the effort in the new version of *PolyCAFe* was directed to the improvement of the analysis of the discussion forums. First of all, the system was extended in order to accept discussion forums in the format provided by the University of Manchester (the first version of *PolyCAFe* accepted discussion forums for analysis in a restricted format). This format is exported from Blackboard (the learning environment used in Manchester) to a CSV (comma separated values) input file that stores a discussion thread from the forum. Secondly, the "Community of Inquiry" model of Garrison, Anderson and Archer (2000), used at the University of Manchester for analyzing forums was also introduced in the analysis.

For the forum analysis and feedback, a similar approach with that for chat conversations is used. It was taken into consideration that *PolyCAFe* is intended to be used, at least its current version, only for discussion threads that involve small groups of up to 15-20 students. This was a requirement for the University of Manchester forums where the students are split into teams of 10-20 students that discuss on a given topic (e.g. safe prescribing, professional behavior in Medicine). Three major changes appeared in the system in order to provide feedback for the students involved in the discussion forums: 1) the communication with WP2's LSA framework in order to use the same latent semantic spaces that are trained in R for Medicine; 2) providing additional feedback at the utterance level in order to label each message in the forum with the corresponding categories in the Community of Inquiry model; 3) provide a list of concepts that are semantically similar to the ones present in the conversation, but that did not appear in it.

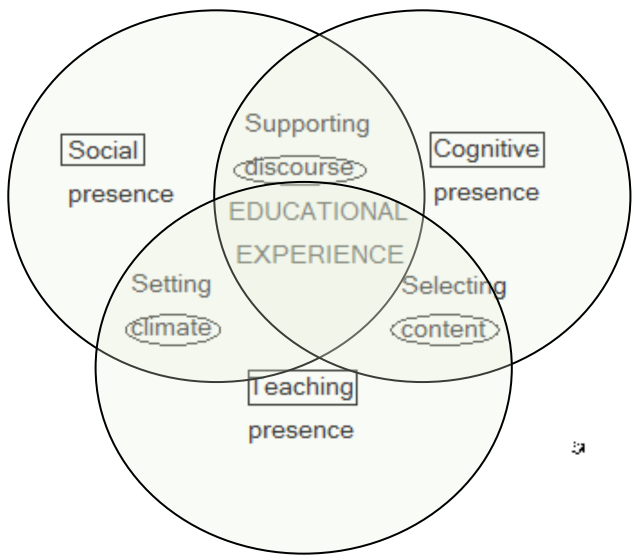

Figure 3 — The Community of Inquiry model from Garrison and Anderson (2003).

As shown in Garrison, Anderson and Archer (2000), a worthwhile education is embedded within a "Community of Inquiry". The educational process has two key players (participants): teachers (tutors) and students. As it is shown in Figure 3 the "Community of Inquiry" model considers that the learning occurs within the community based on three main elements: "social presence", "cognitive presence", "teaching presence"

Garrison, Anderson and Archer (2000) describe each element as a main structure in learning process and created for each element some categories that can appear in discussions (see also Table 5 in Appendix 1). For Cognitive presence they described as categories: triggering event, exploration, integration and resolution. In each message only one of these categories could appear.

Social presence has emotional expression, open communication and group cohesion. These categories can appear many times in a message. Teaching presence is including instructional management, building understanding and direct instruction. These categories can appear only one of them per message.

In the *PolyCAFe* analysis of forums the three main elements of the model of Garrison et al. are searched and investigated for providing a personalized feedback. The messages in the forum are processed and metadata about the three special presences (social, cognitive and teaching) are added. For this analysis, the PatternSearch module (see deliverable D5.2) looks for specific cue phrases described in Appendix 1. For example, the Group Cohesion category of the Social Presence may identified by patterns like "let's talk/do/try …", "who can …", the usage of pronouns like "us", "we", "our", etc.

The rule-based approach was tested on a corpus of about 400 messages from the University of Manchester. The results were 75% for Social Presence, 69% for Cognitive presence and 58% for Teaching Presence. A machine learning approach based on a bag-of-words model was also experimented with several techniques and the results were very good (see next section and Appendix 6).

An example of visualization and analysis of a discussion thread extracted from a forum discussing professional behavior in medicine is presented in Figure 4 and Figure 5. The visualization is similar to that of chats (as described in D5.2 and the later improvements). In Figure 4 we can see the interventions (postings) of each student (as small rectangles on the line associated to the student) and the links among them. The two different colors of the links are reflecting the explicit/implicit distinction: The explicit links are reflecting the "reply to" relation and the implicit ones are detected by the system, using semantic tools for detecting important concepts (LSA, keywords and synonyms). Under the part describing each student's contribution and the links among posts, similarly with the chat case, the collaboration graph is included—this being an improvement of the version presented in D5.2, see Trausan-Matu et al. (2010).

In Figure 5 is presented a conversation visualization widget for the same forum in which the user has selected to see the threads of the words "conduct" and "professionalism", which were two of the main topics of discussion.

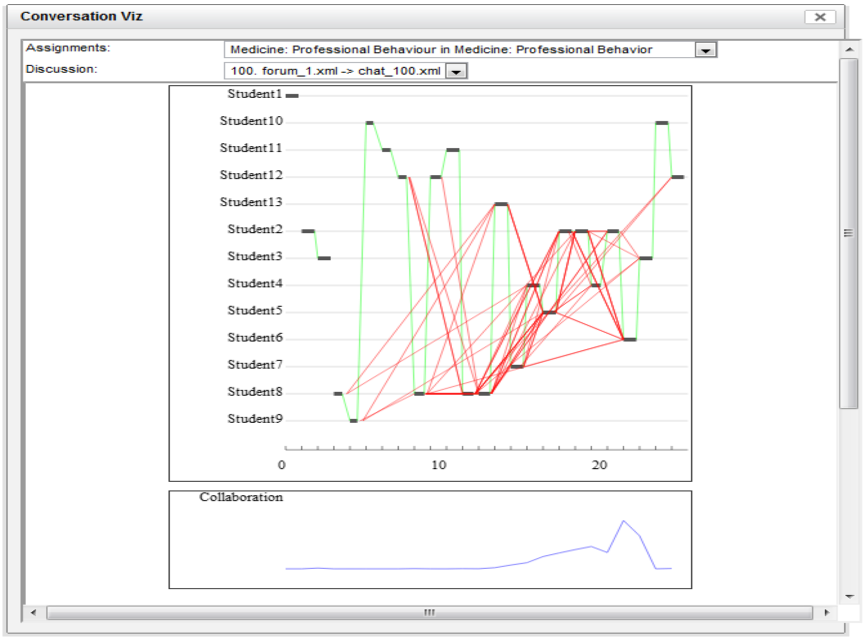

Figure 4 —Visualization of the discussion threads in a discussion forum about professional behavior in medicine (explicit links are green and implicit ones red).

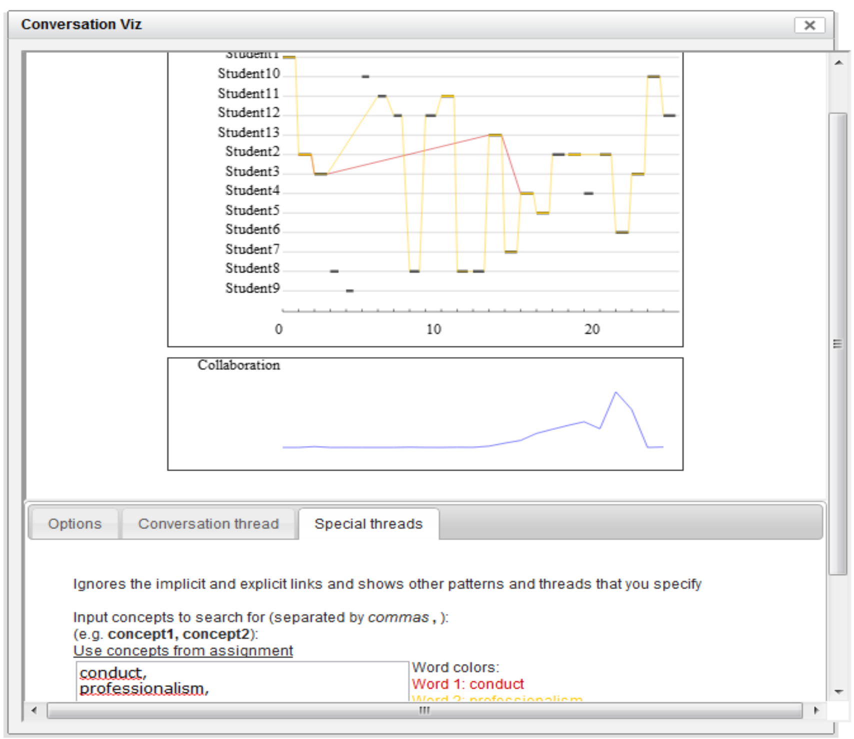

Figure 5. Visualization of specific threads of words

An example of feedback for forums including categories of the community of inquiry model is presented in Figure 6. The results of the annotation are included in the "Inquiry class" column.

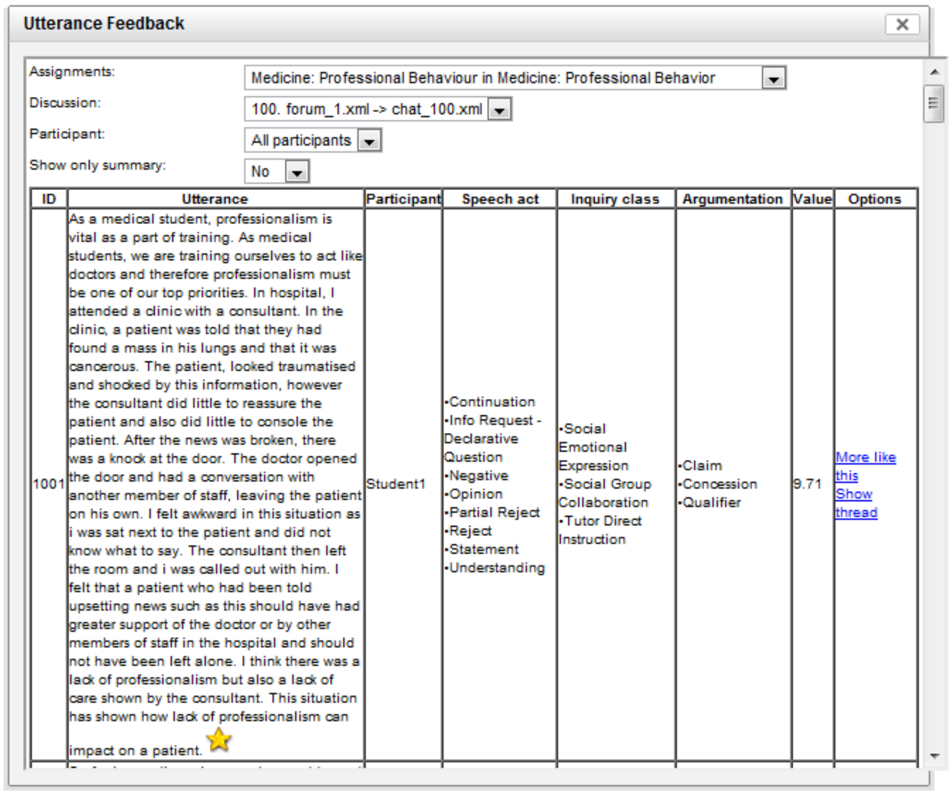

Figure 6. Model of inquiry feedback

Evaluation of the Performance of Speech Acts and Other Annotations

Several verification experiments have been performed for determining the accuracy of different types of annotations performed by *PolyCAFe.*

Speech acts identification: Two alternatives where tested for this task: a rule based-approach and a machine learning approach using TagHelper. The rule-based one is used by *PolyCAFe* by using a special linguistic pattern search module that was implemented in version 1.0 and described in details in D5.2. The machine learning approach using the *TagHelper* tools is based on the *Weka* machine learning framework. The verification of the speech acts detection using linguistic rules was made on a corpus of two annotated chat conversations, consisting of more than 700 utterances that were manually labeled by a tutor using a special tool that was developed for the verification of the automated feedback offered by *PolyCAFe* (Figure 7). For most speech acts, the precision is high, although the recall is limited that shows that the feedback is rather correct, although it is not able to detect all the utterances that contain a certain speech act (Table 2).

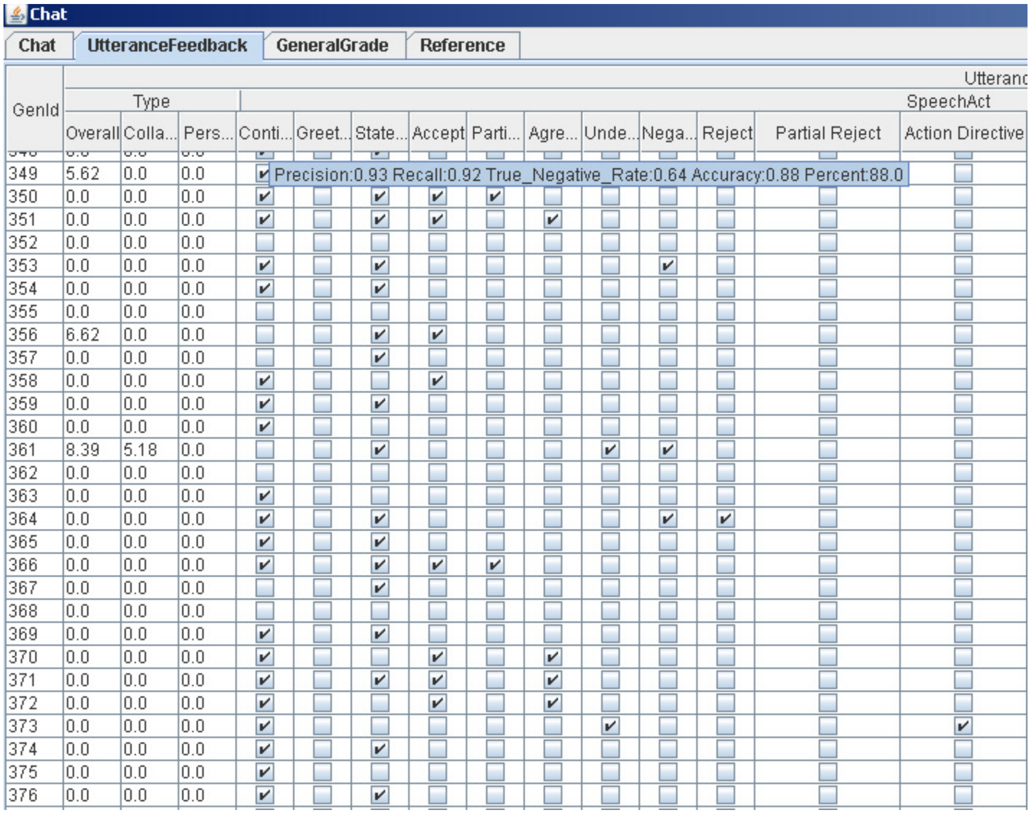

Figure 7. Tool for verification of the feedback

Some examples of linguistic rules (each rule can be saved in a variable and used to create more complex rules) are:

```
- "salut" | "hello" | "hi" | "halo" | "bye" | "goodbye" | "see you" | "c ya" | (<j> ("morning" | "afternoon" | "evening") )

- "maybe" | "may" <v> | "might" <v> | "should" <v> | "would" <v> | <D "believe"> | <D "suggest"
```

The machine learning latter offered results between 21-100% for precision and between 6-73% for recall depending on the speech act class (Dulceanu, 2010). For example, accept speech acts had a precision of 46% and a recall of 64%, while the reject class had only 35% precision and 31% recall. For more details, consult Appendix 6. A comparison of the two approaches shows that the results for speech acts identification using the linguistic rules are slightly better than using machine learning. A viable option to combine the results in the future.

Table 2. Precision and recall for speech acts detection using linguistic rules

Speech act - label	Precision	Recall
Continuation	93%	92%
Statement	94%	93%
Greeting	100%	80%
Accept	92%	80%
Partial accept	71%	55%
Agreement	90%	51%
Understanding	96%	58%
Negative	97%	78%
Reject	73%	82%
Partial reject	35%	27%
Action directive	75%	90%
Info request	100%	71%
Thanks	100%	100%
Maybe	100%	69%
Conventional	66%	50%
Personal opinion	100%	36%
Sorry	66%	75%

Model of inquiry classification: Similar to the experiment for speech acts, two methods have been investigated that were tested on an annotated corpus of discussion threads provided by the University of Manchester. However, due to the small number of annotated messages (below 400), the results may be over-fit (especially for statistical machine learning). The first experiment used the *PatternSearch*-based lexical rules and achieved a precision of 75% for social presence, 69% for cognitive presence and just 58% for teaching presence. The second method was based on the *WEKA* software, testing various alternatives (Bayes, SVM, decision trees and decision tables). Although the best results are obtained for over-fitted decision tables, a Bayes classifier that is less likely to over-fit obtained a precision of only 33% for social presence, 75% for cognitive presence and a good result of 84% for teaching presence. Therefore, the best alternative seems to be to use the rule-based method for social and cognitive presence and the machine learning one for teaching presence.

POS tagging for chat conversations: The experiment involved POS tagging chat conversations in Romanian with a POS tagger trained on a standard corpus for Romanian (George Orwell's *1984* book) that was then improved by manually annotating 5 chats and training the Markov model on these conversations. It has been shown that this resulted into an improvement of 10% for precision (from 66% to 73%) and 6.7% for recall (Chiru, Rebedea, & Ionita, 2010). More details are presented in Appendix 6.

Improvement of the Evaluation

The evaluation of chats and forums has been improved by taking into consideration the results provided by the validation of version 1.0 of *PolyCAFe* (see D7.3). Thus, the utterance grading process and the overall chat evaluation mechanisms have been modified, enabling a deeper insight into the actual discussion threads.

The first aspect that was improved focuses on the technical and the functional architectures behind the *PolyCAFe* system. Better modularization with loose coupling between different modules, higher internal cohesion with enriched functionalities, multilayered analysis focused on different layers of linguistics subdomains are some of the aspects that were taken into consideration. LSA used for evaluating chats was improved by means of POS tagging and participant segmentation. LSA plays a central role in the semantic evaluation of utterances with regards to the overall discussion, on-topic relevance and is used for quantifying the echo attenuation and the voices' impact between interlinked utterances. Also, a generalized utterance graph was built upon implicit and explicit links. This offers better understanding of chat discussions and enables deeper utterance evaluation because of the social perspective taken into consideration.

On another hand, the actual grading process was enhanced being oriented on a multilayered analysis with regards to quantitative, qualitative and social dimensions. Besides the previous improvements, new features regarding collaboration assessment, extractive summarization and semantic search were added to the *PolyCAFe* system. Collaboration is evaluated starting from Bakhtin's polyphonic theory and provides the basis for identifying intense collaboration zones and voice inter-twining in the analyzed discussion.

A complete description of the evaluation process is provided in Appendix 2, which provides information and formulas about all the indicators that are computed and used in the feedback delivery and grading process.

Improvement of Visualization

The visualization was adapted in order to be able to display information for discussion threads with a limited number of participants (below 15 per thread) as those used at the University of Manchester. Moreover, the special threads view mode has been improved to be able to display threads not only for given concepts, but also for more complex expressions as requested for the medicine validation scenario for *PolyCAFe* v. 1.5. A graphical representation of the collaboration below the diagram of participants *x* utterances was included (see Figure 4 and Trausan-Matu et al., 2010). The *x* axis of this graph is common with that of the above graphical representation of the conversation, showing the timeline of the conversation.

Major Changes in Pensum

Since our aim is to foster self-regulated learning processes, the possibilities for the learner to control the feedback offered is of great interest, since it enables the learner to *evaluate* the validity of the delivered feedback (see D 5.1 § 2.3.2) and not only to take the feedback for granted. *Pensum* is used as a cognitive tool insofar it provides the learners who are engaged in reading or writing with feedback, which in turn can be finely controlled by the latter (e.g., in the same way a novel word form that can be added to a spellchecker dictionary). For instance, if a sentence of the synthesis is prompted as "off-topic", the learner can either accept the feedback, tune its severity, or even declare a link between the synthesis and the course text as an evidence for a semantic relation. It is worth noting that this relation is used as input only for the learner his- or herself and not transferred for the other users (see section 5.4). Table 3 offers an overview of the main features of software implemented so far during the LTfLL Project, as well as the most important changes with regard to the previous version.

Table 3. Overview of the main Changes of T 5.2-related Services (PensumPensum).

Service	Deliverable Description	Activity Loops	Main Features	Main changes compared to the previous version
Apex 2.0 showcase	D 5.1 § 4.2	Course reading and summary Writing	Single course text read compared to learner summary. Judgment of understanding.	N/A
Pensum 1.0	D 5.2 § 2.3	Course reading and Synthesis writing	Multiple course texts compared to learner synthesis.	Judgment of understanding removed. Coherence assessment and relevance in synthesis. Topic coverage in course texts. Synthesis writing, multiple source text comparison.
Pensum 1.5	D 5.3 § 3.2	Course reading and synthesis writing	Multiple course texts compared to learner synthesis.	Software architecture Severity management Learner judgment on feedback

A system on which a learner can act makes its functioning more transparent, thus fostering critical understanding of the provided feedback. This principle is in accordance with formative assessment (Sadler, 1989), which lets the learner be one of the protagonist of his or her assessment process in prompting hints rather than definite judgements. The transparency of the functioning also permits the stakeholders to use the software for its actual capabilities, not for its imagined ones, thus improving their evaluation of the software (Murray & Barnes, 1998), but also making better use of the system (Bax, 2003). This principle, redefined in line with the previous reports, the verification/validation processes as well as the reviewers' recommendations, is at the core of many of the improvements made for *Pensum* v 1.5.

Control over Feedback

In order to make the user aware of the role of *Pensum*, we thought it relevant to grant them *Control over feedback*: each of the pieces of feedback can be invalidated by the learner, thus acknowledging the possible fallibility or indicative nature of the feedback. The user can justify his/her choices by linking explicitly a sentence from the learning material and one from the synthesis. The actions of the user are stored in the database, if a user has questioned a feedback item, it will remained questioned until he/she decides to change his/her mind. When justifying the questioning of a piece of feedback in the course (resp. synthesis), the user can link the sentence with any sentence of the synthesis (resp. feedback). Should any modification in the synthesis trigger previously inexistent feedback concerning on the synthesis sentence, it will directly appear as questioned. Additionally, the storage of successive versions of the synthesis, the corresponding feedback and their status can be very valuable information for the later improvement of the system or the analysis of the writing process. Questioned feedback could be an especially important information to help the tutor focus on higher level feedback: one can suspect that when the learner acknowledges the relevance of the feedback offered by *Pensum*, the tutor most likely does not need to intervene, whereas when a disagreement arises it could point to useful information regarding the learning process.

Depending on the content or the texts, the strictness, or severity, of the feedback may be too large, prompting all sentences as incoherent and conversely. For that reason, we have implemented an *adaptable strictness* feature: the learner or teacher can adjust the feedback strictness using a slider.

Finally, all these information provided by the user need to be fully interpreted by the system, so that if a user indicates that a piece of feedback considered inadequate won't be proposed again to the learner (even when with a different severity). We therefore decided to store every action of the user through a *synthesis versioning* feature. All versions of the syntheses are now stored with their feedback. Additionally, this allows the learner to start exactly where he/she left off, instead of needing to invoke feedback whenever starting a new session on the same synthesis (see *improved processing time* in this section). Eventually (roadmap), the tutor will be provided with simple display of the list of all the different versions of the syntheses, along with their related feedback. This functionality is especially interesting for the tutors to monitor learners' work and progress, as well as for later improvement of the precision of the system.

Usability

The addition of a new status for feedback (questioned vs. unquestioned) has prompted us to perform various changes in the GUI, which was also an opportunity to improve known issues. The previous version of *Pensum* made excessive use of tooltips, which are often difficult to read and to interpret. The tabular view of feedback (tab named "Vue 2" in version 1.0), which had been made necessary by the tediousness of using the tooltips, has thus been removed. We believe that changing the way of representing feedback has made this view unnecessary. The tab made the reading of the text as a whole more difficult and the inherent necessity to display one sentence per line had prompted us to indicate in a separate column both coherence gaps before and after. Each single gap thus being represented twice. The coherence comparison between contiguous sentences is now displayed through more readable icons ('∦' and '∥' signs depending on whether the feedback has been questioned by the user, see item "control over feedback" below). The new interface is meant to improve conciseness and readability (despite the likeliness of a slightly steeper learning-curve, the user still depending on help and tooltips to master the new representation). Moreover, in the perspective of "widgetization", icons and menus have seen their size reduced, making the lateral and lower bars unnecessary.

In previous rounds, it appeared students were upset by being addressed as "*tu*" (informal) by the system. Correcting those messages has proved a good opportunity to implement *language localization*, which was especially important since one of the main aims of the LTfLL project has been to create services in different languages (moreover, the long thread prescribes that the LTfLL services use English as a common language, see D3.3).

Both English and French messages are stored independently from the structure of the display and *Pensum* can now provide interface in both languages (this also makes the addition of more languages easier). The handling of content in more languages and domains is on the other hand more time consuming (but *Pensum* is meant to handle them), as it relies on the development of new vector spaces.

Another issue concerning *Pensum*'s usability has been the processing time required, which we could consider solved for the use case "resume work" (see *control over feedback* section above), but not yet for others. Work has therefore in this direction. We started by integrating an R-LSA web service was conceived conjointly by both WPs (and implemented by WP 2), in order to diminish the number of operations both tests and computation of vectors (a particularly costly operation) by taking advantage of our versioning feature. Regarding the saving of the synthesis, it is very basic, in that it only stores a sentence in the database if it has never been written by the user. This allows the web service to keep the already computed vectors.

A first test processing feedback for a synthesis of about 420 words of a set of four course texts totaling around 1,320 words showed an improvement of 65% (41 s with *Pensum* v 1.5 vs. 1 min 57 s for *Pensum* v 1.0). Plus this measure was made for a first time analysis of the synthesis, depending on how much the text changes the improvement can be more drastic (25 s with *Pensum* v 1.5 vs. 2 min 08 s with *Pensum* v 1.0: 80% when adding two sentences to the aforementioned synthesis). The improvement in the second case in mainly due to the re-using of already computed vectors. It constitutes in our opinion a great advance in usability, as it allows the user to really ask for feedback at any time, since it seems, the most often he or she will request it the less time he or she will wait. Still, the processing time could be further improved (see roadmap § 5.3).

Low Level Improvements

Working on efficiency has shed into the light the need to improve the inner structure of the code in order to make *Pensum* more *scalable* and *compliant with standards*.

A lot of work has been dedicated to making the code more robust, more conform to W3C recommendations and more readable, which seems especially important for the further dissemination of the software considering that it is an open source application. By its web application nature, *Pensum* intertwines various types of sources and scripts interacting with each other in various programming and structuring languages, which makes its understanding more difficult. A lot of time has been devoted to cleaning up the code, especially for scalability and transferability needs, replacing hard coded paths and messages dispersed over the code with configuration files (multilingual in the case of messages). We also lowered the number of internal structures by directly providing html code out of AJAX calls instead of using an intermediary formalism and subsequent interpretation routines. The structure of the main document has been simplified and all format information stored separately. The same type of work has been performed on the storage of texts, the form of which will be later on less constrained by the structure of the database (preset list of format attributes replaced by structure indicators and CSS class handling). Finally, the use of an external javascript library (Jquery) has replaced most browser dependent interface calls making the code more readable and improving cross browser portability. The long thread version will also feature facilities to allow every user to select their own texts.

Interoperability

We have taken advantage of development of version 1.5 to prepare for interaction with (i)FLSS and CONSPECT. For self-regulating learning purposes, adding a text will be doable independently from the work of other WP in the upcoming widget version, but also by feeding off FLSS/iFLSS. Taking advantage of our versioning functionalities, we will also provide CONSPECT with one RSS feed per synthesis. Each paragraph/version can be one item to compare with CONSPECT. Inter-widget communication can make known to CONSPECT that new data is available from *Pensum* on each saving of the synthesis (see also section 5.2 for more information on both functionalities). Concerning interoperability with other LTfLL tools actually implemented in version 1.5, the main improvement of *Pensum* v 1.5 has been integrating R-LSA (Wild, 2007) (see usability section), an objective for WP 5.2 as stated in D5.2 (p. 26).

4.4. Brief Description of the Services' Interfaces

Interface of PolyCAFe

In this section, the focus shall be on the widgets that are effectively used by both learners and tutors. Therefore, the services used by tutors, teachers and administrators to setup and manage the course, assignments and conversations are not discussed. Thus, five web widgets were designed to use the web services and provide feedback and support for each conversation to students and tutors:

- Textual feedback on the whole conversation (based both on content and collaboration);
- Textual feedback for each participant of the conversation;
- Feedback and indicators computed at the utterance level (e.g. speech acts, importance of an utterance, etc.);
- An improved graphical visualization for a conversation designed for studying inter-animation;
- A semantic search widget for a discussion.

For a complete discussion about the features of each widget, please consult Appendix 4.

Interface of Pensum

After logging in the user interface allows the user to (all of these functionalities can be called at any time): ask for help, start/resume a synthesis or log out.

Once the user has chosen a set of learning material (and thus a synthesis), he is able to:

- Take notes for their benefit or the tutors' ("Notepad");
- Go through the various course text available;
- View them in full window;
- Edit the synthesis;

- Store it;
- Set the feedback tolerance;
- Launch the feedback;
- Question (and possibly link a sentence of the course texts and a sentence of the synthesis to justify)/unquestion the result of the said feedback.

Consult Appendix 5 for further explanations and screenshots.

5. Next Steps

5.1. Verification and Validation Rounds

The version 1.5 of the WP 5 services will be subject to verification and validation pilots, fully described in the upcoming D7.4. However, we can here describe their overall organization.

PolyCAFe *v. 1.5 Verification and Validation Processes*

The next verification/validation round of *PolyCAFe* v. 1.5 is planned to be as follows (detailed information shall be provided in D7.4).

- *verification process*: at least six tutors will annotate chat conversations (two for each tutor) and discussion threads at utterance level using exactly the same indicators that are provided by the system: importance/grade of an utterance, speech acts for chats and community of inquiry classes for forums. The results shall be used to further determine (supplementing those presented here) the performances of the system and are intended to be used in order to make further improvements for the roadmap.
- As can be seen from the comparison of the methods for the detection of the categories of the community of inquiry model for discussion forums (see section 4.3 and Appendix 6 of this deliverable), each approach has its own advantages and disadvantages – the choice for the rule-based approach is that it does not require a training corpus and is more domain-independent compared to the machine learning approach. These results shall be further investigated in the next validation round under “Prototypical Validation Topic 1: Verification of accuracy”.
- *validation process*: more than 25 students per experiment shall be using PolyCAFe in two different contexts and domains: HCI chats at Politehnica University of Bucharest and discussion forums for Medicine at the University of Manchester. Moreover, control groups shall be used and tutors shall also use *PolyCAFe* at PUB, as the UNIMAN context only uses student facilitators (there are not any tutors involved).

Pensum *v. 1.5 Verification and Validation Processes*

The next verification/validation round of *Pensum* 1.5 is planned to be as follows (more information to come in next D7.4). The *verification process* will be a two-fold verification session with about 18 2nd year Master students in educational sciences: half of them will play the role of students using *Pensum*, the rest playing the role of teachers/tutors in assessing already-made syntheses. The following scenario, close to an e-learning situation in real settings will be used:

- You (*Role 1*) are a student attending a course in ICT at a distance. You are engaged in a *1 hour and 20 minutes* session to revise a given topic of this course.
- Your teacher (*Role 2*) just gave you the task to write a synthesis of a bulk of documents about the "great digital divide", as an assignment. This task is very complex, since you don't only have to understand each document individually, but also to organize the main ideas in an understandable document. You can use *Pensum*, a web service that provides you feedback for that kind of task. Very briefly (see more information in *Pensum's* tutorial), *Pensum* gives information on the relevance of written sentences as well as their inter-sentence coherence. This information is not to be taken "as is", so you can act on each piece of feedback if you don't agree with it.
- Your task will be to connect yourself to *Pensum*, write out the required synthesis using the three kinds of feedback it can provide, then warn your teacher you finished your task. The teacher then connects to *Pensum* to read your work, the feedback you received and you modified. The teacher opens the notepad function of *Pensum* to write his or her comments on your production, and gives you a grade as well as some comments for improving your synthesis.
- You then can revise your synthesis according to the teacher's comments.

The *validation process* is not fully determined yet, since negotiations with the CNED (*Centre national d'enseignement à distance*) in Lyon (France) have been undertaken. It is planned to involve BA degree students in educational sciences at a distance, integrating ICT courses. A real-settings experiment is planned, close to the verification study, in which students will be given the task to write one or many syntheses during a decade, assisted by tutors.

5.2. Long Thread Integration

One of the main goals of this final report is to shed some more light on the possible integration between LTfLL services (more information in D3.3). This section is devoted to the description of the information flow as input/output towards a full integration of the LTfLL services, whose aim is to devise the long thread pilot study (as recommended in the second year review report, p. 3). WP6-related information from this table is taken from D6.3.

*Integration with WP4 (*CONSPECT*)*

PolyCAFe will input from CONSPECT the list of important concepts, which should be discussed by students in forums and/or chats (see also D4.3.2). The output from *PolyCAFe* will be the feedback about how the participants in the chat or forum used the concepts they were supposed to discuss.

Pensum can deliver RSS-based syntheses to CONSPECT (at different steps of the learning), which can in turn analyze them in order to reveal learners' conceptual development. Input for *Pensum*: User ID. Output generated by *Pensum*: Syntheses. Changes required: create a RSS feed as output.

Intra-WP 5 Integration

PolyCAFe and *Pensum* are respectively devoted to assess collaborative and individual knowledge building, so every pedagogical situation containing two contiguous learning phases (individual then collaborative work or reversely) would benefit from the use of the both services. However, since the input and output of these two services weakly match to each other (*PolyCAFe* only accepts chat and forum discussions as input while *Pensum* only deliver free texts), no specific integration project is possible between these services.

Integration with WP 6 (see also D6.3)

PolyCAFe identifies important concepts that have not been mentioned or that have not been mentioned enough by the learner. These concepts will be sent to iFLSS, which will provide learning resources for them. The WP5.1 widgets will send the request containing the query terms to the ontology-based search service and the query terms and the username to the social search service. In order to take advantage of the social search services the user should already use social networking tools and has to be registered as a user of the social search service.

Pensum's use would also benefit from social learning tools like (i)FLSS, since the proposition of course texts to read is static, and could become more dynamic with the interaction with (i)FLSS. Input for *Pensum*: Link to course texts retrieved from a query within (i)FLSS. Output generated by *Pensum*: facilities for the synthesis of the course texts. Changes required: Allow the learner to add course texts of their own.

5.3. Transferability (Pedagogical, Technical, Organizational)

PolyCAFe

We considered transferability of *PolyCAFe* in several directions. First of all, we considered in what domains may be used *PolyCAFe*. Our experience with using chats and forums for CSCL covers mathematics (in the VMT project, see Stahl 2006, 2009), Human-Computer Interaction (Rebedea et al., 2010; Trausan-Matu, 2010), Algorithm Design, Adaptive and Collaborative Systems, and Natural Language Processing. In almost all of these cases only text was used. The whiteboard of the VMT chat system offers the possibility of using also shared drawings but the actual version of *PolyCAFe* does not analyze the graphical interactions.

Due to the usage of LSA, all domains where textual descriptions of a descriptive knowledge are used (with few or no images, formulas, or numbers) are well suited. Several areas of computer science, literature, psychology, education, social and human sciences may be used. Conversely, there are domains where *PolyCAFe* is not suited to be used due to the needed graphics and images: several areas of geography, medicine, mathematics, physics, engineering, etc. All other domains for which it is difficult to have a large corpus of relevant text material for LSA training are also to be avoided.

Pedagogical transferability. PolyCAFe can be used in a variety of collaborative contexts: The types of conversations considered are very well suited for role-based discussions and debates, problem solving (in mathematics and design), creative discussions (brainstorming), etc. More specifically, we consider the following contexts:

- Use *PolyCAFe* together with chat or forums for revising for exams;
- Use *PolyCAFe* together with chat or forums for finding collaborative solutions to problems that can be described without the importance of a sequence of steps (PBL);
- Use *PolyCAFe* together with chat or forums to further investigate a given topic of interest to the learner (Self-Regulated Learning).

However, the system is not suited for the following situations:

- Use *PolyCAFe* together with chat or forums in a setting that involves scripted collaboration;
- Use of *PolyCAFe* together with chat or forums in a setting that does not involve or require collaboration (that is designed to be solved individually).

Technical transferability:

The most important technical problems are related to the following barriers:

- Functioning of LSA for the new domain and language;
- Support of Wookie-based widgets by the VLE, PLE, LMS used by the beneficiary organization;
- The integration of the NLP pipe for new languages is time-consuming and difficult.

Organizational transferability:

There are not a lot of limits for transferring *PolyCAFe* in another organization, however the following issue should be taken into consideration: privacy concerns about the use of chat conversation and discussion forums of the students by the services. In order to overcome these problems there are several solutions: anonymizing, do not show information about students to peers without their consent, asking for students consent before using the services.

Pensum

Pedagogical transferability: *Pensum* will be useful in educational contexts involving summaries or syntheses as activities for fostering understanding, either individually or collaboratively. Its integration has already been discussed in the "long thread scenario" (see D3.3 § 5.2 and § 5.2 of this Deliverable). Table 4 below details some possible additional pedagogical transfers, in individual and collective learning activities. For each of these activities the respective source text and learner production are mentioned (after Bonk & Dennen, 2003). It is worth noting that *Pensum* can be used for triggering collaborative or collective activities, as a brainstorming or a way to highlight the most important ideas from texts. This point has been particularly emphasized in the second year review report (§ 1.3): "The challenges suggest that the most fruitful avenue for development would be to focus on the tool's uses as a mechanism for triggering debate and other collaborative activities."

In the long run, following Rabardel's (2002) concept of instrumental genesis, another lead among others would also be to explicitely go in the direction of an instrumentation of *Pensum*. Rabardel defines instrumentation processes as "relative to the emergence and evolution of utilization schemes and instrument-mediated action: their constitution, their functioning, their evolution by adaptation, combination coordination, inclusion and reciprocal assimilation, the assimilation of new artifacts to already constituted schemes, etc." (Rabardel, 2002, p. 103).

Table 4 – Main Pedagogical Activities involving *Pensum*'s Transferability

Modes	Activity	Source texts	Learner Production
Individual	Case Study	Several cases	Analysis of the cases
	Portfolio Analysis	Portfolio	Analysis of the documents of the portfolio
	Lecture Notes Analysis	Lecture notes	Synthesis of the lecture content
	Synthesis production	Several documents	Written synthesis
	Learning academic writing	Texts on a given domain	Written synthesis
	Understanding a tutorial	Tutorial	Explanations on the tutorial content
Collective	Collaborative writing	Several texts (1 per learner)	Collaborative synthesis
	Brainstorming or debate	Case study	New ideas from the case study, main points of a debate
	Community of Inquiry-based pedagogy	Inquiry documents	Notes from integration and resolution phases (see Appendix 1)

Technical transferability:
The major barriers for the adoption of *Pensum* are stated in the document; many are mainly technical (for the problem of acceptance by teachers and institutions is not specific to *Pensum*):

- the functioning and validation of LSA for the new domain and language;
- corpus retrieval and adaptation for working with specific domains;
- processing time (see § 2.2 and § 4.3)
- the lack of a teacher interface and the size of the display zones, for instance the synthesis editing text area (see § 5.3)
- possible lack of validity for feedback (depending on the type of course texts).

All of these problems are being addressed, but most are so as part of the roadmap. It is difficult to foresee the impact of these main technical limitations on *Pensum*'s transferability. However, the recent adaptation of *Pensum* to R-LSA engine would allow an easier adaptation of its algorithms.

Organizational transferability:

Once the likely technical issues solved (see previous point), there are no specific issues for transferring *Pensum* to another organization. This service provides sufficiently fine-grained activities to be adapted in a wide range of organization and purposes. The reason why is two-fold: the flexibility of the tools offered in *Pensum* (e.g., both the feedback and the "notepad" can be used either by learners or tutors); the possibility for learners (as well for tutors) to reject the feedback if it appears they don't have enough confidence on it. To extend organizational transferability (see § 5.4), a more extensive administrative interface for both researchers or teachers would unlock the source of data that the recording of learners' actions on feedback is (both on LSA and learning processes). However, as stated in D7.3 § 3.6.2, some organizational barriers may hinder the transferability of *Pensum*. First of all, the compatibility between services issue is not fully addressed yet, as long as the long thread pilot study is not fully completed. Both language portability and interaction between services remain to be fully addressed for a better organizational transferability of *Pensum*.

5.4. Road Map

The lifecycle of LTfLL is limited; moreover, the development cycle of the second version has been reduced (delivering v. 1.5 instead of v. 2.0). The last validation round should feed the next iterations, which happen to be outside our project. The roadmap described below is the development agenda for that next iteration.

Although it offers a lot of functionalities, *PolyCAFe* has to be enhanced in several dimensions:

- *Improving the content of the utterance and conversation feedback*. This will be done considering the results of the final validation round.
- *Improving the evaluation of collaboration*. This will be done starting from data of the final round of verification/validation.
- *Including the modules needed for using new languages*. This was started for Romanian (a POS Tagger was developed) and will be completed after the end of the LTfLL project by developing speech act tagging and other annotation for Romanian.
- *Improving the forum analysis* for discussion forums that involve more than 20 participants. However, in this case the polyphony paradigm should be developed to consider the case of a very large number of participants. Moreover, the social network analysis influence in scoring the participants should be extended as central participants become more important. It should also be taken into consideration that in this cases, a user is part of more than a single discussion thread as is the case of the discussion forums used by the University of Manchester.

The following functionalities may be added to *Pensum* for improving learners' and tutors' activity.

- *Tutor/administrator* interface: to this date, there is no specific interface for teacher, tutors or administrators (or even researchers) enabling them to get students' tracks or syntheses. Part of this interface may be implemented by integrating the existing user interface of course selection (common to all LTfLL services) but other parts are more application-specific such as the handling of the various traces of the student's actions. We have chosen not to include this in version 1.5 as it seemed more important that the system should be acceptable for learners, before we deployed facilities for complex administrative duties.
- *Dynamical thresholds*, accounting for text variability. Since the types of texts submitted to students during courses is highly variable, it could be useful to add a 'calibration' functionality, which may calibrate the thresholds of coherence and topic selection/rejection beforehand. The results of this functionality, of course, have to be compared against human judgments. The data gathered during the use of *Pensum* might be useful in this task.
- *Choice of vector space* according to domain could also improve system results. Since roadmap version will use a different space for English and French handling, only minor changes will be required to obtain this feature as part of the administrator interface.
- *Adaptation of the learner interface*, the research literature (Olive, Rouet, François, & Zampa, 2008) highlights the importance of the setting up of the work spaces (summarized text and synthesis), concluding that showing both windows at the same time "allows for better control over the summary writing task". But the quantity of text readable might be an issue as well, which added to the space issue in the context of widgets prompts for interface rethinking in terms of single versus multiple widgets. The same authors also mention the potential help provided by highlighting and annotation tools. Those could extend the feedback actions implemented in v. 1.5. A careful thinking of display and action options is necessary in order to provide useable information.
- *Performance*, despite improved could still be made slightly faster. Especially, when using *Pensum* to summarize longer texts, storing in the database every comparison result (each associated to a given space) would allow to query LSA only for new sentences, thus reducing a little further execution duration.
- *Cross-learner actions on feedback*: We have so far focused on associating to each learner their actions and taking them into account. An improvement could be to view user actions as global improvement to the system and make the latter learn from students' actions in storing all their actions and judgements. For instance, if a first learner links a sentence of his or her synthesis to this of the course, the system could then automatically link this latter sentence to all sufficiently close synthesis sentences of learners to come—users being able to react on that kind of feedback like on any other, the system thus using a so-called Web 2.0 approach, especially in "Harnessing Collective Intelligence" (O'Reilly, 2007) to LSA.

6. Conclusion & Open Points

As noted in Dessus et al. (2010) learning activities engaged in collaborative e-learning contexts share some specificities. First, they are based on writing. Second, their manifestations are both at individual and collective level. Third, their aim is twofold: at covering (learning) a given knowledge domain but also at leading a pedagogy-related activity. Fourth, they require to be analyzed in order to provide an adequate feedback. Fifth and last, the stakeholders to be considered are not only the learners and the teachers, but also the researchers studying the activity. 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 (Dessus & Trausan-Matu, 2010).

This three-year long research on relations between writing and learning has led to some results:

- Reading, writing, chatting are intertwined activities leading to learning and proposing automated feedback to learners for these activities in formal or informal settings.
- Chat and forum semantic analyses of utterances can be partly automated and provide useful information for learners and teachers to help them build knowledge collaboratively.
- Learners can write out course summaries and syntheses and be given automated feedback to help them build knowledge individually.

There also are open points that remain to be investigated in further research programs:

- How Bakhtin’s notions, often thought-provoking though sometimes vaguely formulated, can be further operationalized with Natural Language Processing techniques;
- How learners can cope with the organization of a set of widgets without being cognitively overloaded?
- How learners can manage their work collectively and use the services as cognitive tools?
- Which kind of specific interface to dedicate to teachers for them to monitor learners’ processes?

7. Appendices

Appendix 1 — Categories in the Forum Analysis

The categories describing the main elements of "Community of Inquiry" theory can be identified in forum messages using the definition of each category, pattern matching and other solutions. Table 5, proposed by Garrison, Anderson and Archer (2000), describes the categories of each element.

Table 5 Main categories used by the "Community of Inquiry" model

Elements	Categories	Code	Indicators
Cognitive presence	Triggering Event	Cte	Sense of puzzlement
	Exploration	Ce	Information exchange
	Integration	Ci	Connecting ideas
	Resolution	Cr	Apply new ideas
Social presence	Emotional Expression	See	Emotions
	Open Communication	Soc	Risk-free expression
	Group Cohesion	Sgc	Encouraging collaboration
Teaching presence	Instructional management	Tim	Defining and initiating discussion topics
	Building understanding	Tbu	Sharing personal meaning
	Direct Instruction	Tdi	Focusing discussion

The cue phrases that *PolyCAFe* considers for each are described below.

Cognitive Presence

Triggering Event – cte – identifying and recognition of a problem, sense of puzzlement, questions and explanations

- Triggering an event: "to start the discussion…", "I would like to start with/by…", "the next issue to discuss …", „we should start by clarifying" , „if you have ... questions", "I am here to guide the group"

Exploration – ce – focused on exchanging information and clarifying situations or terms

- Questions: "why .. ?", "where ... ?", "We should also consider about"
- Exploring : "I am wondering who is responsible…", "We should also consider/think about…", " I think this is an interesting topic, but others have a different perception of it, to mine and we have to take this into account"

Integration – ci – integration of knowledge and ideas

- Connecting ideas: “another thing I have noticed is…”, “I believe your idea is perfect”, “I guess”, “I feel professional behaviour” , “I have come across”
- Searching for answers and solutions: “to solve this I would…”, “I believe”…`
- Giving examples: “for example”, “the reason why”

Resolution – cr – checking the founded solutions, exploring consensus, agreements or disagreements

- Applying ideas: “you can do this by…”, “to achieve that you/we can”
- Evaluating the solutions : “the best thing to do is…”, “I think the best way to do this...”, “I think this shows...”, “Perhaps the best way of approaching this problem is not by direct confrontation, but by asking questions in such a manner as to be probing but not offensive, which would improve inappropriate practices by a senior professional by making them think about their actions”
- Asking for answers: “I did not understand why…”,”could you explain…”

Social Presence

Emotional Expression – see - sharing and expression of feelings, emotion, humour, irony, etc.

- Key words that express emotions (“angry”, “anger”, ”happy”, “happiness “, ”sad”, “sadness”, ”surprised”, “unhappy”, “fear”, etc
- Emoticons: ☺ ☹ :P >:)
- Expressions: “I’m only human”, “in your dreams”, “you must be joking”, etc.
- Words that generate or have an influence on emotions: “anxious”, “arrogant”, “fierce”, etc.

Open Communication - soc – referring to other posts or other persons posts, acknowledging other or their ideas, pushing others to participate in conversation

- Cite a person : ”Alex said that...”, “you say ...”
- Referring others: “you”, “yours”, “your”
- Opinions or comparisons : “agree”, ”disagree”, “I think”, “I don’t think”, ”I believe”, “I like …better”, “X is … than Y”, “in my opinion”

Group Cohesion - sgc – encouraging collaboration and group interchanges and it is based on accepting differences of opinion and a use of designated group by using the “us” term.

- Address to the group: “us”, “we”, “our”
- Encouraging group discussions: “let’s talk/do/try…”, “let’s all say something about…”
- Addressing questions to the group: “who can..”, “where..”+”?”, “am I right?”, “does anyone know”, “can anyone…”, “if anyone agrees, let me know”, “I would appreciate if X would reply/add a few words”
- Greetings: “hello”, “good morning/afternoon/evening”, “good bye”, “see you soon”, “hi”, “hey”, “hi all”, “hello all”, etc.

Teaching presence

Instructional Management – tim - Facilitating establishment of group organisation and guidelines, initiating discussions, facilitating choice of topics, establishing ground rules and netiquette

- Discussions about tasks and their management: “We also need to include evidence ... of our participation ...”, “look in.../at page...”, “you should consider..”, “We must finish this discussion by Friday…”
- Presentation of other helping ideas: “Your task is to…”, “Your assignment…”, “the first team will have to…”

Direct Instruction – tdi - when discussion or group has reached a dead end, teacher will help by giving references to other (outside) knowledge or other helping questions

- Direct instructions: …”If you want to upload an attachment just click on….”
- Dead ends: “since there’s not more to say, let’s talk…”, “why don’t we change the main idea”
- Encouraging students: “X had the best…”, “I would have appreciated if …"
- Technical problems: ”you will be able to post messages only after 10 PM”
- Indicators about quality of learning process: “What an interesting discussion”

Building understanding – tbu – summarising, facilitating group collaboration and discussion, focusing on agreements and disagreements, encouraging others to participate to discussion, appreciating the useful comments from others.

- Encouraging communication: “I am curious, what X’s opinion is…”, “…maybe X could tell us”, “Any thoughts on this issue?”, “Anyone care to comment?”
- Key questions : “Would a sensible suggestion be if you ...”,”what if we ...from X’s point of view”

Appendix 2 – The evaluation process in *PolyCAFe*

System Architecture

Technical architecture

For an enhanced and holistic assessment of participants, a multi-layered architecture has been proposed with the following tiers (Rebedea et al., 2010):

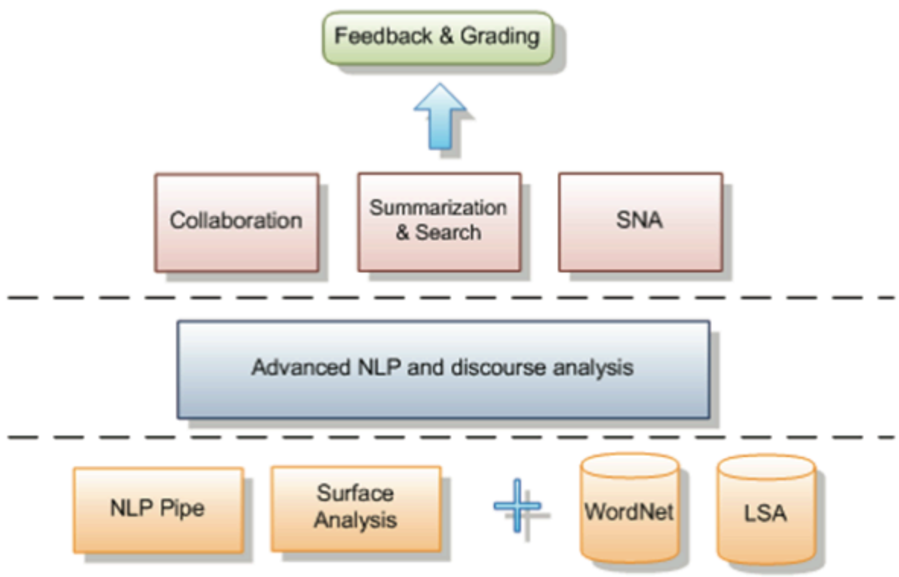

Figure 8 — System's technical architecture

1. First layer with surface evaluation, basic NLP processing and predefined concept definitions

The first step in analyzing the raw data is the NLP pipe which covers spelling correction, stemming, tokenizing, Part of Speech Tagging. Next surface analysis is performed consisting of metrics derived from Page's essay grading techniques (Page & Paulus, 1968) and readability measures. The semantic sub-layer defines concepts using both linguistic or domain specific ontologies (e.g., Wordnet) and Latent Semantic Analysis (LSA). These two approaches form the basics for a semantic evaluation of the participants' involvement and evolution. In contrast with surface analysis, which is based only on quantitative measurements, ontologies and LSA enable a qualitative assessment of the overall discussion and of the participants involved.

2. Second layer centered on advanced NLP and discourse analysis

For a deeper insight of the discourse, techniques as speech acts identification, cosine similarity between utterances are used for identifying interactions among participants. In addition, each utterance is evaluated using the semantic vector space from LSA in order to determine its importance regarding the overall discourse.

3. The third layer regarding collaboration, social networks built on determined interactions and polyphony

Social network analysis takes into account the social graph induced by the participants and their corresponding interactions that were discovered in the previous layer. Collaboration also plays a central role in the discourse for highlighting the involvement and implications of each participant. This is essential from the perspective of Computer Supported Collaborative Learning because a chat with high collaboration has greater impact and is more significant in contrast with others in which the discourse is linear and the utterances are not intertwined in a polyphonic manner. Extractive summarization and search are easily achievable by interaction with other components and come as features.

The final step in the analysis consists of combining the results of previous sub-layers in order to offer textual and graphical feedback and a grade proposal for each participant of a chat discussion.

Functional architecture

Communication between participants in a chat is conveyed through language in a written form. Lexical, syntactic, and semantic information are the three levels used to describe the features of written utterances (Anderson, 1985), and will be taken into account for the analysis of a participant's involvement in a chat. Therefore, the multi-layered architecture can be presented also by using the following model:

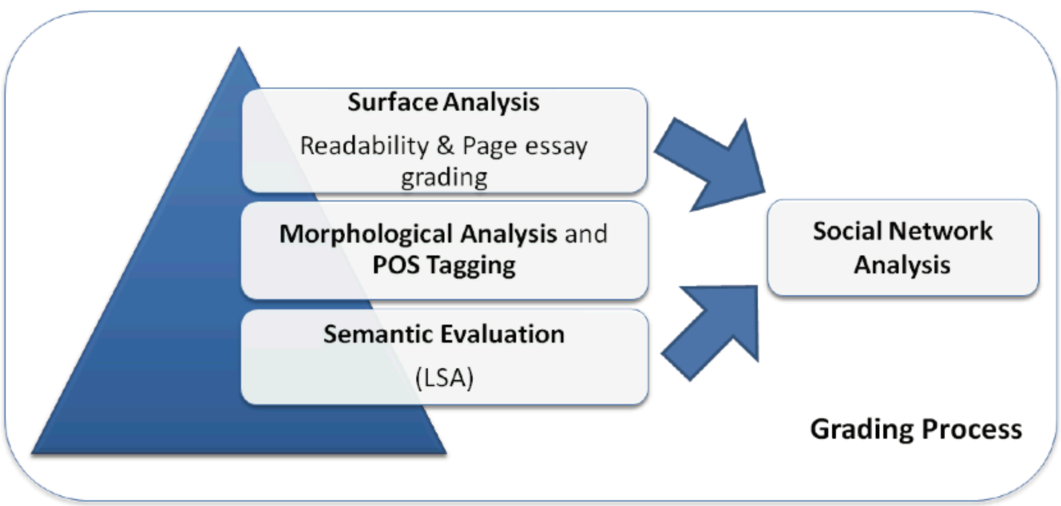

Figure 9 — System's functional architecture

First, surface metrics are computed for all the utterances of a participant in order to determine factors like fluency, spelling, diction or utterance (Page & Paulus, 1968). All these factors are combined and a mark is obtained for each participant without taking into consideration a lexical or a semantic analysis of what they are actually discussing. At the same level readability ease measures are computed.

The next step is morphological analysis (spellchecking, stemming and tokenization) and part of speech tagging. Eventually, a semantic evaluation is performed using Latent

Semantic Analysis. For assessing the on-topic grade of each utterance a set of predefined keywords for all corpus chats is taken into consideration.

Moreover, at the surface and at the semantic levels, metrics specific to social networks analysis are applied for properly assessing participants' involvement and contribution with regards to similarities with the overall chat and with the predefined topics of the discussion.

Tagged LSA and semantic space visualization

Latent Semantic Analysis starts from the *vector-space model,* later on used also for evaluating similarity between terms and documents, now indirectly linked through concepts after projection is performed (Landauer, Foltz, & Laham, 1998a; Manning & Schütze, 1999).

The usage of LSA in our system is based on a *term-document matrix* build upon a corpus of chats with terms extracted from utterances after applying stop words elimination and spellchecking. Term Frequency – Inverse Document Frequency is applied and the final transformations are singular value decomposition and projection after the optimal considered empiric value for the number of dimensions *k* of 300, a value agreed by multiple sources (Lemaire, 2009).

Two important aspects must be addressed: *tagging* and *segmentation*. POS tagging is applied on all remaining words and a specific improvement is the reduction of existing verb forms by applying stemming only on verbs. According to Lemaire (2009) and Wiemer-Hastings and Zipitria (2001), stemming applied on all words reduces overall performance because each form expresses and is related to different concepts. Therefore the terms included in the supervised learning process are made up of original words/stems plus their corresponding part of speech.

Segmentation divides chats into meaningful units by considering cohesion and unity between utterances. In the current implementation we use the utterances of the same participants because of their inner consistency and fixed non-overlapping windows for defining the maximum document length. A future improvement will include segmentation based on thread identification.

A small modification in the determination of the corresponding vector of each utterance is the use of the logarithmic function for smoothing the impact of a concept in a single utterance.

One immediate application of LSA is for *topic identification*, which relies on the cosine similarity of each word with the whole document. The maximum value reflects the best correlation between the concept and all terms used in the discussion.

Another important aspect is the visualization of the resulted vector space. Therefore a network that is very similar to a social graph is generated by using the terms as nodes and the similarity between terms as the strength of the edge between vertices. This network is obtained by applying Breath First Search from the starting term (manually introduced by the user into a form with auto-complete options) and by inducing two heuristics maximum *k*-nearest neighbors selection and a minimal threshold for inter-word similarity for massively reducing weak links between terms.

The grading process of chat utterances

The first step in the grading process involves building the utterance graph in which two types of utterances can be identified: explicit ones, added manually by participants during their conversations by using a facility from the conversation environment – in our case, *Concert Chat*; the second is based on implicit links automatically identified by means of co-references, repetitions, lexical chains and inter-animation patters. Each utterance is a node and the weights of edges are given by the similarity between the utterances multiplied by the trust assigned to each link. By default, each explicit link has a trust equal to 1 and each implicit link has an inner trust related to the factor used for identification scaled from 0 to 2 multiplied by an attenuation factor – in our current implementation 0.5 for allowing further fine-tuning of the grading process. The orientation of each edge follows the timeline of the chat and the evolution of the current discussion in time. The link is actually a transpose of the initial referencing from current utterance to the previous one to which it is logically linked explicitly or implicitly.

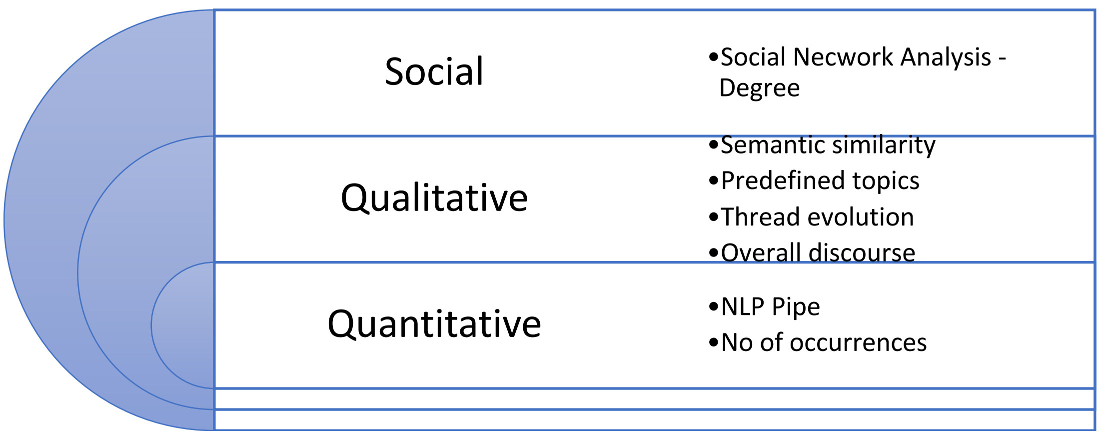

Figure 10 — Hierarchy of utterance evaluation

The actual grading of each utterance has *3 distinct components*: a *quantitative*, a *qualitative* and a *social* one.

The ***quantitative perspective*** evaluates on the surface level the actual utterance. The assigned score, only from the quantitative view, considers the length in characters of each remaining word after stop words elimination, spellchecking and stemming are applied. For reducing the impact of unnecessary repetitions used only for artificially enhancing the grade, we use the logarithm of the number of actual occurrences of each word. Another improvement in evaluating each utterance involves applying the logarithm function on the previously obtained result, therefore considerably reducing the impact of oversized, but not cognitive rich utterances.

A more interesting dimension is the *qualitative* one (Wiemer-Hastings & Zipitria, 2001), which involves the use of LSA in determining:

Thread coherence

Based on the previously identified utterance graph, thread coherence for a given utterance represents the percentage of links, starting from that specific utterance, which share a similarity above a given threshold. In order to ensure inner cohesion and continuity within a thread, similarities between any adjacent utterances must exceed the specified threshold (in our case 0.1);

Future impact

Future impact enriches thread coherence by quantifying the actual impact of the current utterance with all inter-linked utterances from all discussion threads that include the specified utterance. It measures the information transfer from the current utterance to all future ones (explicitly or implicitly linked) by summing up similarities above the previously defined threshold. In terms of Bakhtin's (1993) philosophy, this can assimilated to voice inter-animation and echo attenuation in the sense that multiple voices from current utterance influence directly linked utterances (both explicitly and implicitly). Future impact, therefore the echo of a given voice, is estimated by measuring similarity between the two linked utterances. The term 1 in the final formula expresses current utterance (inner strength), while all other factors represent a bonus due to the influence measured by means of cosine similarity over future utterances.

Relevance

Relevance expresses for each utterance the importance and the coherence with the entire discussion. This can be easily measured by computing the similarity between the current utterance and the vector assigned to the entire chat, therefore determining the correlation with the overall discussion.

Completeness

Because each discussion has a set of predefined topics that had to be followed and which should be in the center of each chat, completeness measures the actual coverage of each utterance of those keywords. This can be obtained by evaluating the similarity between the utterance and the vector of the specific set of keywords specified by the tutor or teacher as important topics for the discussions. This measure can also be used with regards to automatically determined topics obtained through LSA.

The ***social perspective*** implies an evaluation from the perspective of social networks analysis performed on the utterance graph. In the current implementation only two measures from graph theory are used (in-degree and out-degree), but other metrics specific to SNA (for example betweenness) and minimal cuts will be considered. The centrality degree from social networks analysis is not very relevant because all the links follow the flow of the conversation and therefore they are all oriented in the same direction.

By combining all previous dimensions, the formulas used for marking each utterance are:

$$mark(u) = quantitative(u) \times qualitative(u) \times social(u)$$

$$quantitative(u) = \log \left(\sum_{remaining\ words} length(stem) \times (1 + \log(no_{occur}) \right)$$

$$qualitative(u) = thread\ coherence(u) \times future\ impact(u) \times relevance(u) \times completness(u)$$

$$thread\ coherence(u) = \frac{no\ of\ future\ links\ with\ similarity\ above\ threshold}{no\ of\ future\ links} \quad \text{(E1)}$$

$$future\ impact(u) = 1 + \sum_{\substack{future\ link\ l \\ v \to u}} sim(u, v)$$

$$relevance(u) = Sim(u, vector(entire\ discurssion))$$

$$completness(u) = Sim(u, vector(predefined\ keywords/topics))$$

$$social(u) = \prod_{all\ SNA\ factor\ f} (1 + \log(f(u)))$$

Participant grading

The in-degree, out-degree, closeness and graph centrality, Eigen-values and rank factors are applied on the matrix with the number of interchanged utterances between participants and the matrix which takes into consideration the empiric mark of an utterance instead of the default value of 1. Therefore, in the second approach quality, not quantity is important (an element [*i*, *j*] equals the sum of mark(*u*) for each utterance from participant *i* to participant *j*), providing a deeper analysis of chats using a social network's approach based on a semantic utterance evaluation.

Each of the analysis factors (applied on both matrixes) is converted to a percentage (current grade/sum of all grades for each factor, except the case of Eigen centrality where the conversion is made automatically by multiplying with 100 the corresponding Eigen-value in absolute value). The final grade takes into consideration all these factors (including those from the surface analysis) and their corresponding weights:

$$final\ grade_i = \sum\nolimits_k weight_k * percentage_{k,i} \quad \text{(E11)}$$

where k is a factor used in the final evaluation of participant i and the weight of each factor is read from a configuration file.

After all measures are computed and using the grades from human evaluators, the Pearson correlation for each factor is determined, providing the means to assess the importance and the relevance compared with the manual grades taken as reference. General information about the chat – for example overall grade correlation, absolute and relative correctness – are also determined.

Collaboration assessment in a chat conversation

Collaboration in a chat environment is assessed based on the following measures: social cohesion and collaboration, quantitative, mark based and gain based collaborations which will be thoroughly described in the next sub-sections (Dascalu, Rebedea, & Trausan-Matu, 2010).

Social cohesion and collaboration

Starting from the analysis of the social network performed both at the surface level and at the semantic one with participants as nodes of a graph, the following metrics is derived for assessing equitability and cohesion for all participants. An generally accepted remark regarding a collaborative environment is that the more the participants have equal involvements (analyzed at surface level using only the number of interchanged utterances) and knowledge (relative to the topics taken into consideration and evaluated upon the marking process with LSA as support for semantic analysis) the more they collaborate being on the same level of interest and common/interchanged knowledge.

In order to have a normalized measure for the spread of each factor taken into consideration for social network analysis, the coefficient of variation is computed for each metric. The overall result represents 100% minus the mean value of all partial results because it is considered that the more the participants have similar involvements, the better the collaboration. A later improvement will involve using weighted influences for each factor in the final collaboration assessment (for example in-degree represents more interest in the current speaker, whereas out-degree expresses gregariousness in a certain manner; therefore their corresponding weights should have different values because of the nature of interactions expressed by these factors).

Quantitative Collaboration

The most straightforward approach for assessing the collaboration level of a chat is by using the number of explicit links and of the implicit ones with a trust coefficient assigned to them. Collaboration is essentially illustrated when certain ideas or viewpoints present within the current utterance are up-taken or are transferred to a latter one, issued by a different speaker. In other words, our opinion is that collaboration can be measured using linked utterances that have different speakers.

Starting from this idea, a quantitative collaboration score can be computed using the following formula:

$$quantitative = \frac{\sum_{\substack{all\ links\ l \\ with\ \boldsymbol{different}\ speakers}} attenuation(l) \times trust(l)}{total\ number\ of\ links\ (explicit\ or\ implicit)} \quad \textbf{(E2)}$$

where:

- attenuation(l) is used for assigning a different importance to the explicit links that are considered more valuable (score of 1) relatively to the implicit ones (score of 0.5);
- trust(l) is the assigned trust for an implicit link (for example, in the case of direct repetition this value is set to 2); for all explicit links, trust is set to 1.

By combining attenuation (values of 0.5 and 1.0) with trust (values in [0..2]), all explicit links have a score of 1 for their contribution, whereas implicit links can achieve the same maximum value only if their corresponding trust is maximum.

Gain based collaboration

The following two measures are computed starting from the assessment of the following formulas for any utterance *u*:

$$gain(u) = personal(u) + collaborative(u) \quad \text{(E3)}$$

$$personal(u) = \sum_{\substack{links\ v \to u \\ \boldsymbol{diffent}\ speakers}} \Big(\big(mark(v) + gain(v)\big) * Sim(u, v) * atten * trust(l) \Big) \quad \text{(E4)}$$

$$collaborative(u) = \sum_{\substack{links\ v \to u \\ \boldsymbol{same}\ speakers}} \Big(\big(mark(v) + gain(v)\big) * Sim(u, v) * attent * trust(l) \Big) \quad \textbf{(E5)}$$

The fundamental ideas are derived from information theory (Shannon, 1951) in the sense that each utterance has an overall gain composed of both personal evolution (links to previous personal utterances) and collaborative building (ideas shared, future implications), with transfer of knowledge, ideas or concepts through the identified links. Therefore, for each utterance *u* we take into consideration all previous utterances *v* where a link (explicit or implicit) exists and we try to determine the amount of information that is transferred. Each utterance *v* has a score assigned after the marking process and a gain computed using its previous inter-linked utterances. Starting from these two components (cumulative build for gain and current importance using the corresponding mark), LSA is used for assessing the actual transfer, impact and correlation, in other words similarity between the two utterances (*u* and *v*).

The main difference between the two gains is that personal gain assumes that inner knowledge is built, whereas collaborative gain is expressed relative to others and expresses working together, sharing ideas and influencing the other participants of the discussion.

Our ideas are correlated with Bakhtin's dialogistic theory (Bakhtin, 1993), in the sense that to achieve genuine collaboration, true polyphony, with an actual dense interlacing of voices is needed. The concept of voice at Bakhtin and in our approach has a larger extent than the usual acoustic one (Trausan-Matu & Rebedea, 2009).

Links, both explicit and implicit, express the idea of inter-changing concepts or ideas, which are later processed with a new voice as result. In the end, a particular sense is given to each voice, taking into consideration echoes from previous voices related by the determined links. From this point, gain can be considered the echo of all previous inter-linked voices, attenuated corresponding with similarity and trust, and the mark is the actual implication, strength and importance of the current voice. Personal gain expresses and measures implicit inner dialog (individual voices), whereas the collaborative gain addresses explicit dialog (by definition between two participants), also highlighting external voices.

Using the previous formulas, the following two metrics are derived for assessing collaboration in a chat conversation:

Formula (E6) is used for estimating the percentage of overall utterances' importance/marks relatively to information build/transferred in a collaborative manner.

$$mark\ collab = \frac{\sum_{all\ utterance\ u} collaborative(u)}{\sum_{all\ utterance\ u} mark(u)} \quad \text{(E}\mathbf{6}\text{)}$$

Formula (E7) is used for assessing collaboration relatively to overall gain (practically excluding inner build).

$$mark\ collab = \frac{\sum_{all\ utterance\ u} collaborative(u)}{\sum_{all\ utterance\ u} gain(u)} \quad \text{(E}\mathbf{7}\text{)}$$

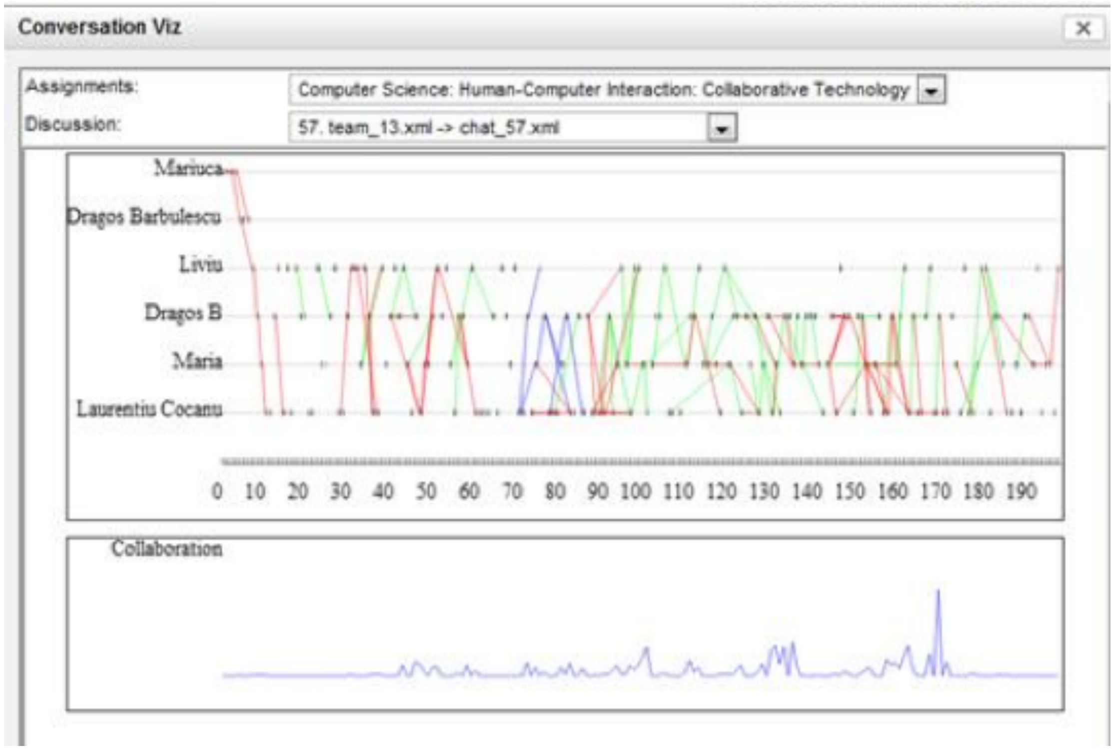

Figure 11 — Collaboration assessment and chat evolution visualization.

The overall collaboration mark is obtained by multiplying the four previously defined metrics:

$$\text{overall collaboration} = \text{social cohesion} \times \text{quantitative} \times \text{gain based} \times \text{mark based} \quad (E8)$$

Semantic extractive summarization

The most important aspect that can be identified regarding the summary generation is the tight link between obtaining a summary and evaluating chat participants. Therefore, in order to obtain a good extractive summary each utterance must be assessed – in our case, the marking process has been previously described and it provides the basis for identifying the key utterances.

A set of predefined percentages of extraction is provided, allowing the user to select the level of interest at which the overall view of the chat is desired. Starting from this percentage the number of utterances to be displayed is determined. After that, utterances are extracted in decreasing order of importance. A very important fact is that this process has a *gain based selection* (utterances are ordered based on the sum of corresponding marks and gains – both personal and collaborative ones), therefore combining collaboration and the marking process of each utterance. The approach of granting bonuses to the next level utterances in explicit threads and correspondingly weighted with trust for the implicit ones is reflected by the previously described gain which takes into account both implicit and explicit links, and fundaments its measure on semantic similarity.

The impact of using the automated summary system was the reduction with more than 30% of the time spent by the tutor for the evaluation of a single chat.

Semantic search

A distinctive and yet important component of the system addresses enhanced search capabilities within a conversation. Starting from a given query, two types of results are being assessed relatively to a single chat: a classification of participants with the criteria of best overall performance with regards to the given query and an utterance list in descending order of scores. In our case, these search facilities are centered on a scoring mechanism of each utterance.

In order to evaluate the relevance of each utterance to the given query, three steps are performed. First, the query is syntactically and semantically enriched by using synsets from *WordNet* and by selecting the most promising neighbors from the vector space model computed using LSA. For reducing the number of possible newly added items to the query, a threshold and maximum number of neighbors have been enforced. The final scope is to obtain a list of words and a query vector.

The next step addresses the syntactical level by evaluating the number of actual occurrences in each utterance. In this process, different weights for original/ enriched words and stems are also considered for reflecting the importance of each utterance.

The third step focuses on a semantic assessment by measuring the cosine similarity between the query vector and each utterance vector. The final score is obtained by multiplying the semantic score with the social one and with the sum of mark and gain determined in the previous analysis of the chat. These last two factors are important by taking into consideration the actual importance and cumulative gain of each utterance.

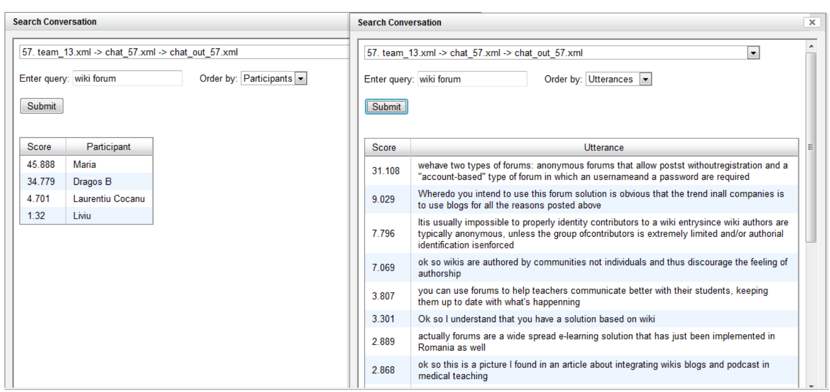

Figure 12 — Web widget displaying the results of the semantic search

The results can be viewed in Figure 12 in two different windows with regards to the searched factor – participants or utterances.

Appendix 3 – The Writing-to-Learn Pedagogical Approach and the LTfLL Services

Overall, the pedagogical orientation of distance learning tasks often pertains to three intertwined fields of research and theoretical assumptions:

- *the writing-to-learn approach*, which assumes that writing on a given subject matter is a good first step toward learning it (Emig, 1977);
- *the multiple-source writing*, which assumes that if learners are able to capture the gist of a couple of documents, they undertake a conceptual transforming activity closely akin to knowledge building (Segev-Miller, 2004) ;
- *the self-regulated learning approach*, which assumes that learners are able to regulate the way they learn in controlling some aspects of their cognition, behavior or affects (Boekaerts, 1997), notably through control- and monitoring-based strategies of their activity (either reading and writing).

The purpose of this Appendix is to describe more extensively the first approach, since the two latter have been extensively described in D5.1 (§ 2.1) and D5.2. (Appendix 1). Very roughly, we assume that learners can transform their writing in knowledge when they use the two services described above. This viewpoint is named “writing-to-learn” in the research literature (Bangert-Drowns, Hurley, & Wilkinson, 2004) and D5.1 § 2.1. The way the writing activity can promote learning is not fully uncovered, but, as Langer and Applebee (1987, pp. 135–136), put it:

> “1. Writing activities promote learning better than activities involving only studying or reading.
>
> 2. Different kinds of writing activities lead students to focus on different kinds of information.
>
> 3. In contrast to short-answer responses, which turn information into discrete small pieces, analytic writing promotes more complex and thoughtful inquiry but on a smaller amount of information. (Langer & Applebee)“

PolyCAFe and *Pensum*, as well as some tools implemented in this research project can support the first assumption. The second can be the possibility to cite all the kinds of writing activities students can carry out with our services (e.g., for *PolyCAFe* and *Pensum*: notes, keywords, essays, syntheses, chat utterances, definitions), for which forms of knowledge and which processes. By the way, we could state how and to what extent a given form of writing contributes to learning. Which are the cognitive strategies at stake within this approach? According to Klein (1999, p. 211), four non-contradictory strategies can be seen (Figure 13 depicts the four strategies graphically):

- *spontaneous*: “writers generate knowledge ‘at the point of utterance‘, that is, by generating language, without planning or revision” (see the knowledge telling mode, (Bereiter & Scardamalia, 1987)). This strategy converts tacit knowledge into explicit knowledge in making semantic and syntactic choices for his or her writing activity, as if speech was writing down. Klein suggests that this mode plays a “real but limited effect on learning“ (Klein, 1999, p. 220).

- forward search: “[...] writers externalize their ideas in text, then reread this text, and make new inferences based upon it“ (Klein, 1999). The written text plays the role of an external memory, enabling the writer to store ideas and reflect on them afterwards. In reviewing the text written so far, the initial goals of the writer are reconsidered, transformed, to form a new goal, and operations like making inferences, evaluate, organize are carried out to transform and revise the initial text according to the learner's goals. In that strategy, what counts is the current writing goal, and how it can be developed and modified, as justified by the term “forward search”.
- Genre-related hypotheses: writers use genre structures (i.e., kinds of discourse, rhetorical intention) to organize relationships among elements of text, thereby linking elements of knowledge. This genre affects, according to Klein (1999, p. 231), both the likelihood and the nature of their learning. Reading a text on a topic in a given genre and writing another one (same topic) in a different genre can construct new relationships among ideas, therefore knowledge.
- Backward search: writers set rhetorical goals, then from these they derive (backward) content subgoals, and transform their knowledge to accomplish these. This strategy is mostly used by expert writers, and taught to novices, and is closely akin to Bereiter and Scardamalia (1987)’s “knowledge-transforming“ mode. Writers using this strategy have in mind a set of ideas beforehand, and revise their texts to incorporate new ideas. In turn, and not so clearly determined by the research, writers can build knowledge from this strategy.

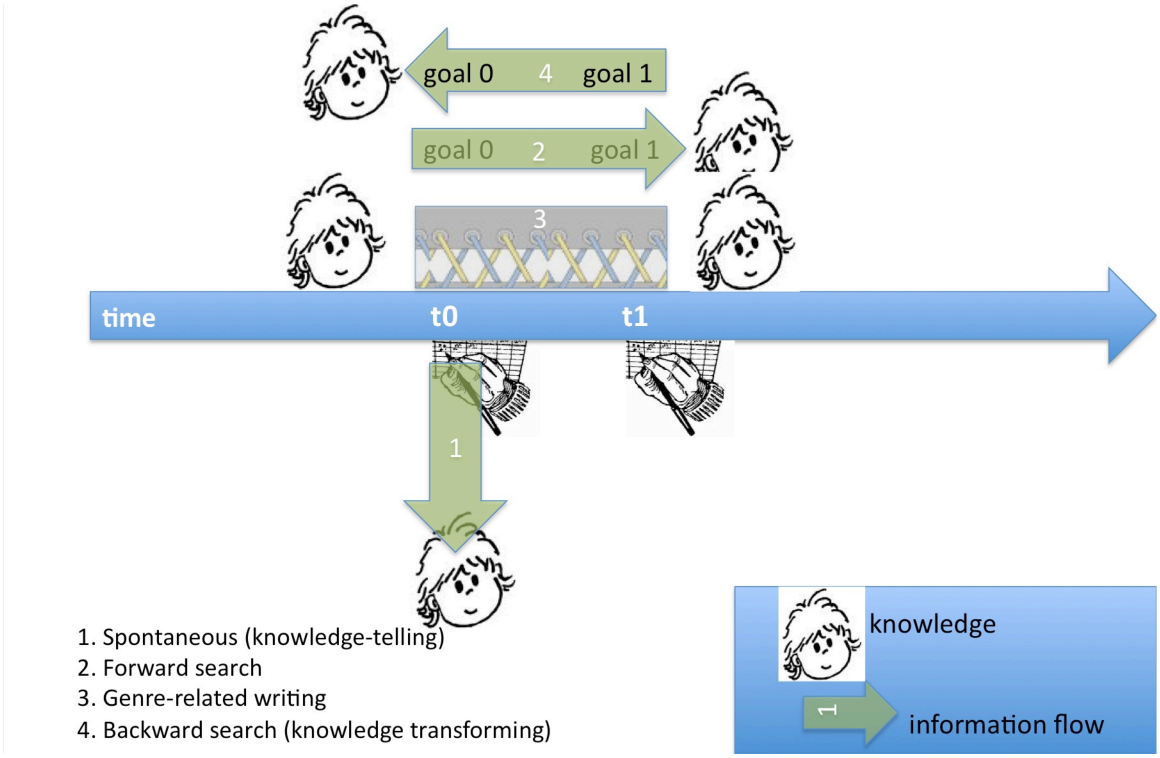

Figure 13 — Pictorial representation of four writing-to-learn strategies (after Klein, 1999).

To these four categories, which all are at an individual level, it would be worth to add a more collective one, related to the collective knowledge building cycle (from Stahl, 2006). All theses strategies are not contradictory, and can all serve learning. One could see, for each service/widget, which strategy it is intended to promote. For instance:

- *Spontaneous*: chats (service 5.1), notepads (service 5.2), blogs, all direct and immediate ways to write down ideas without reflecting on them immediately.
- Forward search (review-based): synthesis writing (service 5.2), resource annotation (service 6.1) all feedback-based activities, all ways to represent, code some form of knowledge on a relational form promote forward search.
- Genre-related: the fact that all our services promote various genres serves this strategy. Some of them (e.g., *PolyCaFe*, service 5.1) could even be used to analyse which genre is used, in letting students to “crisscross” a given topical landscape (Klein, 1999; McGinley & Tierney, 1989). The genre of each written production could be very precisely determined in order to scrutinize students’ strategies.
- Backward search: the conceptual map given in service 4.2, and the ontology-based resources (services 6.1/6.2) could serve at implementing such a strategy: the 'expert' view helps students to set up goals, as well as the resource materials.

Appendix 4 – Detailed Description of the Interface of *PolyCAFe*

The *conversation visualization* widget (presented in Figure 14 and Figure 15) offers an alternative view of the discussion compared to a normal text view enabling a better understanding of the collaborative processes and of the discussion threads. It has three areas – at the top, there is the conversation graph with the participants on the vertical axis and the utterances on the horizontal axis. The middle pane is used to display the collaboration graphics that has a value that varies during the evolution of the conversation and measures the quality of the collaboration as described in section 3. Other measures that vary after each utterance can also use this pane, like the evolution of the participants score related to the content of the utterances, or related to the social role (acceptance, agreements with the other participants seen as bonuses and reject, disagreements working as penalties). In the lower area of the widget, there are three tabs: the options tab is used to control the conversation graph's settings (zoom, colors, etc.), the second one – called "Conversation thread" – allows the user to see the utterances that are part of the same discussion thread. Thus, by clicking on any utterance in the conversation graph, the entire thread the selected utterance is part of is colored in blue in the graph, while the text messages are shown in this tab. The "Special threads" is the last tab and it is used to see how certain concepts – usually defined by the tutor when adding a new chat assignment, but others can be added – are covered in distinct parts of the conversation by various participants. This offers an alternative view on the discussion that is useful to determine if the threads are inter-animating or not and how did the participants cover each of these concepts during the whole conversation. An example is shown in Figure 15.

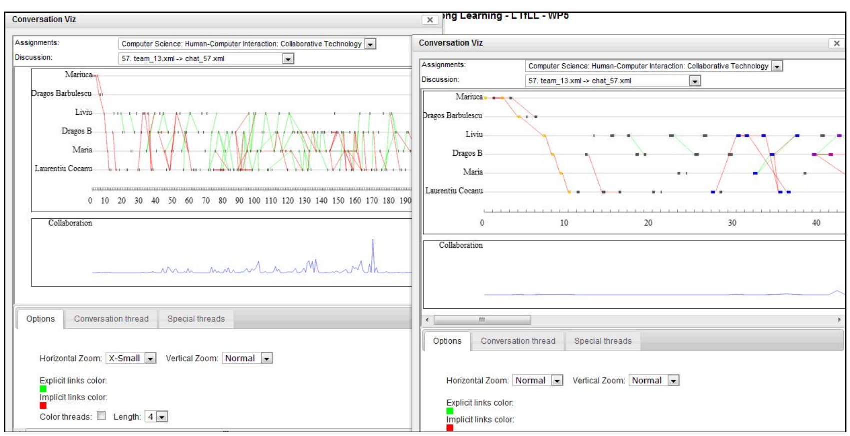

Figure 14 — The conversation visualization widget for collaborative chats, with the conversation graph at two different zoom levels showing the implicit and explicit links between utterances.

.

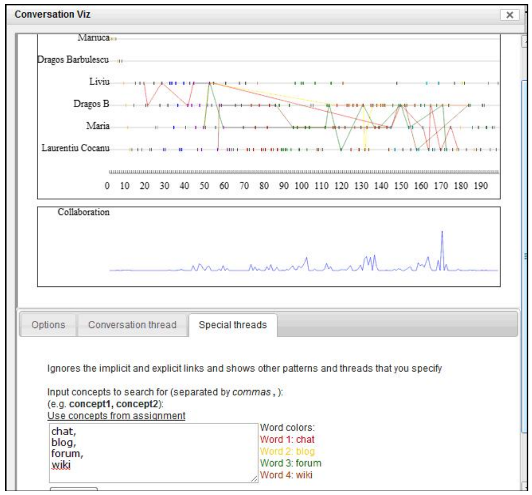

Figure 15 — The conversation visualization widget in the "Special threads" mode depicts the threads and inter-animation for the concepts defined by the tutor for the given assignment.

The *utterance feedback* widget (see Figure 16) is valuable for finding information for each message in the conversation. It also has some useful filtering capabilities based on the information determined for each utterance, such as viewing only the most important messages in the conversation, thus providing a useful summary. At this moment, the widget provides mostly content-based information about each utterance, such as the speech and argumentation acts and a score computed mainly by taking into account the lexical and semantic information of the text in the message. The only collaboration feature of the widget is the possibility to highlight the conversation thread that each utterance is part of.

Figure 16 — Utterance feedback widget (on the right) showing a discussion thread (colored in gray) next to the visualization widget with the same thread (colored in blue).

Of course, there is a need for a textual feedback about the conversation that highlights the key elements. The *conversation feedback* widget is used for this task: it provides information both on collaboration and on content of the conversation. The key aspects of the content-based feedback are: the list of most frequent concepts in the discussion determined using synonyms, hyponyms and hypernyms from *WordNet*; the list of the most relevant concepts that were discussed using LSA by measuring the cosine similarity between all used terms and the overall document vector. The feedback for collaboration (shown in Figure 17) provides information about the discussion as a collaborative process, focusing on aspects such as number of total links between utterances, explicit links, implicit links, links between different participants, percentage of utterances that contain relevant speech acts (elaborations, requests for information, accept or reject statements, etc.) and argumentation.

Similarly, the *participant feedback* widget presents information relevant to each student in the conversation. However, the analysis is done according to the content and level of the discussion and it is not compared with the entire corpus of documents. Thus, each participant is provided with statistics related to the content, the domain-related information in his posts, to the collaboration and involvement in the conversation, plus general linguistic features (such as utterance structure, fluency, spelling, etc). These are organized into 10 indicators that are computed by taking into the consideration the mean and standard variation of the score of each participant, thus dividing the feedback into 5 classes for each indicator accordingly to the variation from the mean of a given score: very good, good, normal, bad and very bad. The participant feedback widget is presented in Figure 18.

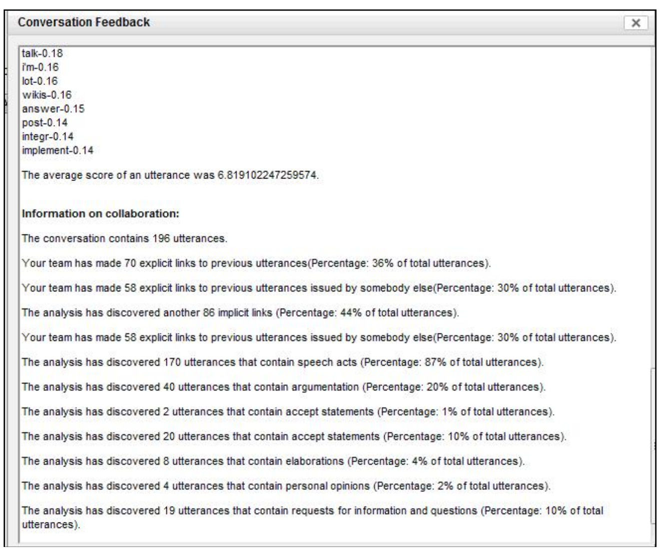

Figure 17 — The conversation feedback widget presents general information about the conversation, structured on content and collaboration feedback.

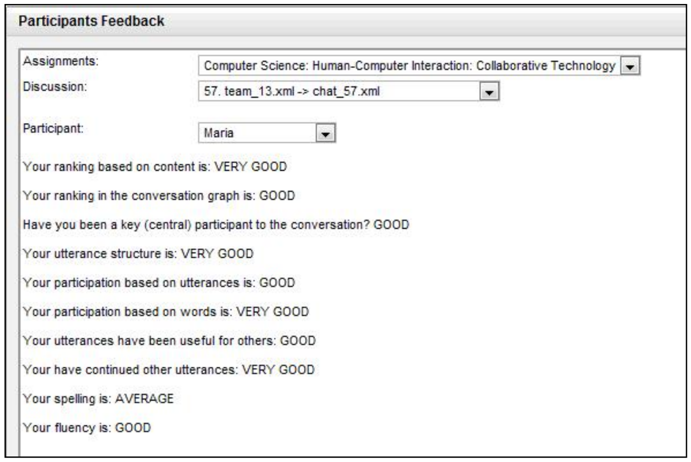

Figure 18 — The participants' feedback widget offers information about the quality of the content, the collaboration and the role in the conversation for each participant.

In certain situations, it is useful to get a quick overview and answer regarding a conversation with regard only to a given query that usually consists of one or more topics that the discussion should have covered. This feature is especially useful for discussion forums, when one wants to find the most relevant messages, but also the most relevant users, on a given topic. Nevertheless, it is as useful for running queries and getting simple and fast answers for any kind of conversation by taking into account both a lexical and a semantic score. This way, searching for "chat" in a discussion, would also return messages that contain LSA-similar concepts, like "real time", "communication" or "real-time" and "communication". One of the most powerful features of the *search conversation* widget presented in Figure 19 is that it provides a ranking not only of the utterances given the query, but also of the participants. In Figure 19, it is presented a comparative view of the search for "blog" ordered by participants ranking "cristian" first, "madalin" second and "delia" third with a very low score. From the "Special threads" view of the conversation visualization, it is easy to spot that indeed "cristian" and "madalin" were the only participants that tackled this subject, while "delia" has discussed about concepts semantically similar to it.

In Figure 19 two instances of the search conversation widget (on the right): the upper instance shows the results of the search for "blog" ordered by participant, while the latter shows the most important utterances related to this query.

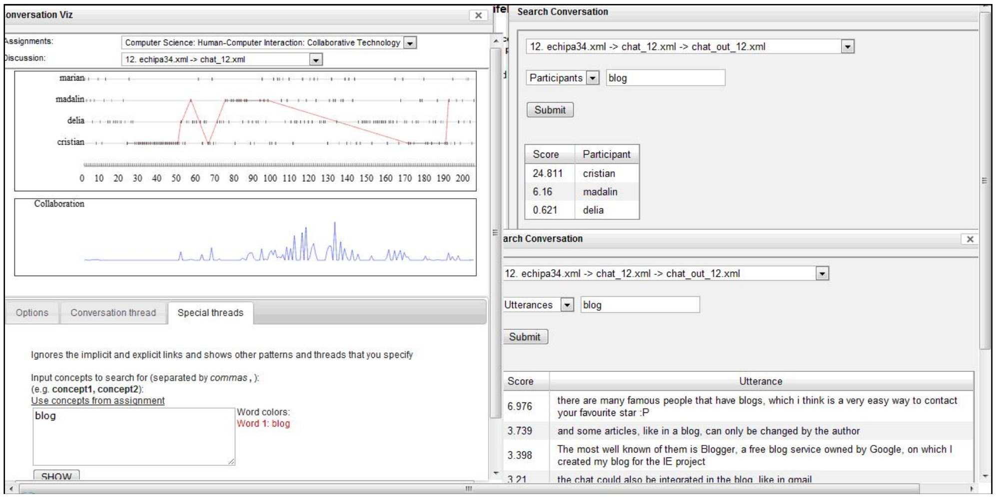

Figure 19 — Two instances of the search conversation widget

Appendix 5 – Detailed Description of the Interface of *Pensum*

Figure 20 is a screenshot of *Pensum* v. 1.5 (see also D5.2 § 2.). The upmost text field displays the selected source course, while the other one (at bottom) displays the learner's synthesis (who can toggle between editing and feedback mode). Just above the latter field are the main functionalities of *Pensum*. At any moment, the learner can click:

- On "*Feedback tolerance*" to finely tune the severity of the feedback (new functionality). We are perfectly aware of the lack of clarity of the slider in version 1.5 (its use will be explained thoroughly to participants in all verification and validation activities), but left it as is, for it is going to be addressed during widgetization.
- On "*Launch Feedback*" to launch the assessment of the synthesis;
- Additionally, the menu over the source course enables the learner to select a given text among all forming the course.
- Read the selected text in full screen mode ( )

On top of the screen, the remaining button allow the learner:

- To read/synthesize a new set of course documents ( );
- To ask for help ( );
- To go to another page to jot down transient information on the way the learner is going to understand the course ( );
- To log out ( ).

The synthesis on the bottom of Figure 20 shows some of the feedback prompts, using the conventions introduced for version 1.5.

- ∦ means there is a coherence gap detected between two sentences;
- || means there is a coherence gap detected between two sentences, but the user has questioned it;
- a sentence in red font means that this sentence is semantically far from these of the course texts and could be removed (off-topic).
- a sentence highlighted in blue means that it has firstly been assessed as "off-topic" by *Pensum*, nonetheless this feedback has been rejected by the learner. The learner can decide to justify this by linking it to a given sentence of the course text. An asterisk designates linked sentences (a tooltip over a linked sentence of the synthesis indicates to which course text it is linked). All linked sentences to those of the currently displayed course text can be toggled all at once using or one by one by clicking on it. Each set of linked sentence is then associated with a given color (cf. Figure 20) and both visible texts decorated. It is noteworthy that any French speaker will realize that the "learner" has clearly mislinked the blue highlighted sentences in Figure 20 (whereas the green highlighted sentences are coherent).
- reversely, a sentence highlighted in the course text has been linked to the synthesis. A questioned feedback displayed in blue font in the text is in fact present in the synthesis (linked) whereas if it is not link, this is because the

user considers it irrelevant and it is displayed in gray, in order to remind them, they have considered the sentences as not important.

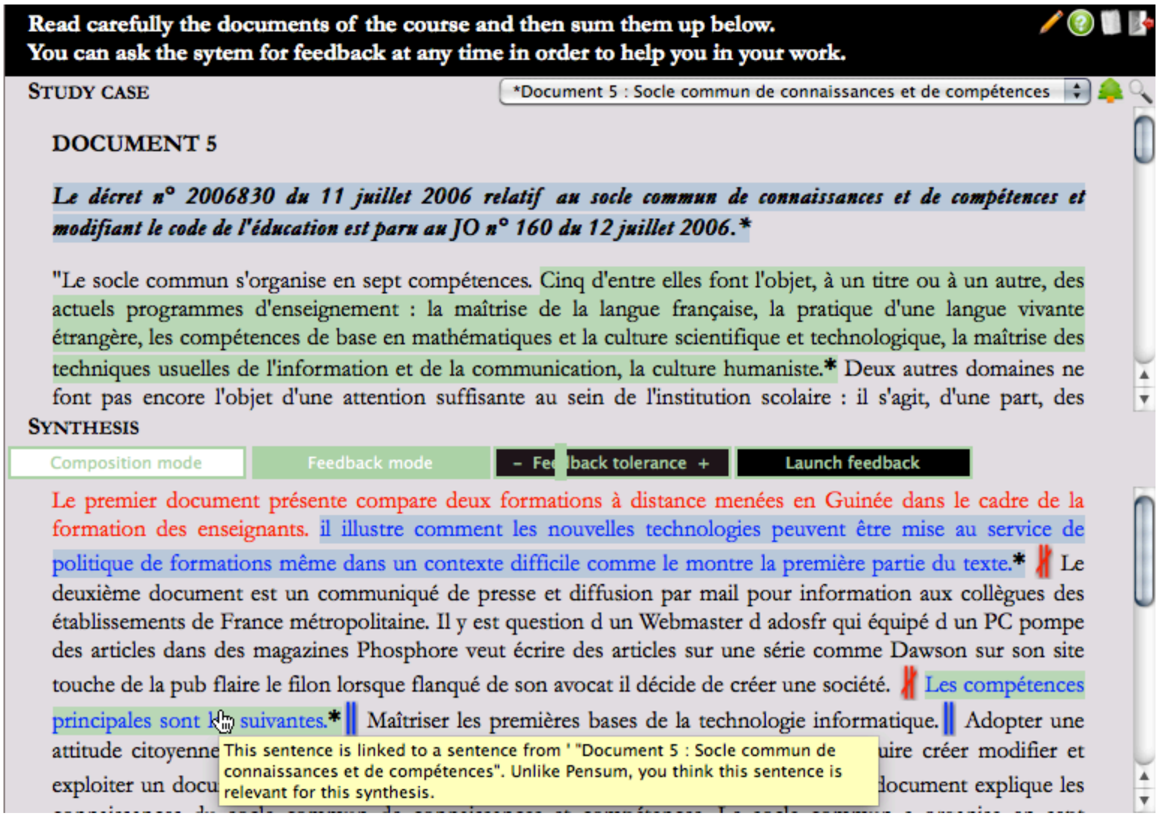

Figure 20 – Screenshot of *Pensum* v. 1.5. The upmost field contains the course text. The bottom field contains the synthesis.

Figure 21 depicts the main contextual menus allowing the learner to reject *Pensum*'s feedback. A different context menu handles each type of feedback. The enabled/disabled items in these context menus are different depending on whether the feedback is *questioned* or not. At any moment the learner can:

- *Question/reject pertinence feedback*, using a right click on the flagged course sentence, which the system has recognized as not taken into account in the synthesis (context menu):
  - By considering the content of the sentence irrelevant; should a link reach an accepted feedback, it will be automatically set to 'questioned';
  - By linking it to the sentence of the synthesis that renders its main idea (see upper right and lower screenshots in Figure 21);
- *Question/reject off-topic feedback*, using a right click on the flagged synthesis sentence, that have been considered irrelevant to the topic (context menu):
  - By considering the sentence relevant;
  - By explicitly linking it with the corresponding course sentence; should a link reach an accepted feedback, it will be automatically set to 'questioned' (see lower screenshot in Figure 21).

- *Decide to reconsider questioning the feedback* (see middle and upper-left screenshots in Figure 21). The link created from this sentence is erased (unlike links to this sentence that originate from the other type of feedback—absence, resp. relevance—, which remain until the feedback it originates from is accepted);
- *reject/question a coherence gap feedback*, in indicating (right click on the icon) that the two given sentences do not have an important coherence gap (in the upmost left screenshot the user "unquestions" such a rejected feedback).

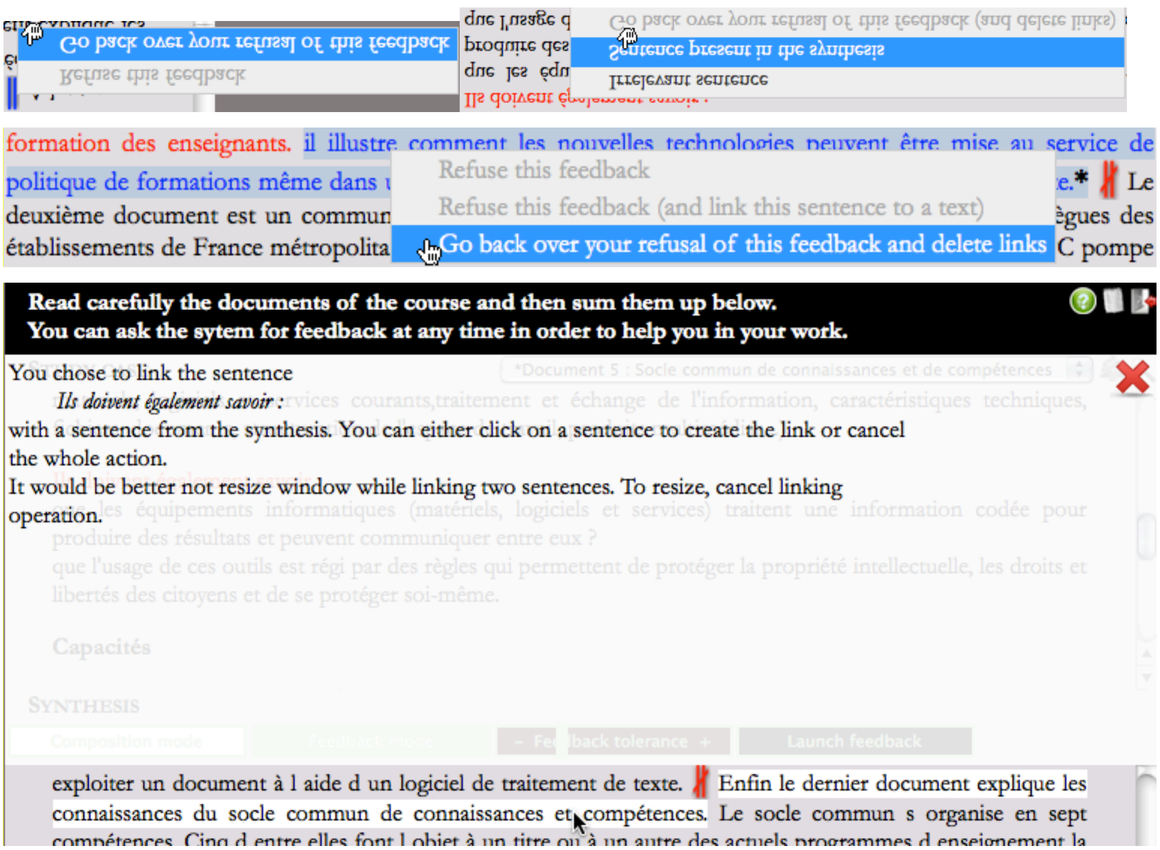

Figure 21 – Screenshot of the contextual menus of *Pensum* v. 1.5. Upmost left: for rejecting coherence feedback. Upmost right: for rejecting pertinence feedback. Middle: Accepting a previously questioned off-topic feedback. Bottom: creating a link between a course sentence and one of the synthesis'.

Appendix 6 – Preliminary Verification Results of *PolyCAFe*

Speech Act Classification

Tables 6 and 7 show the complete results of using a machine learning approach with *TagHelper* for labeling speech and argumentation acts in a chat corpora. The training corpus consisted of six files with dialogs in English summing up 2408 utterances, while the testing corpus had three files summing up 1215 utterances. There were cases in which certain utterances were not linked to previous ones; hence all three columns were labeled "N/A" – Not Applicable. Firstly, it should be noted that the precision of this experiment must be considered of a greater importance than its recall, because the conversations from *ConcertChat* could not be treated as a whole. The reason was that, for those conversations in which the referred utterances were quite a few, the un-referred utterances were labeled too, therefore the model learned after training was disturbed by the un-referred, but annotated utterances which were not in a vast amount in the test corpus. Secondly, the results for precision and recall metrics presented in the tables below were influenced by a limitation of *TagHelper* system which does not parse multiple values for a column, but considers them a whole. As an illustration, the annotation consisting of three different labels "confirm, deliberate, suggest" is used as a "block" in the test corpus providing rather poor results.

Table 6 – Precision and recall for "SPEECH ACT" column

Label	Precision	Recall
Thank *	-	-
Greet *	-	-
Introduce *	-	-
Bye *	-	-
Request-comment	85%	26%
Suggest	56%	73%
Reject	35%	31%
Accept	46%	64%
Request-suggest	100%	18%
Init *	-	-
Give-reason	42%	23%
Feedback *	-	-
Deliberate		
Confirm	44%	48%
Clarify *	-	-
Digress *	-	-
Motivate	60%	43%
Garbage *	-	-
Completion	21%	6%
Repeat-rephrase	N/A	0%

** Labels marked with star were used for rather few utterances or were not used at all in the correct coding of test corpus; therefore they are not relevant for this experiment*

Table 7 – Precision and recall for "ARGUMENTATION" column

Category	Precision	Recall
Reason	N/A	0%
Condition	N/A	0%
Consequence	100%	12%
Contrast	N/A	0%
Elaboration	57%	31%
Claim	75%	4%
Question	93%	30%
Answer	8%	23%
Rebuttal	43%	42%
Cohesion	38%	57%

Community of Inquiry Model Classification

The rule-based approach described in Appendix 1 for the classification of forum messages according to the community of inquiry model was tested on a corpus of discussion threads annotated by the University of Manchester. The results for precision are shown in the figure below: test 1 contains all the messages (close to 400), while test 2 contains only a subset of 40 messages.

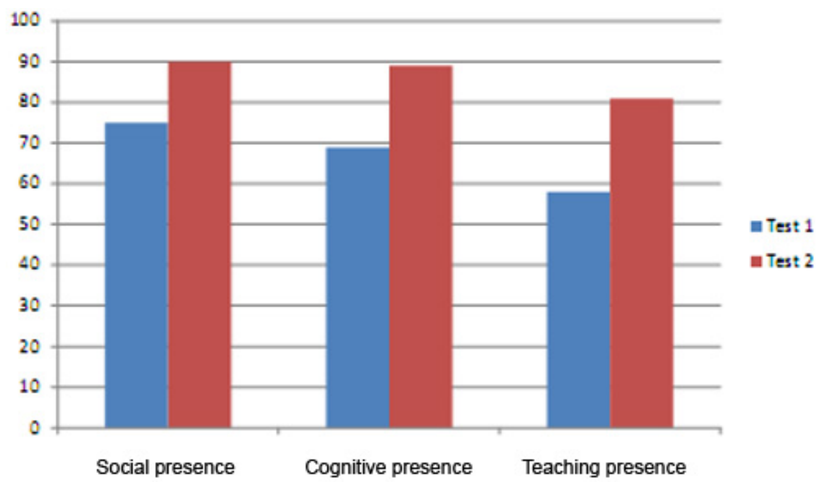

Figure 22 – Precision for the community of inquiry classification task using linguistic rules

On the other hand, the results for the machine learning classification using various algorithms from WEKA are displayed in Figure 23. Although, the results for decision tables and trees seem to be the best, it should be taken into consideration that these are over-fitted on the data and the Bayesian classifier provides more accurate results in this case.

Decision Trees		CE	CI	CR	CTE	SEE	SGC	SOC	TBU	TDI	TIM
Decision Table	Accuracy	98.275	82.758	91.379	87.931	36.206	48.275	60.344	89.655	86.206	94.827
	Precizie	0.983	0.829	0.892	0.876	0.263	0.319	0.499	0.88	0.846	0.946
JRip	Accuracy	100	68.965	87.931	79.31	44.827	43.103	51.724	81.034	87.931	86.206
	Precizie	1	0.693	0.868	0.793	0.379	0.268	0.432	0.765	0.832	0.877
Ridor	Accuracy	86.206	79.31	89.655	82.758	29.31	39.655	63.793	82.758	89.655	93.103
	Precizie	0.862	0.793	0.834	0.819	0.293	0.335	0.588	0.828	0.834	0.936
Rules		CE	CI	CR	CTE	SEE	SGC	SOC	TBU	TDI	TIM
REPTree	Accuracy	86.206	82.758	91.379	75.862	43.103	43.103	65.517	87.931	91.379	91.379
	Precizie	0.853	0.828	0.835	0.704	0.228	0.232	0.523	0.773	0.835	0.922
J48	Accuracy	68.965	56.896	89.655	79.31	34.482	25.862	68.965	82.758	87.931	93.103
	Precizie	0.644	0.569	0.834	0.783	0.317	0.22	0.662	0.767	0.832	0.931
Decision Stump	Accuracy	79.31	67.241	91.379	70.689	48.275	46.551	67.241	93.103	84.482	94.827
	Precizie	0.629	0.735	0.835	0.634	0.402	0.241	0.53	0.926	0.829	0.946
Other		CE	CI	CR	CTE	SEE	SGC	SOC	TBU	TDI	TIM
LibSVM	Accuracy	79.31	62.069	91.379	77.586	48.275	48.275	46.551	87.931	91.379	86.206
	Precizie	0.629	0.631	0.835	0.602	0.233	0.233	0.217	0.773	0.835	0.743
Bayes	Accuracy	79.31	70.689	89.655	82.758	48.275	48.275	67.241	87.931	87.931	93.103
	Precizie	0.629	0.707	0.834	0.828	0.233	0.237	0.53	0.773	0.832	0.927

Figure 23 – Precision and recall for the community of inquiry classification task using machine learning

POS Tagging Chat Conversations in Romanian

First, a POS model for novels has been build starting from the translation of the novel "1984", by George Orwell. The tag set consisted of 154 tags that included all the Romanian POS and some of their grammatical categories. Then, we have built the new model in a semi-supervised manner. The first step was to apply the already build model on a corpus of chats in order to get an idea on accuracy and precision of that model. Using this approach, the tagger assigned a tag to every word from these chats, using the probabilities learnt from the "1984"model. Naturally, a large percent of the words were incorrectly tagged. As we wanted to create a new model that would be suitable for chats, we manually corrected the wrong tags, and the obtained corpus has been considered the new input text for the application. The corpus of chats that has been used for this step consisted of 5 chats in *ConcertChat* format, each having 4-5 participants. These chats consisted of approximately 700 utterances and 15,000 words. The improvements of using this chat POS model are shown below. The testing of the new model was made on a corpus consisting of another four chats that debated the same topic and contained the same number of participants. The corpus contained about 500 utterances and 12,000 words.

Table 8 – Comparative precision of the two models

Tag	Novel Model Precision (%)	Chat Model Precision (%)	Improvement (value and %)
Overall	66	73	7 (10.6 %)
Substantives	54	69	15 (27.7 %)
Numerals	66	69	3 (4.54 %)
Adjectives	59	53	-6 (-10.1 %)
Pronouns	66	67	1 (1.5 %)
Verbs	68	77	9 (13.2 %)
Abbreviations	35	81	46 (131.4 %)
Prepositions	85	82	-3 (-3.5 %)
Conjunctions	68	91	23 (33.8 %)
Prefixes	33	50	17 (51.5 %)
Articles	70	83	13 (18.5 %)
Adverbs	62	86	24 (38.7 %)
Interjections	56	89	33 (58.9 %)
Auxiliary	63	76	13 (20.6 %)

Table 9 – Comparative recall of the two models

Tag	Novel Model Recall (%)	Chat Model Recall (%)	Improvement (value and %)
Overall	59	63	4 (6.7 %)
Substantives	48	54	6 (12.5 %)
Numerals	66	43	-23 (-34.8 %)
Adjectives	48	56	8 (16.6 %)
Pronouns	61	63	2 (3.2 %)
Verbs	60	61	1 (1.6 %)
Abbreviations	26	58	32 (123 %)
Prepositions	82	90	8 (9.7 %)
Conjunctions	68	84	16 (23.5 %)
Prefixes	100	100	0 (0 %)
Articles	63	73	10 (15.8 %)
Adverbs	80	77	-3 (-3.7 %)
Interjections	66	34	-32 (-48.4 %)
Auxiliary	54	52	-2 (-3.7 %)

Appendix 7 – The list of publications of WP5 *(as of 11th October, 2010)*

WP 5.1 Publications

1. Trausan-Matu, S. (**2010**). The Polyphonic Model of Hybrid and Collaborative Learning, In Wang, F., L., Fong., J., Kwan, R.C., *Handbook of Research on Hybrid Learning Models: Advanced Tools, Technologies, and Applications*, Information Science Publishing, Hershey, New York, pp 466-486.
2. Trausan-Matu, S. (**2010**). Automatic Support for the Analysis of Online Collaborative Learning Chat Conversations, in Philip M. Tsang, Simon K. S. Cheung, Victor S. K. Lee, Ronghuai Huang (Eds.): *Hybrid Learning, Third International Conference, ICHL 2010*, Beijing, Lecture Notes in Computer Science 6248, Springer, pp. 383-394 (Invited Talk)
3. Trausan-Matu, S., Rebedea, T. (**2010**), A Polyphonic Model and System for Inter-animation Analysis in Chat Conversations with Multiple Participants, in A. Gelbukh (Ed.), *CICLing 2010*, LNCS 6008, Springer, pp. 354–363
4. Dascalu, M., Rebedea, T., Trausan-Matu, S. (**2010**), A Deep Insight in Chat Analysis: Collaboration, Evolution and Evaluation, Summarization and Search. in Darina Dicheva and Danail Dochev (Eds.), *Proceedings of AIMSA 2010*, Lecture Notes in Computer Science 6304, Springer, pp. 191-200.
5. Trausan-Matu, S., Mihai Dascalu, Rebedea, T., Alexandru Gartner (**2010**), Corpus de conversatii multi-participant si editor pentru adnotarea lui, *Revista Romana de Interactiune Om-Calculator*, Vol.3, No.1, pp. 53-64, ISSN 1843-4460.
6. Rebedea, T., Trausan-Matu, S., Costin Chiru (**2010**), Automatic Feedback System for Collaborative Learning Using Chats and Forums, *Proceedings of CSEDU 2010 - International Conference on Computer Supported Education*, Valencia, pp. 358-363.
7. Rebedea, T., Dascalu, M., Trausan-Matu, S., Dan Banica, Alexandru Gartner, Costin Chiru and Dan Mihaila (**2010**), Overview and Preliminary Results of Using PolyCAFe for Collaboration Analysis and Feedback Generation, in *Proceedings of ECTEL 2010*, LNCS 6283, Springer, pp. 420-425
8. Rebedea, T., Mihai Dascalu and Trausan-Matu, S. (**2010**). PolyCAFe: Polyphony-Based System for Collaboration Analysis and Feedback Generation, in Stefan Trausan-Matu, Philippe Dessus (eds.), *Proceedings of the Second Workshop on Natural Language in Support of Learning: Metrics, Feedback and Connectivity,* ISBN 978-973-755-634-9, Ed. MatrixRom, Bucharest, pp. 21-34
9. Trausan-Matu, S., Traian Rebedea (**2010**), Inter-animation and polyphony in computer-supported collaborative learning, in *Annals of the Academy of Romanian Scientists*, Series on Science and Technology of Information, vol.3, nr.1, pp. 113-124.
10. Trausan-Matu, S., Traian Rebedea, Mihai Dascalu (**2010**), Analysis of discourse in collaborative Learning Chat Conversations with Multiple Participants, in Tufis, D., Corina Forascu (eds.), *Multilinguality and Interoperability in Language Processing with Emphasis on Romanian*, Editura Academiei, pp. 313-330, (2010)
11. Trausan-Matu, S., Posea, V., Rebedea, T. and Chiru, C., (**2009**). Using the Social Web to Supplement Classical Learning, in M. Spaniol et al. (Eds.): Proceedings ICWL 2009, Lecture Notes in Computer Science (Vol. 5686, pp. 386–389), Springer
12. Trausan-Matu, S. & Rebedea, T. (**2009**), Polyphonic Inter-Animation of Voices in VMT, in Stahl.G. (Ed.), *Studying Virtual Math Teams* (pp. 451–473). Boston, MA: Springer US.
13. Rebedea, T. & Trausan-Matu, S. (**2009**). Computer-Assisted Evaluation of CSCL Chat Conversations. In A. Dimitracopoulou, C. O'Malley, D. Suthers, & P. Reiman, *Computer Supported Collaborative Learning Practices - CSCL2009 Community Events Proceedings* (pp. 183-185). Rhodes: Lulu.
14. Dascalu, M. & Trausan-Matu, S. (**2009**) Ch.A.M.P. – Sistem pentru evaluarea şi modelarea contribuţiei participanţilor la un Chat*, Revista Română de Interacţiune Om-Calculator 2*, 131-146.
15. Trausan-Matu, S., Posea, V. & Rebedea, T. (**2009**), Utilizarea instrumentelor informatice colaborative de interfaţare în învăţământ, în Interacţiune Om-Calculator, volumul de lucrări ale celei de a şasea. In D. Gorgan & A. M. Guran (Eds.), *Conferinţe Naţionale de Interacţiune Om-Calculator ROCHI 2009* (pp. 19–24). Bucharest: Ed. MATRIX ROM , ISSN 1843-4460.

16. Mihai Dascalu, Stefan Trausan-Matu (**2009**), C.H.A.M.P. – A Program for Chat Modelling and Assesment, *Annals of the Academy of Romanian Scientists*, Series on Science and Technology of Information, vol.2, nr. 1, pp. 95-106.
17. Dascalu, M., Chioasca, E., & Trausan-Matu, S. (**2008**). ASAP- An Advanced System for Assessing Chat Participants. In *Artificial Intelligence: Methodology, Systems, and Applications*, Lecture Notes in Computer Science (Vol. 5253, pp. 58–68). Berlin: Springer.
18. Rebedea, T., Trausan-Matu, S., & Chiru, C. (**2008**). Extraction of Socio-semantic Data from Chat Conversations in Collaborative Learning Communities. In *Times of Convergence. Technologies Across Learning Contexts*, Lecture Notes in Computer Science (Vol. 5192, pp. 366–377). Berlin: Springer.
19. Dascalu, M., Chioasca, E.-V. & Trausan-Matu, S. (**2008**), ASAP – Sistem avansat de evaluare a participantilor la un chat, in A. Marhan, I. Juvina (eds.), *Proceedings la a 4-a Conferinta Nationala de Interactiune Om-Calculator (RoCHI 2008)* (pp. 105–112). Bucharest: MATRIX ROM.

WP 5.2 Publications

1. Dessus, P. (**2008**, 14th November). Learning to summarize and summarizing for learning: Some computer-based supports. *Symposium around the inaugural address of Peter Sloep*. OUNL: Heerlen. [Invited Talk].
2. Dessus, P. (**2009**). An overview of LSA-based systems for supporting learning and teaching. In V. Dimitrova, R. Mizoguchi, B. du Boulay & A. Graesser (Eds.), *Artificial Intelligence in Education. Building learning systems that care: From knowledge representation to affective modelling (AIED2009)* (pp. 157–164). Amsterdam: IOS Press.
3. Dessus, P. (**2009**, 26th August). The four pillars of educational ICT research. *Education & Technology Project Summer School*. Strasbourg (France) [Invited Talk].
4. Dessus, P. (**2010**, 1st June). Un EIAH d'aide à la rédaction de synthèses de cours dans un contexte d'enseignement universitaire à distance [An ILE that helps write course synthèses in a universty Distance Learning context]. *Première Journée du Pôle Grenoble Cognition*. Grenoble. [Invited Talk]
5. Dessus, P. (**2010**). Des théories de l'apprentissage pour concevoir des environnements d'apprentissage informatisés [Psychological theories for Instructional Design]. In B. Charlier & F. Henri (Eds.), *Apprendre avec les technologies*. Paris: P.U.F., coll. Apprendre.
6. Dessus, P., Mandin, S. & Zampa, V. (**2008**). What is teaching? Cognitive-based tutoring principles for the design of a learning environment. In S. Tazi & K. Zreik (Eds.), *Common Innovation in e-learning, Machine learning and Humanoid (ICHSL'6)* (pp. 49–55). Paris: IEEE/Europia.
7. Villiot-Leclercq, E. & Dessus, P. (**2009**, 23 June). Les contraintes de l'activité de tutorat à distance [Activity constraints in distance tutoring]. In V. Guéraud & P. Leroux (Eds.), *Atelier "Instrumentation des activités du tuteur" de la conférence EIAH'09* (pp. 45–52). Le Mans.
8. Villiot-Leclercq, E., Mandin, S., Dessus, P. & Zampa, V. (**2010**, 5-7th July). Helping students understand courses through written syntheses. An LSA-based online advisor [Short paper]. *Proc. Tenth IEEE Int. Conf. on Advanced Learning Technologies (ICALT 2010)(pp.* 341–343). Sousse (Tunisia): IEEE.
9. Zampa, V. & Dessus, P. (**2010**, 6-8th December). Écrire des résumés pour apprendre des cours : un système d'aide à l'apprentissage autorégulé [Writing summaries to learn courses: an SRL-based system]. In *7e Colloque Technologies de l'Information et de la Communication pour l'Enseignement (TICE 2010)*. Nancy.

WP 5.1 and 5.2 Joint Publications

1. Dascalu, M., Trausan-Matu, S. & Dessus, P. (**2010**, 2-3rd September). Evaluarea si sumarizarea automata a conversatiilor chat. *7-a Conferinta Nationala de Interactiune Om-Calculator (RoCHI 2010)*. Bucharest (Romania).
2. Dascalu, M., Trausan-Matu, S. & Dessus, P. (**2010**). Utterances assessment in chat conversations. *Research in Computing Science, 46,* 323–334. [From Proc. Conf. CICLing 2010, Iasi, Romania, 21-27th March].
3. Dessus, P. & Trausan-Matu, S. (**2010**). Implementing Bakhtin's dialogism theory with NLP techniques in distance learning environments. In S. Trausan-Matu, & P. Dessus (Eds.), *In Proc. 2nd Workshop on*

Natural Language Processing in Support of Learning: Metrics, Feedback and Connectivity (NLPsL 2010)(pp. 11–20). Bucharest: Matrix Rom.

4. Dessus, P., Trausan-Matu, S., Mandin, S. Rebedea, T., Zampa, V., Dascalu, M. & Villiot-Leclercq, E. (**2010**). Assessing writing and collaboration in learning: Methodological issues. *International Reports on Socio-Informatics, 7*(1), 86–94. [From Proc. Workshop "Analysing the quality of collaboration in task-oriented computer-mediated interactions", held in conjunction to the 9th International Conference on the Design of Cooperative Systems (COOP 2010), Aix-en-Provence, France, 18 May]
5. Dessus, P., Trausan-Matu, S., Zampa, V., Rebedea, T., Mandin, S. & Dascalu, M. (**2009**). Vers un environnement-tuteur d'apprentissage dialogique [Towards a Dialogical Learning ILE]. In C. Develotte, F. Mangenot & E. Nissen (Eds.) *2e Colloque Echanger pour Apprendre en Ligne (EPAL'09)*. Grenoble: LIDILEM/INRP.

Whole Project-related publications

1. Dessus, P., Trausan-Matu, S., van Rosmalen, P. & Wild F. (**2009**). Proc. of Natural Language Processing in support of Learning: Metrics, feedback and connectivity. *Workshop held in collaboration with the 14th Int. Conf. on Artificial Intelligence in Education (AIED'09)* (48 p.).
2. Trausan-Matu, S. & Dessus, P. (Eds.). (**2010**). *2nd Workshop on Natural Language Processing in Support of Learning: Metrics, Feedback and Connectivity (NLPsL 2010)*. Bucharest (Romania): Matrix Rom Publisher, 97 p.

8. References

Adams, P. H., & Martell, C. H. (2008). *Topic Detection and Extraction in Chat*. Paper presented at the Proceedings of the 2008 IEEE International Conference on Semantic Computing.

Anderson, J. R. (1985). *Cognitive psychology and its implications*. New York: Freeman.

Avouris, N., Fiotakis, G., Kahrimanis, G., & Margaritis, M. (2007). Beyond Logging of Fingertip Actions: Analysis of Collaborative Learning Using Multiple Sources of Data. *Journal of Interactive Learning Research, 18*, 231–250.

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays*. Austin: The University of Texas Press.

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

Bakhtin, M. M. (1993). *Problems of Dostoevsky's poetics*. Minneapolis: University of Minnesota Press.

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*(1), 29–58.

Bax, S. (2003). CALL–Past, present and future. *System, 31*(1), 13–28.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Illsdale: Erlbaum.

Boekaerts, M. (1997). Self-regulated learning : a new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*(2), 161–186.

Bonk, C. J., & Dennen, V. (2003). Frameworks for research, design, benchmarks, training and pedagogy in web-based distance education. In M. G. Moore & W. G. Anderson (Eds.), *Handbook of distance education* (pp. 331–348). Mahwah: Erlbaum.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion online writing service. *AI Magazine, 25*(3), 27–36.

Chiru, C., Rebedea, T., & Ionita, M. (2010). *Chat-Adapted POS Tagger for Romanian Language.* Paper presented at the The Second Workshop on Natural Language Processing in Support of Learning: Metrics, Feedback and Connectivity.

Dascalu, M., Rebedea, T., & Trausan-Matu, S. (2010). *A deep insight in chat analysis: Collaboration, evolution and evaluation, summarization and search.* Paper presented at the AIMSA 2010.

Dessus, P., & Trausan-Matu, S. (2010). Implementing Bakhtin's dialogism theory with NLP techniques in distance learning environments. In S. Trausan-Matu & P. Dessus (Eds.), *Proc. 2nd Workshop on Natural Language Processing in Support of Learning: Metrics, Feedback and Connectivity (NLPsL 2010)* (pp. 11–20). Bucharest: Matrix Rom.

Dessus, P., Trausan-Matu, S., Mandin, S., Rebedea, T., Zampa, V., Dascalu, M., et al. (2010). *Assessing writing and collaboration in learning: Methodological issues*. Paper presented at the Workshop "Analysing the quality of collaboration in task-oriented computer-mediated interactions", held in conjunction to the 9th International Conference on the Design of Cooperative Systems (COOP 2010).

Dong, A. (2006). Concept formation as knowledge accumulation: A computational linguistics study. *Artif. Intell. Eng. Des. Anal. Manuf., 20*(1), 35–53.

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

Dowell, J., Tscholl, M., Gladisch, T., & Asgari-Targhi, M. (2009). Argumentation scheme and shared online diagramming in case-based collaborative learning. *Computer Support for Collaborative Learning*, 567–575.

Dulceanu, A. (2010). *Speech Acts Identification Using TagHelper.* Paper presented at the The Second Workshop on Natural Language Processing in Support of Learning: Metrics, Feedback and Connectivity.

Dyke, G., Lund, K., & Girardot, J.-J. (2009). Tatiana: an environment to support the CSCL analysis process. *Computer Support for Collaborative Learning*, 58–67.

Emig, J. (1977). Writing as a mode of learning. *College Composition & Communication, 28*(2), 122–128.

Garlan, D., & Shaw, M. (1993). An Introduction to software architecture. In V. Ambriola & G. Tortora (Eds.), *Advances in Software Engineering and Knowledge Engineering* (Vol. 2). Singapore: World Scientific Publishing Company.

Garrison, D. R., & Anderson, T. (2003). *E-Learning in the 21st century. A framework for research and practice*. New York: Routledge.

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *Internet and Higher Education, 2*(2-3), 87–105.

Harrer, A., Hever, R., & Ziebarth, S. (2007). Empowering researchers to detect interaction patterns in e-collaboration. *Frontiers in Artificial Intelligence and Applications; Vol. 158*, 503–510.

Hmelo-Silver, C. E., Chernobilsky, E., & Masto, O. (2006). *Representations for Analyzing Tool-mediated Collaborative Learning.* Paper presented at the Proceedings of ICLS 2006, New Brunswick.

Hübner, S., Nückles, M., & Renkl, A. (2006). Fostering the cycle of self-regulation in writing learning protocols. In G. Clarebout & J. Elen (Eds.), *Avoiding simplicity, confronting complexity. Advances in studying and designing (computer-based) powerful learning environments* (pp. 155–164). Rotterdam: Sense Publishers.

Joshi, M., & Rose, C. P. (2007). *Using transactivity in conversation summarization in educational dialog.* Paper presented at the Proceedings of the SLaTE Workshop on Speech and Language Technology in Education.

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

Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review, 11*(3), 203–270.

Kontostathis, A., Edwards, L., Bayzick, J., McGhee, I., Leatherman, A., & Moore, K. (2009). *Comparison of Rule-based to Human Analysis of Chat Logs.* Paper presented at the 1st International Workshop on Mining Social Media Programme, Conferencia de la Asociación Española para la Inteligencia Artificial CAEPIA 2009

Koschmann, T. (1999). *Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration.* Paper presented at the Computer Support for Collaborative Learning (CSCL'99), Palo Alto.

Kumar, R., Chaudhuri, S., Howley, I., & Rosé, C. P. (2009). VMT-Basilica: an environment for rapid prototyping of collaborative learning environments with dynamic support. *Computer Support for Collaborative Learning*, 192–194.

Lajoie, S., & Azevedo, R. (2006). Teaching and learning in technology-rich environments. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2e ed., pp. 803–821). Mahwah: Erlbaum.

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. (1998a). An introduction to Latent Semantic Analysis. *Discourse Processes, 25*(2/3), 259–284.

Landauer, T. K., Laham, D., & Foltz, P. W. (2000). The Intelligent Essay Assessor. *IEEE Intelligent Systems, 15*(5), 27–31.

Langer, J. A., & Applebee, A. N. (1987). *How writing shapes thinking: A study of teaching and learning*. Urbana (IL): National Council of Teachers of English.

Law, N., Lu, J., Leng, J., Yuen, J., & Lai, M. (2008). *Understanding Knowledge Building from Multiple Perspectives.* Paper presented at the Proceedings of Workshop on Interaction Analysis, ICLS 2008, Utrecht.

Lemaire, B. (2009). *Limites de la lemmatisation pour l'extraction de significations.* Paper presented at the Proc. 9es Journées Internationales d'Analyse Statistique des Données Textuelles (JADT 2009), Lyon (France).

Lemaire, B., Denhière, G., Bellissens, C., & Jhean-Larose, S. (2006). A computational model for simulating text comprehension. *Behavior Research Methods, Instrument and Computers, 38*(4), 628–637.

Linell, P. (2009). *Rethinking language, mind, and world dialogically*. Greenwich: Information Age Publications.

Lukashenko, Graudina, & Grundspenkis. (2007). *Computer-based plagiarism detection methods and tools: An overview.* Paper presented at the Int. Conf. on Computer Systems and Technologies (CompSysTech'07).

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

McGinley, W., & Tierney, R. J. (1989). Traversing the topical landscape. *Written Communication, 6*, 243–269.

Murray, L., & Barnes, A. (1998). Beyond the "wow" factor – Evaluating multimedia language learning software from a pedagogical viewpoint. *System, 26*(2), 249–259.

O'Reilly, T. (2007). What Is Web 2.0? Design patterns and business models for the next generation of software. *International Journal of Digital Economics, 65*, 17–37.

Olive, T., Rouet, J.-F., François, E., & Zampa, V. (2008). Summarizing digital documents : Effects of alternate or simultaneous window display. *Applied cognitive psychology, 22*(4), 541–558.

Page, E. B., & Paulus, D. H. (1968). *Analysis of essays by computer. Predicting Overall Quality*: U.S. Department of Health, Education and Welfare.

Rabardel, P. (2002). *People and Technology -- a cognitive approach to contemporary instruments*. Paris: University of Paris-VIII.

Rebedea, T., Dascalu, M., Trausan-Matu, S., Banica, D., Gartner, A., Chiru, C., 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.), *Sustaining TEL: From Innovation to Learning and Practice* (Vol. 6383, pp. 420–425): Springer Berlin / Heidelberg.

Reed, E. S. (1991). Cognition as the cooperative appropriation of affordances. *Ecological Psychology, 3*(2), 135–158.

Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., et al. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 3*(3), 237–271.

Rowley, K., & Meyer, N. (2003). The effect of a computer tutor for writers on student writing achievement. *Journal of Educational Computing Research, 29*(2), 169–187.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144.

Schmidt, A. P., & Stone, T. K. M. (1993). *Detection of topic change in IRC chat logs*.

Segev-Miller, R. (2004). Writing from sources: the effect of explicit instruction on college students' processes and products. *L1-Educational Studies in Language and Literature, 4*(1), 5–33.

Sfard, A. (2000). On reform movement and the limits of mathematical discourse. *Mathematical Thinking and Learning, 2*(3), 157–189.

Shannon, C. (1951). Prediction and entropy of printed English. *The Bell System Technical Journal, 30*(50–64).

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

Stahl, G. (2009). *Studying Virtual Math Teams*. New York: Springer.

Suthers, D., Dwyer, N., Medina, R., & Vatrapu, R. (2007). *A framework for eclectic analysis of collaborative interaction.* Paper presented at the Proceedings of Computer Support for Collaborative Learning 2007.

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

Teplovs, C. (2008). The Knowledge Space Visualizer: A Tool for Visualizing Online Discourse.

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28*, 129–160.

Toulmin, S. (1958). *The Uses of Arguments*: Cambridge Univ. Press.

Trausan-Matu, S. (2010). 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.), *Hybrid Learning, Third International Conference, ICHL 2010* (pp. 383–394). 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). New York: Springer.

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

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): Editura Academiei.

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

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

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research, 10*(2), 1–24.

Wegerif, R. (2007). *Dialogic education and technology : Expanding the space of learning*. New York: Springer.

Wiemer-Hastings, P., & Zipitria, I. (2001). *Rules for syntax, vectors for semantics.* Paper presented at the Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society.

Wild, F. (2007). *An LSA package for R*. Paper presented at the Mini-Proceedings of the 1st European Workshop on Latent Semantic Analysis in Technology-Enhanced Learning.

Winne, P. H., & Hadwin, A. F. (1998). Studying as Self-Regulated Learning. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp. 277–304). Mahwah: Erlbaum.