

# Sharing Corpora, Analysis and Tools for CSCL Interaction Analysis

Christophe Reffay, Marie-Laure Betbeder

► **To cite this version:**

Christophe Reffay, Marie-Laure Betbeder. Sharing Corpora, Analysis and Tools for CSCL Interaction Analysis. Pre-conference Workshop "Common Objects for Productive Multivocality in Analysis". Computer Supported Collaborative Learning, Jun 2009, Rhodos, Greece. <edutice-00410225>

**HAL Id: edutice-00410225**

**<https://edutice.archives-ouvertes.fr/edutice-00410225>**

Submitted on 18 Aug 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sharing Corpora, Analysis and Tools for CSCL Interaction Analysis

**Christophe Reffay, Marie-Laure Betbeder**, *Computer Science Laboratory of the Franche-Comté University, France.*

## Interaction analysis

In previous research, presented in CSCL2003, we built an automatable computing process based on Social Network Analysis to evaluate the cohesion of a learning group using e-mail and forum messages (Reffay & Chanier, 2003). Another work, mainly conducted during the phd thesis of A. Mbala (Mbala, 2003) and presented in ITS2002 (Intelligent Tutoring System) (Mbala et al., 2002), is a multi-agent system that tries to predict if some learners are less and less participating, in order to prevent their resignation. We have also worked on the analysis of data resulting from a tailorable framework to support collective activities in a learning context (Betbeder & Tchounikine, 2003). We analysed the content of the acts and worked out the proportion of acts related to the following four categories: activity achievement, group organization, environment tailoring and socialization.

In collaboration with a research team in linguistics, we are also involved in the field of multimodal interaction analysis (Betbeder et al., 2008), combining quantitative (macroscopic level) and content analysis (microscopic level) on micro actions that each participant can realize in an audio-graphic synchronous collaborative environment (audio talking turns, chat acts, votes, paragraph production in a shared text document, objects in a shared concept map or whiteboard). Our contributions in this field deal with (1) pattern discovering in sequences of such actions (Betbeder et al., 2007), (2) a visualization tool (Betbeder et al., 2008) which emphasizes the intertwinement of the actors' synchronous acts. In such an environment, actors can interact by using different modalities at the same time. Due to the importance of time and sequence of acts, such phenomena are hardly visible through a database representation.

We are currently working on navigation through corpora. This includes the selection of corpora and visualization of archived interaction acts and the context of the learning situation. Once the corpus has been chosen, a system of requests provides selection of parts of the corpus by considering for example: time restriction, selected communication tools, author or group or string search in interaction contents. The resulting set of acts can be visualized in the appropriate form, according to the communication tool they are recorded from. This heterogeneous set of acts is organised by our XSD schema, specified by the Mulce project, presented in (Reffay & al., 2008) and available on the Mulce project site<sup>1</sup>.

To facilitate analysis on the platform, we also want to propose tools originated from our research team or from partnership. For example we have two running collaborations: the Calico<sup>2</sup> project and Tatiana (Dyke et al., 2007). The Calico project aims at proposing different visualization and analysis tools available as servlets on their platform, specialized in discussion forums. Tatiana is a client application that combines and synchronizes various data sources (i.e. videos, logs, etc.). It includes a replayer and an annotator as well as various visualization facilities typically useful for synchronous interactions (face to face, chat, synchronously shared document editors).

---

<sup>1</sup> Mulce data structure available here: [http://mulce.univ-fcomte.fr/metadata/mce-schemas/mce\\_sid.xsd](http://mulce.univ-fcomte.fr/metadata/mce-schemas/mce_sid.xsd)

<sup>2</sup> French national research project coordinated by E. Bruillard (ERTÉ: Technical Research Team in Education) <http://calico.inrp.fr/>

## Mulce project

The Mulce<sup>3</sup> Project aims at developing a server that would allow researchers to share Learning and Teaching Corpora (Letec). Our main goals are: first, to facilitate the work of researchers in the CSCL field by reusing existing corpora instead of creating new experiments, collecting and organizing data, and secondly, to connect these corpora to shared analyses and visualization tools. We also hope that this process of sharing would deepen and widen the validity of research tools and analyses.

Its main objectives are (1) to define the data structure of a “Learning and Teaching Corpus”, (2) to specify and develop a technical support to effectively share such corpora, integrating OLAC specifications (Open Language Archives Community) and OAI-PMH (Open Archives Initiative’s Protocol for Metadata Harvesting) and (3) to rebuild 2 of our global corpora (Simuligne and Copeas) to be accessible throughout this platform, and documented according to the specifications of a “Learning and Teaching Corpus”.

We propose a formalism to describe learning and teaching corpora and a platform to share them among the research community. The formalism defines the information which can be contained in a corpus and the structure of the data. Through the platform, researchers can share their corpora with the community and access the data shared by other members of the community. To share a corpus, a researcher has to provide metadata describing the corpus’ components and upload a file describing each component. While accessing a corpus, an identified researcher is provided with a variety of tools to browse the corpus components, to navigate through the contextualized interaction data, to visualize and to analyze them. Our major goal is to develop efficient tools and technical environments to help the wide variety of actors involved in online teaching and learning.

## REFERENCES

- Betbeder, M.-L., Ciekanski, M., Greffier, F., Reffay, C., and Chanier, T. (2008). Interactions multimodales synchrones issues de formations en ligne : problématiques, méthodologie et analyses, *In STICEF journal*, 15, (to appear).
- Betbeder, M.-L., & Tchounikine, P. (2003). Symba: a tailorable framework to support collective activities in an learning context. In J. Favela and D. Decouchant, (Eds.), *Procs of the 9th Int. Workshop on Groupware (CRIWG 2003)*, Springer-Verlag, Autrans, France, 90-98.
- Betbeder, M.-L., Tissot, R., Reffay, C. (2007). Recherche de patterns dans un corpus d’actions multimodales. In Nodenot, T., Wallet, J., Fernandes E. (Eds.), *EIAH’2007 Conference: Environnements Informatiques pour l’Apprentissage Humain*, Switzerland, 533-544.
- Dyke, G., Girardot, J.-J., Lund, K. & Corbel, A. (2007). Analysing face to face computer-mediated interactions. EARLI (European Association for Research, Learning and Instruction), 12th Biennial International Conference, University of Szeged, Eötvös Lorand University, Hungarian Academy of Sciences,.
- Mbala, A., Reffay, C., and Chanier, T. (2002). Integration of automatic tools for displaying interaction data in computer environments for distance learning. In S.A. Cerri, G. Guardères, and F. Paraguaçu (Eds.), *ITS’02, Intelligent Tutoring System conference, LNCS*, Springer-Verlag, France, 841-850.
- Mbala, A. (2003). Analyse, Conception, spécification et développement d’un système multi-agents pour le soutien des activités en formation à distance. Phd thesis of University of Franche-Comté.
- Reffay, C. & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning? *In proceeding of CSCL’2003*, Bergen.
- Reffay, C., Chanier, T., Noras, M. and Betbeder, M.-L. (2008). Contribution à la structuration de corpus d’apprentissage pour un meilleur partage en recherche. *In STICEF journal*, 15, (to appear).

---

<sup>3</sup> *Multimodal Learning Corpus Exchange* (2007-2009). <http://mulce.univ-fcomte.fr/axescient.htm#eng>