Tatiana: an environment to support the CSCL analysis process

Gregory Dyke, Ecole Nationale Supérieure des Mines de Saint-Etienne, 42023, St. Etienne cedex 2, France, Dyke@emse.fr

Kristine Lund, ICAR, CNRS, University of Lyon, ENS-LSH, 15 Parvis René Descartes, BP 7000, 69342 Lyon cedex, France, Kristine.Lund@univ-lyon2.fr

Jean-Jacques Girardot, Ecole Nationale Supérieure des Mines de Saint-Etienne, 42023, St. Etienne cedex 2, France, girardot@emse.fr

Abstract. The analysis of multimodal computer-mediated human interaction data is difficult: the diverse nature of this data and its sheer quantity is challenging enough, but a further obstacle is introduced by the complex nature of these interactions. In this paper, we describe the kinds of activities performed by researchers wishing to analyze this data. We present a model for analysis based on these activities. We then introduce Tatiana (Trace Analysis Tool for Interaction Analysts) as an environment based on this model and designed to assist researchers in managing, synchronizing, visualizing and analyzing their data by iteratively creating artifacts which further their understanding or exhibit their current understanding of their data. We show how Tatiana can be used to perform analyses and its potential for sharing corpora and analyses within the research community.

Introduction

The socio-cognitive study of human computer-mediated interactions can be performed through recordings of these interactive activities, particularly if they are not limited to interaction log files but also include audio and video recordings (Avouris, Fiotakis, Kahrimanis, Margaritis, & Komis 2007). Cox (2007) encourages researchers to use computers and the various techniques they offer (visualization, data mining etc.) to perform their analyses of interactions in what he calls "process data". However, corpora of human interaction, particularly when these interactions are both face-to-face and mediated by computers are difficult to manage from a technological standpoint and complicated to understand and analyze due to the multiplicity and variation of source data. Indeed, it is not enough to look at individual data streams; different media streams must be combined to achieve a global understanding of the interactions that occurred (Goodman, Drury, Gaimari, Kurland, & Zarrella, 2006). Furthermore, it is often necessary to perform analysis as a team, be it in order to validate the analysis method through inter-coder reliability (De Wever, Schellens, Valcke, & Van Keer, 2006), to extend applicability of an analytical method to a new domain of application (Lund, Prudhomme, & Cassier, 2007), to spread the workload (Goodman et al., 2006), or to combine the insights of several analysts (Prudhomme, Pourroy & Lund, 2007).

The difficulties described above suggest the necessity of tools which provide not only the means to manage this variety and quantity of data, but also to allow visualization and analysis within a common framework and in a way that can be shared with other researchers (cf. Reffay, Chanier, Noras, & Betbeder, 2008 for work on structuring learning corpora for sharing purposes).

In this paper, we will give an overview of a selection of the actions that are performed by researchers when analyzing human interaction data, particularly in the case where it is computer mediated. We will then present a simple model to describe this analysis activity and show how the analysis tool Tatiana (Trace Analysis Tool for interaction analysis) is designed to support this model, which we see as generalizable. More specifically, we will demonstrate how Tatiana can assist analysts in performing the actions we have presented. Finally, we will conclude by exploring how Tatiana is intended to assist new kinds of analysis and foster sharing and collaboration in CSCL research.

How do we analyze?

A typical CSCL corpus might assemble video and audio recordings, computer interaction log files, pre- and post- tests, interviews, field notes and experiment descriptions (design, context, etc.). In designing a tool to support the analysis of such a corpus, the question we have attempted to answer is "what do researchers do with this data?"

While we could approach this question from a methodological standpoint, achieving a reasonable coverage of existing practices in the CSCL and other closely related communities would require performing a meta-analysis of a very large body of research and to do so in a satisfactory way is beyond the scope of this article. In addition, if we did carry out such an analysis, it would show us what researchers are doing but it would not necessarily show how they are doing it, as that information is rarely reported. As designers of tools to support this process, this information has been difficult to come by on a larger scale. In this section, we will present our understanding of analysis activity by exploring a subset of analysis themes and the artifacts that a

select number of researchers create in relation to these themes. To support our claims we will draw variously on case studies, papers describing methodological issues or experiments and tools that exist to support analysis. Finally we will present the model of analysis on which we have based the design of Tatiana.

Case Studies

In the context of the Lead(1) project (Lead, 2006), we have had the opportunity to follow the activities of four research teams with respect to the design and analysis of experiments related to the project. All these experiments focus on the use of computer-mediated communication in face-to-face situations (communication can be verbal or computer mediated). Analysis in these cases presents particular problems (Dyke *et al.*, 2007) one of which being the necessity of synchronized replay of data provided by computer log files and data provided by audio and video recordings in order to fully understand the situation. We were able to work in continued collaboration with one of the teams and had the opportunity for on-site visits with two other teams where we recorded and discussed their analysis activities. In presenting these case studies we cannot go into great detail as to the hypotheses and theoretical assumptions because they relate to as yet unpublished studies. We will however give an overview of the procedure that each team followed for some portion of their analysis (the part that is relevant to support our claims regarding how analyses are performed). The parallel development of Tatiana means that at least some of their analysis was performed with Tatiana, but particularly during on-site visits, we discovered requirements that we could not meet at the time. In this section we will be interested in the actual analysis steps that each team wanted to undertake rather than in the usage (or not) of Tatiana.

Lyon case study

We were our own main users of Tatiana during the initial development phase (variously wearing our tool development, usability tester and CSCL researcher caps). We observed nine dyads over the course of three to five meetings with their teacher for an introductory-level computer-programming project. These meetings took place face-to-face with the assistance of a chat and a real-time shared text editor (both of which the two students and the teacher had access to, on their laptops). The students were encouraged to take notes in the shared text editor.

In one analysis, we were interested in the provenance of these notes. Very often they are reformulations of utterances. We wanted to gain a deeper understanding of this reformulation from a linguistic viewpoint and explore possible pedagogical consequences. In order to do this, we first transcribed the dialogues. We transformed the interaction log data from the shared text editor into something we could understand. The nature of the shared text editor (Corbel, Girardot, & Jaillon, 2002; Corbel et al., 2003) means that the log data it produces consists of events which are recorded each second (provided changes have been made) and contain the full text, the name of the author who did the changes and the position of their cursor at the time of the change. We desired a much higher level of analysis, consisting of some kind of semantic unit. These events were therefore grouped together manually into writing units. We then annotated the reformulation links between the transcription and the writing units. Finally this was transformed into visualizations (cf. Figure 1) which allowed us to see certain interesting phenomena and to gain intuitive insight into the reformulation process.
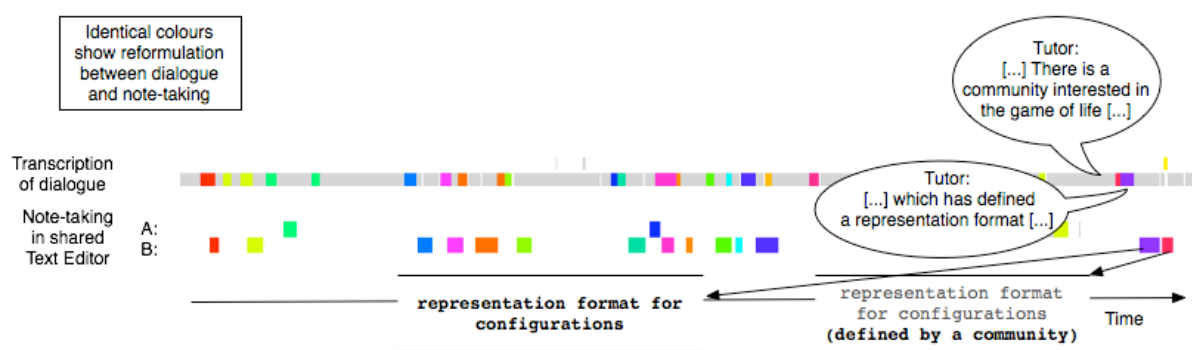

Figure 1. An example of a visualization of reformulation

Paris case study

In this second case, we were on site for two days, discussing their current practices and examining how Tatiana could usefully modify these practices. This resulted in audio and video recordings of our discussions along with usability testing of Tatiana which we will not report here. One of their analyses involved the long-term study of classroom adoption of the CoFFEE software (De Chiara, Di Matteo, Manno, & Scarano, 2007), developed during the LEAD project. They followed pairs of dyads who worked together in a shared workspace. Each dyad worked behind one computer and each pair of dyads (sitting at adjacent computers) worked in a common workspace . In this analysis they were specifically interested in the kinds of interactions (Bernard, 2008) that

happened within a dyad (tool-focused vs. subject-matter focused, verbal vs. computer mediated, etc.) hoping to observe the link between the kinds of interactions and the dyad-dyad collaboration.

In order to do this, they wished to replay the video synchronously with the interaction log data, marking blocks as different types of interaction took place in each dyad. These interactions would then be categorized according to a coding scheme describing the different kinds of interactions thus enabling the side-by-side comparison of the interactions of each pair of dyads in relation to the computer-mediated interaction.

Utrecht case study

In this third case, we were again on site for two days, discussing current practices, and examining whether Tatiana was adapted to them. This also resulted in audio and video recordings of our discussions. Their work focuses mainly on the production of argumentation diagrams (e.g. Overdijk & van Diggelen, 2008). One of their analyses methods consists in isolating from the log files only the events where boxes are created and then re-ordering these events to show the threading in the diagram, rather than the order of production (rather like threading occurs in threaded forums). From this point they are able to apply discourse or content analysis methods to the data.

Another analysis process they apply is the creation of visualizations that show how students' involvement across the various threads happens over time. These combined views (time-based, thread-based and student-based) enable a fuller grasp and better understanding of the situation (Lund *et al.*, 2008).

Nottingham case study

In the fourth case we have not had an opportunity for on site visits but have (as with the other partners) had discussions during meetings and via email. One practice of interest to us is their use of statistical methods such as those available in SPSS and Excel. In one study, a follow-up on Gelmini Hornsby, Ainsworth, Buda, Crook, & O'Malley (2008), they were interested in how various conditions influenced change of opinion in a vote module in CoFFEE. The global pattern was first vote, discussion and second vote. Before each vote was closed, students could change their opinion any number of times and this was reflected in the log files. In this case, the researchers had to manually find the last opinion in each vote for each student, in order to observe whether an opinion changed between the first and second vote.

They also expressed interest in automated calculation of general kinds of statistical indicators in the discussion such as words per turn, words per turn per student, turns per student per topic etc. The rationale behind this being that any kinds of abnormal phenomena (e.g. unusually high or low participation) might be a starting point for further exploration. Finally, they confirmed the necessity of any new tool to be able to integrate with their current practices (e.g. Excel and SPSS, but also other analysis tools and transcription tools).

Some analysis themes

As we have previously stated, our aim is not to cover all (or even the majority) of analysis methodologies, their theoretical assumptions and the intricacies of their conditions of application. Rather, we use these analysis approaches to evidence a selection of the kinds of actions that analysts are led to perform.

Coding and counting

Several kinds of analysis make use of coding and counting, most notably content analysis (Strijbos, Martens, Prins, & Jochems, 2006). This is such a common practice that there is an increasing body of literature on various artificial intelligence techniques to help automate the task of coding (Rosé *et al.*, 2008; Erkens, & Janssen, 2008). In the application of this kind of method, an important question is that of *unit of analysis* (De Wever *et al.*, 2006); what is the pertinent granularity: e.g. a dialogue turn, propositional content within a dialogue turn, a succession of dialogues turns? Another question is that of the coding scheme to be applied, which can be drawn from the literature (e.g. Baker, Andriessen, Lund, van Amelsvoort, & Quignard, 2007) or created for the purpose of a specific analysis. Once this scheme is applied the question of *validation* of the coding (De Wever *et al.*, 2006) must be raised. Finally some statistical tests must be applied.

Regardless of the choice of analysis unit, the corpus must be broken up into such units. This segmentation is performed differently, depending on the type of media. Dialogue and human gestures from audio and video is typically transcribed using tools such as Elan (http://www.lat-mpi.eu/tools/elan/). Computer based interaction log data is transformed into a series of actions which are then regrouped or re-segmented. In both cases, the resulting data is in the abstract form of events (or actions) with a series of properties such as time, user/speaker, content, tool, etc. More concretely, such data can typically be found in an Excel^{TM} spreadsheet with one row per event and one column for each property. In the case of existing software support, tools such as Elan, Videograph® (http://www.ipn.uni-kiel.de/aktuell/videograph/enhtmStart.htm) and DRS (Digital Replay System, Greenhalgh, French, Humble, & Tennent, 2007) enable the direct annotation and coding of video in the way suggested by the desired practice of the Paris case study.

Coding or rating schemes are diverse in nature. They are often created through an iterative process involving partial coding of a corpus and subsequent redefining of the scheme, or even through collaborative coding (De Vries, Lund, & Baker, 2002). It is rarely the case that a coding scheme is directly applied without further modification. This point is important for software design as it means that evolving categorization schemes must be taken into account.

Coding can then be seen as adding one or more properties to each event by a coder. The subjective nature of this activity combined with the desirability of subsequently applying statistical methods to coded data necessitates the validation of the coding. It is common for intercoder reliability to be applied (De Wever *et al.*, 2006), which, regardless of the metric used (Cohen's kappa, Krippendorf's alpha, etc.) usually involves comparing two or more codings (complete or partial) side by side.

Data is then collated by counting various codes according to other characteristics (e.g. user, group, tool). By observing the distribution these codes and through the application of statistics certain kinds of hypotheses can be confirmed or infirmed.

Bookmarks, collections and annotations

Other kinds of analysis, most notably through case studies (e.g. Rummel, & Hmelo-Silver, 2008), in the field of conversational analyses (Sacks, Schegloff, & Jefferson, 1974) and analyses based on activity theory (e.g. Avouris et al, 2007), consist of describing corpora by adding annotations, creating collections or groups of events. This can be done manually (in a text file or spreadsheet) or through the direct support of tools such as ActivityLens (Fiotakis, Fidas, & Avouris, 2007), Videograph or Transana (http://www.transana.org/).

Synchronization in time

The analytical necessity of synchronization of different data sources and analysis artifacts is plain: the observed events — at the time of observation — were temporally situated and must be replaced in this context to be understood. Furthermore, different views on the same data (e.g. video and transcription) complement each other well when they are synchronized. In fact, it is a scientific necessity to have this synchronization (or some other means of returning to the primary data); when making a claim the question that is always asked is: "is there evidence in the data to back up this claim?" Analysts frequently make use of artifacts that present the corpus in a way that is more readily understandable, browsable or analyzable (compare transcriptions and the video/audio they transcribe). However, when a claim is postulated based on information found in one of these "secondary" artifacts, it is necessary to verify that the original data also evidences the claim that is being made. Synchronization presents a way to easily refer back to the primary data at any point where confirmation of a claim is needed.

Synchronization of different media sources is further justified by the number of tools that enable it to a greater or lesser degree: Elan, ActivityLens, DRS, Replayer (Morrison, Tennent, & Chalmers, 2006) and ABSTRACT (Analysis of Behaviour and Situation for menTal Representation Assessment and Cognitive acTivity modeling; Georgeon et al., 2007; Georgeon, Mille, & Bellet, 2006) all enable synchronized replay (c.f. Figure 2) of the analysis artifacts which they are designed to create and handle.

Figure 2. In this example from Replayer, the events selected in the left hand view are also highlighted on the map and in the video timeline on the right (http://www.dcs.gla.ac.uk/%7Emorrisaj/Replayer.html).

Graphical Visualizations

As our examples in the Lyon and Utrecht case studies show, graphical visualizations allow researchers to look at data through a different lens and often help in isolating interesting phenomena or give an intuitive insight into what happened. A typical form of graphical visualization is a symbolic representation of events on a horizontal timeline. This feature is pervasive, being present in SAW (Synchronized Analysis Workspace, Goodman et al., 2006), ABSTRACT (cf. Figure 3), Replayer, DRS and others. In this case, visualizations can be used to explore the temporal dimension of the data. Suthers & Medina (2008) augment these visualizations with complex annotations, allowing the visualization to become a means of recording knowledge gained from analysis and for communicating this knowledge to a wider audience. Teplovs & Scardamalia (2007) use the information found

in log files of a threaded forum to generate interactive visualizations which group similar posts on criteria such as common author, common threading and semantic proximity. There are many other kinds of visualizations which aim to convey some kind of analytical knowledge about data such as the generic argumentation diagrams describing design produced by Prudhomme, Pourroy & Lund (2007) or those which are related to awareness tools such as those presented by van Diggelen, Jansen, & Overdijk (2008).
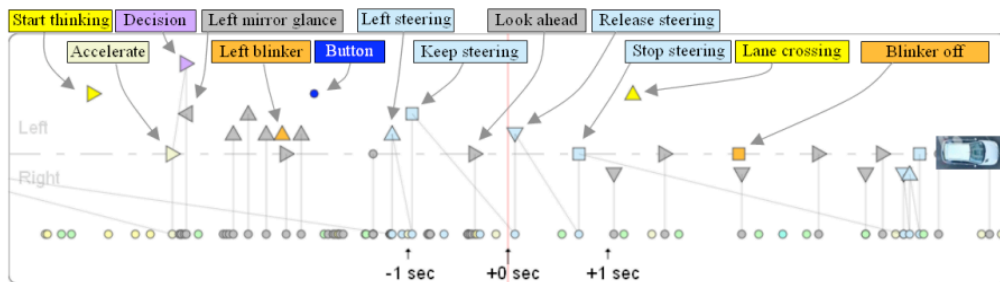

Figure 3. Symbolic graphical representation of events recorded during a lane change on a motorway, viewed in ABSTRACT (http://liris.cnrs.fr/abstract/).

Interoperability and sharing

Reffay *et al.* (2008) note the importance of being able to share CSCL corpora. Kahrimanis, Papasalouros, Avouris, & Retalis (2006) examine how greater interoperability between CSCL log data and analysis tools can be achieved. Our case studies tell us that researchers have existing practices which are tried and trusted, and that they do not feel safe putting all their eggs into the basket of one tool. We are ourselves interested in the possibilities generated through the sharing of analyses as a means of validation (such as the kind provided by inter-coder reliability) or as a means to gain a holistic understanding through the combination of analyses from experts in different domains (e.g. Prudhomme *et al.*, 2007).

A simple model of analysis

Harrer et al. (2007) have modeled the analysis process (cf. Figure 4) with a view of basing the design of analysis tools and interoperable formats on this model. They propose that analysis can be seen as a sequential process going through phases of capture, segmentation, annotation, analysis, visualization and interpretation.
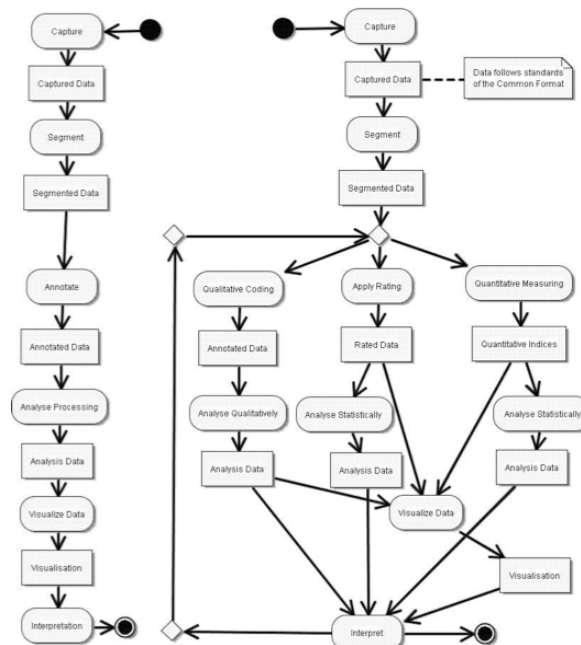

Figure 4. Graphical representation of the Cavicola process model (Harrer, et al., 2007, p. 2)

Our current understanding of analysis, based on our case studies and our past experience (Lund, Rossetti & Metz, 2007; de Vries *et al.*, 2002; Prudhomme *et al.* 2007; Baker, et al. 2007), is that the most important part in this model is the iterative loop: researchers who arrive at the interpretation phase and are not satisfied with their results incrementally improve the analysis of their data until they arrive at a satisfactory

result which can be reported to the scientific community. Contrary to what can be seen in Figure 4, we do not believe it is necessary to go through all the phases in a particular order or to wait until the interpretation phase to iterate. For example, segmentation of data into units of analysis has often been seen in our case studies as being unavoidably entangled with the kind of coding that will later be applied; an inability to apply a coding to a certain unit can lead to a slight change to the segmentation.

We have based our design of Tatiana on a similar (but less detailed) model that puts more focus on the iterative nature of analysis (cf. Figure 5). Analysts constantly evaluate whether their current collection of primary data and secondary artifacts is sufficient. If it is not, they create a new artifact that is intended either to further their understanding of the data or to reify their current understanding of the data.
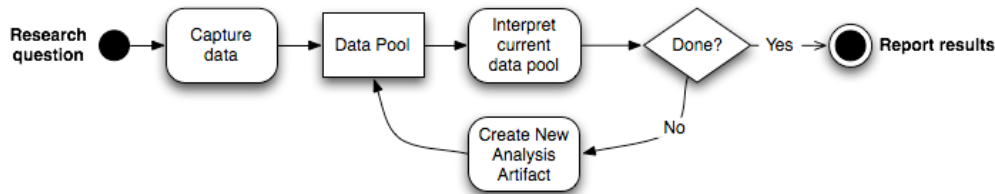

Figure 5. Graphical representation of the analysis model which Tatiana is designed to support

In the analysis themes presented above, we have seen the variety of artifacts that researchers create. Sometimes the creation of these artifacts can be automated (e.g. transforming data into a new representation or performing statistical analysis). Sometimes it is manual (e.g. creating certain kinds of visualizations) and sometimes it is tool-assisted (e.g. transcription, annotation and coding). These artifacts are frequently ways of representing data with some kind of temporal dimension through different lenses: the sequence of events recorded in the observation is made available to the researcher, either by presenting it in a media player (the researcher uses a remote control to navigate the data) or in some kind of graphical or tabular representation (where time is plotted along a vertical or horizontal axis).

These artifacts can be classified into three kinds. *Collations* aggregate data over a time period producing data such as indicators and statistics. *Analyses* are artifacts which add researcher created analytical knowledge (such as codings, annotations and relationships). Finally, we propose that the kinds of artifacts which retain some notion of ordering of events and interactions in time be termed *replayables*. These are objects that can be replayed, synchronized and analyzed. We further propose that the analysis process consists of the iterative creation of new artifacts (such as replayables and analyses) that exhibit researchers' understanding of their data or that allow them to further this understanding. We are particularly interested in replayables as they are the most frequent source for creating new artifacts, other replayables in particular. The transformations from replayables into other artifacts include transcription, annotation, coding, visualization, filtering, synchronization, merging and collation. It should be noted that the creation of analyses might be assimilated to creating a new replayable where each event has a new set of properties. However, by considering analyses as separate entities, they become reifiable objects which can be shared with other researchers who are already in possession of the same corpus and can be overlaid on top of other replayables representing the same data.

Tatiana: a generic analysis environment

In the previous section, we describe how researchers perform their analyses. While many tools currently exist to assist some of these tasks, even the most generic of them such as ActivityLens and Digital Replay System lack several features such as automated transformations, the ability to include new kinds of data and the extensibility to adapt to new kinds of analyses. Tatiana (Trace Analysis Tool for Interaction ANAlysts) is an environment designed for manipulating the kinds of artifacts described above, replayables in particular. In this section we briefly present the features of Tatiana which make this possible.

Tatiana overview

Tatiana is built on a number of core concepts and components (cf. Figure 6). Tatiana replayables can be created either automatically (through import) or by hand. Once created, all replayables in Tatiana benefit from Tatiana's four core functionalities: transformation, analysis, visualization and synchronization.

Transformations

Replayables can be transformed (again, automatically or manually) and exported. Automated import, transformation and export works through the application of what we call *filters*. These are objects which combine *scripts* into a workflow. Scripts are small programs written to perform a specific operation, such as transform a file in the corpus into data Tatiana can understand, exclude certain kinds of events from a replayable, find certain kinds of events in a replayable, combine multiple replayables, etc. As we do not expect

researchers with no programming knowledge to write these scripts, we are currently developing a graphical editor for filters which will allow researchers to customize the execution of these scripts. Such a filter might combine a new script for data import from the interaction log data produced by a new kind of tool with an existing script which only shows the actions of a particular subset of students. Manual transformations include the ability to delete, reorder, re-group and split events.
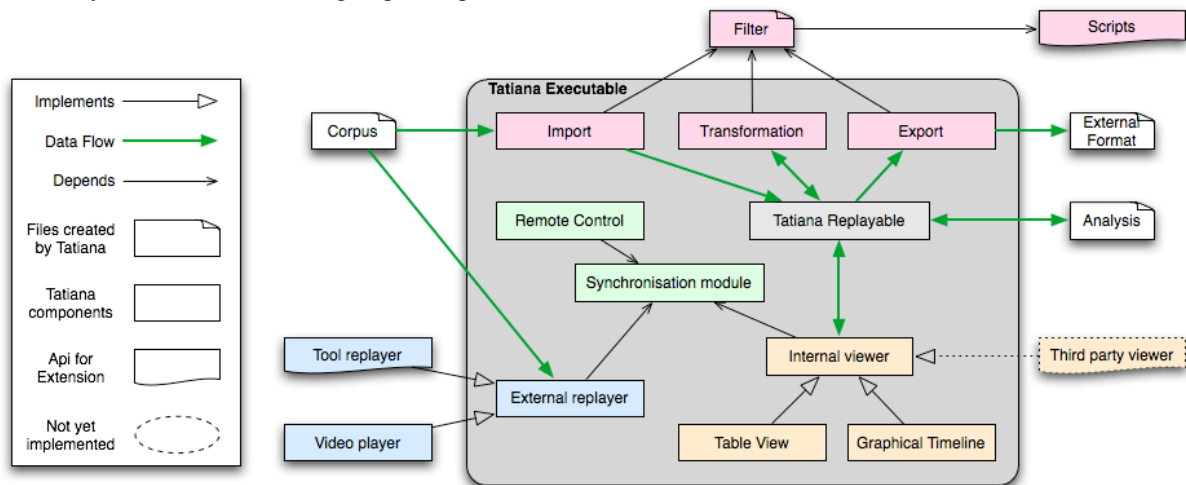

Figure 6. Tatiana architecture showing 1) dependencies between components, 2) components designed with extensibility in mind and 3) future developments.

Analysis

All replayables within Tatiana can be augmented by analysis data generated by the researcher. There are currently two kinds of analyses supported by Tatiana: annotations (for free-form annotation) and categorizations. Categorization is simply a way of adding annotations from a restricted list of words and can be used for coding, labeling and adding keywords. The list of categories available can be edited at any time thus allowing for an evolving analysis scheme.

Visualization

All replayables within Tatiana can be visualized in different viewers. There currently exist two kinds of viewers: a table view in which data is presented as it might be presented in Excel, with one row per event and columns for each of the event's properties and a graphical timeline. The graphical timeline is a first attempt at assisting the automated creation of visualizations. It presents each event as a graphical object whose graphical properties (color, shape, size, position, etc.) can be set according to the properties of the event (user, tool, timestamp, analysis category, etc.). We plan on making Tatiana extensible so that new kinds of visualizations of replayables can be created while benefiting from Tatiana's other core functionalities.

Synchronization

Finally, all replayables in Tatiana can be synchronized with each other and also with data viewed in external replayers such as media players and tool replayers (a special mode of certain CSCL tools which are capable of reading the interaction log data they generated and reproducing on screen what the user saw). Tatiana provides a mechanism to pilot external tool replayers. Synchronized replay means that when a timestamp is selected in the "remote control", the video player (and other external replayers) are instantly navigated to that timestamp, and the events matching that timestamp in the currently visualized replayables are highlighted. Furthermore, selecting an event in a visualized replayable will again navigate all the other views to that moment in time. For example, during analysis of a discussion of genetically modified organisms using Tatiana, if a researcher clicks on the timestamped event in the table view "argument by Alice: they go against ethics" this action causes the replayer to show the state of the diagram immediately after that argument was constructed. Information on the dynamics of the interaction in thus provided, which is oftentimes difficult to discern in static log traces. Zooming in on particular episodes becomes possible. In general, such linking between replayables is very useful for limiting the amount of information displayed in a single visualization, with the knowledge that further information is available in other visualizations on demand.

Sample use of Tatiana

Typical usage of Tatiana consists of iteratively creating replayables and analyses, gradually increasing and exhibiting the researcher's understanding of his or her corpus. To illustrate, we show how we used Tatiana in the Lyon case study to analyze reformulation of dialogue into notes in a shared text editor (cf. Figure 7).
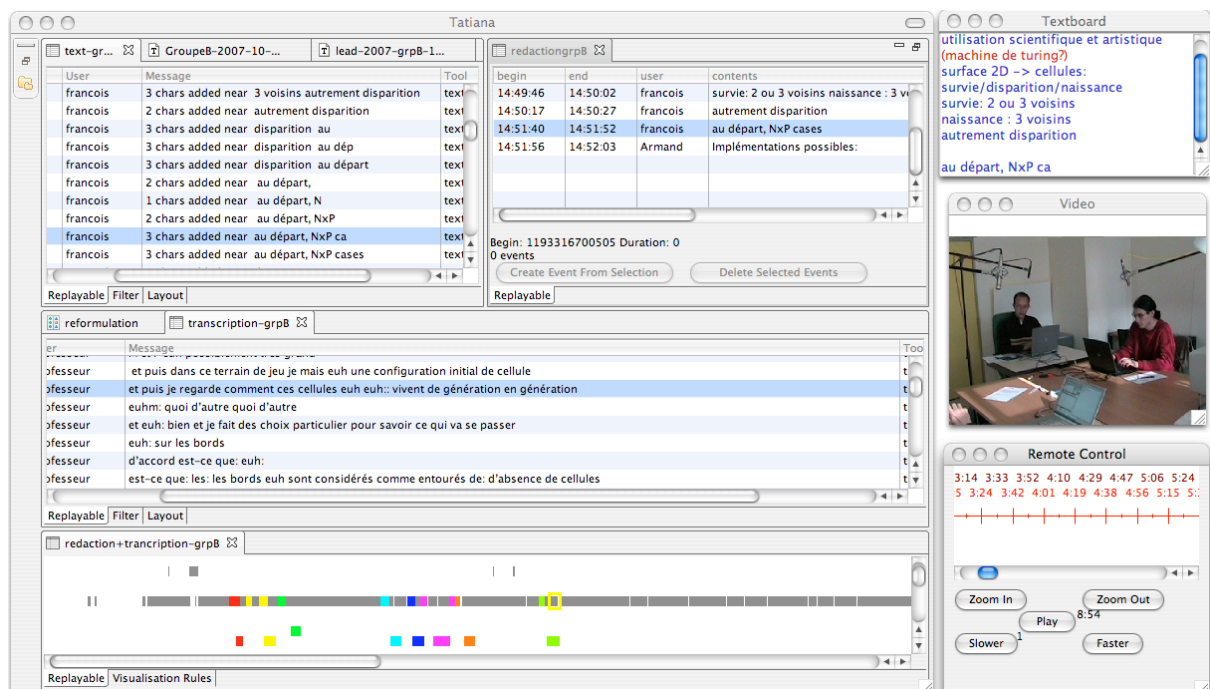
Figure 7. Various replayables in Tatiana: traces of a shared text editor (top left), transcription (middle left), writing units (top center), visualization of reformulation (bottom left), synchronized with external tools, DREW replayer (top right), video player (middle right), remote control (bottom right).

The artifacts created can all be seen under Tatiana synchronization. Transcription was performed outside of Tatiana and then imported as a replayable representing transcription. The interaction log data at low granularity was grouped together into a replayable representing writing units. The writing units and transcription were analyzed, giving identical labels and colors where reformulation occurs. The writing unit and transcription replayables were merged, and visualized as a graphical timeline with transcription on the top row and writing units on the bottom row. Identical colors show which utterances in the transcription have reformulations in the shared text editor. Transcription, grouping and analysis was performed by hand (with tool assistance), but all the other transformations were automated.

In order to give meaning and context to the data produced in the shared text editor, we can observe in the replayer (here DREW is used: Corbel, Girardot & Jaillon, 2002; Corbel et. al. 2003) and in the video exactly what happened at a given time.

Tatiana usage and limitations

Tatiana was developed in parallel to the analyses carried out in our case studies. Its features enabled us to carry out our work in the Lyon case study and are now adequate to assist in performing the analyses in the other cases. Further analyses are expected to be carried out with the assistance of Tatiana in coordination with other tools (such as Excel and SPSS for statistical analysis and Elan for transcription).

There is currently limited support for collations in Tatiana. Such data is difficult to synchronize, analyze and visualize in the same way as for replayables. However, transformations do exist to produce limited statistics such as contingency tables which can then be exported to other formats such as Excel. Once generated, replayables and analyses can be saved and shared with other researchers.

Conclusions and Future work

In this paper, we described many of the activities performed by analysts when analyzing CSCL and related corpora. Based on these activities, we presented a model describing the analysis process. In this model, researchers iteratively create new artifacts which afford them new understanding or exhibit their current understanding of their corpus. We identified a particular time based artifact which we call a replayable. We then presented Tatiana, a tool whose design supports the iterative creation of replayables. We described the main features of Tatiana with regard to creation, transformation, export, analysis, visualization and synchronization of these replayables.

Tatiana responds to the inherently iterative and diverse nature of socio-cognitive interaction analysis by providing flexible data transformations and visualizations and by providing several extension points in order to meet new needs such as creating new transformation filters and providing additional views for the creation and

iterative improvement of replayables. Such new views could be similar to the visualizations suggested by Suthers & Medina (2008), or Teplovs & Scardamalia (2007). The difficulty in recreating what participants experienced during corpus collection on the basis of the recorded data is answered through multiple synchronized visualizations of the data, and through the integrated use of external tool replayers. Finally, through the ability to save and share analyses, Tatiana enables researchers in the human and social sciences to work as a team and to integrate and compare their analyses.

Our future work will involve simultaneously bettering our understanding of CSCL and CSCW researcher's analysis methodologies and further developing Tatiana as an environment for managing replayables (and other analysis artifacts). We will strive to make this framework more coherent, and to assess the extent to which researchers with no programming experience can fully use the power provided by Tatiana.

We hope that in making Tatiana and similar tools more widespread, we will enable researchers to make better sense of their corpora and to share the knowledge they have created with other researchers. We also hope that this will enable us to gain more insight into the process of socio-cognitive interaction analysis, making it easier to evaluate and share methodologies, corpora and analyses throughout the CSCL community.

Endnotes

(1) The European project LEAD (*Technology-enhanced learning and problem-solving discussions: Networked learning environments in the classroom*) is funded by the 6th framework *Information Society Technology* LEAD IST-028027. http://www.lead2learning.org/

References

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

Baker, M., Andriessen, J., Lund, K., van Amelsvoort, M. & Quignard, M. (2007). Rainbow: A framework for analyzing computer-mediated pedagogical debates. *ijcscl* 2 (2-3).

Bernard F.-X. (2008). Un modèle d'analyse des interactions médiatisées par les technologies éducatives : le carré médiatique [An analysis model of interactions mediated by educative technologies]. *Congrès national de la Société Française de Psychologie (SFP)*, France.

Carletta, J., Evert, S., Heid, U., Kilgour, J., Robertson, J., and Voormann, H. (2003). The NITE XML Toolkit: flexible annotation for multi-modal language data. Behavior Research Methods, Instruments, and Computers, special issue on Measuring Behavior, 35(3), 353-363.

Cox, Richard (2007). Technology-enhanced research: educational ICT systems as research instruments. *Technology, Pedagogy and Education*, 16 (3), 337-356.

Corbel, A., Girardot, J.J., & Jaillon, P. (2002). DREW: A Dialogical Reasoning Web Tool*, ICTE2002, International Conference on ICT's in Education. Badajoz*, Spain.

Corbel, A., Jaillon, P., Serpaggi, X., Baker, M., Quignard, M., Lund, K., & Séjourné, A., (2003). DREW : Un outil Internet pour créer des situations d'apprentissage coopérant [DREW : An internet tool for creating cooperative learning situations]. In Desmoulins, Marquet & Bouhineau (Eds.) EIAH2003 *Environnements Informatiques pour l'Apprentissage Humain, Actes de la conférence EIAH* 2003, Strasbourg, 15-17 avril 2003, Paris : INRP, p. 109-113.

De Chiara, R., Di Matteo, A., Manno, I., Scarano, V. (2007). CoFFEE: Cooperative Face2Face Educational Environment. *Proceedings of the 3rd International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2007)*, New York, USA.

De Vries, E., Lund, K. & Baker, M.J. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *The Journal of the Learning Sciences*, 11(1), 63—103.

De Wever, B., Schellens, T., Valcke, M. & Van Keer, H. (2006). Content analysis schemes to analyse transcripts of online asynchronous discussion groups: A review. *Computers & Education*, 46(1) 6-28.

Dyke, G., Girardot, J.-J., Lund, K., & Corbel, A. (2007). Analysing Face to Face Computer-mediated Interactions. *EARLI '07*. Budapest, Hungary.

Erkens, G. & Janssen, J. (2008) Automatic coding of communication in collaboration protocols. *ijcscl* 3 (4).

Gelmini Hornsby, G., Ainsworth, S., Buda, M., Crook, C., & O'Malley, C. (2008). Making your views known: The importance of anonymity before and after classroom debates. *ICLS 2008*, Utrecht, The Netherlands.

Georgeon, O, Mille, A, & Bellet, B.T. (2006). ABSTRACT: un outil et une méthodologie pour analyser une activité humaine médiée par un artefact technique complexe [ABSTRACT: a tool and a methodology for analyzing human activity mediated by a complex technical artifact]. *Ingéniérie des Connaissances. Semaine de la connaissance*. Rémi Lehn, Mounira Harzallah, Nathalie Aussenac-Gilles, Jean Charlet ed. Nantes, France.

Georgeon, O., Mathern, B., Mille A., Bellet, T., Bonnard, A., Henning & N., Trémaux, J.-M. (2007). *ABSTRACT Analysis of Behavior and Situation for menTal Representation Assessment and Cognitive acTivity modelling*, Retrieved April 10, 2008 from http://liris.cnrs.fr/abstract/abstract.html.

Fiotakis, G., Fidas, C. & Avouris, N (2007). Comparative usability evaluation of web systems through ActivityLens, *Proc. PCI 2007*, Patras, Greece

Goodman, B. A., Drury, J. Gaimari, R. D., Kurland, L., & Zarrella, J (2006) Applying User Models to Improve Team Decision Making. Retrieved April 10, 2008 from http://mitre.org/work/tech_papers/tech_papers_07/06_1351/

Greenhalgh, C., French, A., Humble, J. and Tennent, P. (2007) Engineering a replay application based on RDF and OWL, *Online Proceedings of the UK e-Science All Hands Meeting 2007*, September 10-13, Nottingham: NeSC/JISC.

Harrer, A. Zeini, S. Kahrimanis, G. Avouris, N. Marcos, J. A. Martinez-Mones, A. Meier, A. Rummel, & N. Spada, H (2007) Towards a Flexible Model for Computer-based Analysis and Visualisation of Collaborative Learning Activities, *Proc. CSCL 2007*, New Jersey, USA

Kahrimanis G., Papasalouros A., Avouris N., & Retalis S. (2006), A Model for Interoperability in Computer-Supported Collaborative Learning. *Proc. ICALT 2006 - The 6th IEEE International Conference on Advanced Learning Technologies.* IEEE Publ., July 5-7, 2006 – Kerkrade , Netherlands, p. 51-55. *interaction in the classroom*,

Lead (2006). *Problem solving through face to face networked interaction in the classroom*, Retreived April 10, 2008 from http://www.lead2learning.org

Lund, K., van Diggelen, W., Dyke, G., Overdijk, M., Girardot, J.J., & Corbel, A (2008). A researcher perspective on the analysis and presentation of interaction log files from CSCL situations within the LEAD project. *ICLS 2008*, Utrecht, The Netherlands.

Lund, K., Prudhomme, G., & Cassier, J.-L. (2007). Using analysis of computer-mediated synchronous interactions to understand co-designers' activities and reasoning. *Proceedings of the International Conference On Engineering Design*, ICED'07, 28 - 31 august 2007, Cité des Sciences et de l'Industrie, paris, France.

Lund, K., Rossetti, C. & Metz, S. (2007). Do internal factors of cooperation influence computer-mediated distance activity? *Proceedings of the international conference CSCL 2007*, July 16-July 21, New Brunswick, New Jersey, USA.

Morrison A., Tennent P., & Chalmers M. (2006). Coordinated Visualisation of Video and System Log Data. *Proceedings of 4th International Conference on Coordinated & Multiple Views in Exploratory Visualisation*. London, 91-102.

Overdijk, M. & van Diggelen, W. (2008) Appropriation of a shared workspace: Organizing principles and their application. *ijcscl* 3 (2).

Prudhomme, G., Pourroy, F., & Lund, K. (2007). An empirical study of engineering knowledge dynamics in a design situation. *Journal of Design Research, 3,* 333-358.

Reffay, C., Chanier, T., Noras, M., & Betbeder, M.-L. (2008) Contribution à la structuration de corpus d'apprentissage pour un meilleur partage en recherche [Contribution to the structuring of learning corpora for a better sharing in research], *STICEF journal*, Volume 15, ISSN : 1764-7223. Put online 14th October 2008, http://sticef.org.

Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A. & Fischer, F. (2008) Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in CSCL. *ijcscl* 3 (3).

Rummel, N., & Hmelo-Silver, C. (2008). Using contrasting cases to relate collaborative processes and outcomes in CSCL. *ICLS 2008*, Utrecht, The Netherlands.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organiszation of turn-taking for conversation, *Language*, 50, 696-735.

Strijbos, J., Martens, R. L., Prins, F. J., and Jochems, W. M. (2006). Content analysis: what are they talking about?. *Comput. Educ.* 46, 1 (Jan. 2006), 29-48.

Suthers, D. D. & Medina, R. (2008). Tracing Interaction in Distributed Collaborative Learning . Paper presented at the *Annual Meeting of the American Educational Research Association* (AERA), New York, March 24-28, 2008.

Teplovs, C., & Scardamalia, M. (2007). Visualizations for Knowledge Building Assessment. Paper presented at the AgileViz workshop, CSCL 2007.

Van Diggelen, W., Jansen, J., & Overdijk, M. (2008). Analyzing and Presenting Interaction Data: A Teacher, Student and Researcher Perspective, *ICLS 2008*, Utrecht, The Netherlands.