
Tatiana : un environnement d'aide à l'analyse de traces d'interactions humaines

Gregory Dyke* — Kristine Lund** — Jean-Jacques Girardot*

* G2I, Ecole Nationale Supérieure des Mines de Saint-Etienne,
42023 Saint-Etienne cedex 2
{nom}@emse.fr

** ICAR, CNRS, Université de Lyon, ENS-LSH,
15 Parvis René Descartes, BP 7000, 69342 Lyon cedex
Kristine.Lund@univ-lyon2.fr

RÉSUMÉ. L'analyse des interactions humaines multimodales médiatisées par ordinateur est difficile. La variété et la diversité des données recueillies par l'instrumentation de telles activités sont en elles-mêmes impressionnantes, mais un obstacle supplémentaire est introduit par la nature complexe de ces interactions. Dans cet article, nous décrivons les types d'activités effectuées par les chercheurs qui analysent ces données. Nous proposons un modèle de l'analyse reposant sur cette étude, puis nous présentons Tatiana (Trace Analysis Tool for Interaction Analysts), un environnement logiciel basé sur ce modèle et conçu pour aider les chercheurs à gérer, synchroniser, visualiser et analyser leurs données au travers de la création itérative d'artéfacts qui leur permettent d'améliorer ou de concrétiser la compréhension qu'ils ont de leurs données. Nous montrons enfin comment Tatiana peut être utilisé pour réaliser des analyses, et quel est son potentiel pour partager, au sein de la communauté des chercheurs, les corpus et les analyses.

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 a software 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.

MOTS-CLÉS : EIAH, CSCL, CSCW, méthodologie d'analyse, visualisation, traces, interactionnisme socio-cognitif.

KEYWORDS: CSCL, CSCW, analysis methodologies, visualisation, log files, socio-cognitive interactionism.

1. Introduction

L'étude sociocognitive des interactions humaines médiatisées par ordinateur peut s'effectuer au travers de l'enregistrement de ces activités, en particulier si ils ne sont pas limitées aux seules traces produites par les outils, mais incluent également des enregistrements audio et vidéo (Avouris *et al.*, 2007). Cox (2007) encourage les chercheurs à utiliser l'ordinateur et les diverses techniques qu'il permet (visualisation, data mining, etc.) pour effectuer leurs analyses des interactions dans ce qu'il appelle « données du processus » (« *process data* »). Cependant, les corpus d'interaction, en particulier lorsque ces interactions médiatisées se déroulent en face à face, sont difficiles à gérer d'un point de vue technique, et compliquées à appréhender du fait de la multiplicité et de la diversité des sources de données. Il ne suffit pas en effet d'analyser des flots de données séparés ; ces flots de données doivent être combinés pour permettre une compréhension globale des interactions qui ont eu lieu (Goodman *et al.*, 2006). Qui plus est, ces analyses doivent souvent être effectuées en équipe, que ce soit pour valider la méthode d'analyse au travers de la méthode des juges (De Wever *et al.*, 2006), pour répartir la charge de travail (Goodman *et al.*, 2006), ou pour combiner les perspicacités de plusieurs chercheurs de domaines différents (Prudhomme *et al.*, 2007).

Les difficultés décrites ci-dessus impliquent la nécessité d'outils qui puissent non seulement gérer cette variété et quantité de données, mais aussi permettre la visualisation et l'analyse dans un cadre de travail commun qui puisse être partagé avec d'autres chercheurs (cf. Reffay *et al.*, 2008 pour le travail de structuration de corpus d'apprentissage dans un but d'échange et de réutilisation ainsi que Suthers *et al.*, 2009 pour le besoin d'avoir des outils pour gérer le partage d'analyses). Les outils existants (par exemple, Greenhalgh *et al.*, 2007 et Fiotakis *et al.*, 2007) montrent les possibilités de l'analyse assistée par ordinateur mais sont spécialisés à certains types de données ou à certains types d'analyse et nécessitent des développements supplémentaires pour fonctionner de manière interopérable (Kahrimanis *et al.*, 2006). Nous proposons la création d'un outil d'aide à l'analyse qui soit générique, extensible et qui puisse être utilisé en combinaison avec d'autres outils.

Dans cet article, nous donnons un aperçu de quelques-unes des opérations effectuées par les chercheurs qui analysent les traces d'interactivités entre individus, en particulier lorsque ces activités sont médiatisées par ordinateur (nous nous concentrerons sur notre domaine d'application, le CSCL¹, mais ces pratiques s'étendent à d'autres domaines des STIC, notamment le CSCW²). Nous présentons ensuite un modèle simple de description de ces activités, et montrons comment le logiciel Tatiana (*Trace Analysis Tool for Interaction ANALysis*) a été conçu pour soutenir ce modèle, que nous considérons comme généralisable à d'autres types

1. *Computer-Supported Collaborative Learning* ou Apprentissage collaboratif médiatisé par ordinateur.

2. *Computer-Supported Collaborative Work* ou Travail collaboratif médiatisé par ordinateur.

d'analyses.. Nous terminons enfin en montrant, au travers de dix études de cas comment Tatiana est conçu pour s'adapter à de nouvelles formes d'analyses, et encourage le partage et la collaboration dans les recherches sur l'interaction humaine.

2. Comment se déroule l'analyse ?

Un corpus typique d'interactions médiatisées par ordinateur peut contenir des enregistrements audio et vidéo, de fichiers de traces d'interactions médiatisées par ordinateur, des interviews, des notes de terrain, des tests précédant et suivant l'expérimentation, et la description de celle-ci (finalité, contexte, scénario, etc.). Lors de la conception de notre outil destiné à permettre l'analyse d'un tel corpus, nous avons tenté de répondre à la question suivante : « que font les chercheurs de ces données ? »

Bien que cette question puisse être abordée sur le plan méthodologique, parvenir à un recensement raisonnable des pratiques existantes dans le domaine du CSCL (et de celles de communautés proches) aurait conduit à mener une méta-analyse d'un très vaste ensemble de recherches ; le faire d'une manière satisfaisante déborderait largement du cadre de cet article. Lors d'une telle analyse, nous devrions couvrir une variété de méthodologies d'apparences similaires mais épistémologiquement difficilement compatibles (e.g. Law *et al.*, 2009). De plus, nous pourrions nous livrer à une telle analyse, mais elle montrerait ce qu'*obtiennent* les chercheurs, et pas nécessairement de *quelle manière* ils l'obtiennent, dans la mesure où cette démarche est très rarement décrite dans la littérature. Dans cette section, nous présentons notre compréhension de l'activité d'analyse en explorant un sous-ensemble des thèmes d'analyse (c'est à dire des activités analytiques en apparence similaires, malgré d'éventuelles différences épistémologiques) et certains des artéfacts que créent les chercheurs en liaison avec ces thèmes. Nous nous appuyons, pour étayer nos thèses sur des articles décrivant des questions ou des expérimentations méthodologiques, ainsi que sur les outils existants, destinés à aider l'analyse.

2.1. Interrelation entre données, représentations analytiques et épistémologies

Suthers *et al.* (2009) décrivent un cadre analytique pour les interactions médiatisées qu'ils nomment *graphe de contingences*. Ils définissent trois niveaux de description de ce concept : le niveau empirique, le niveau représentationnel et le niveau conceptuel. Au niveau empirique, un graphe de contingences est composé d'évènements (des changements observés dans l'environnement) et de contingences avérées entre ces évènements (e.g. un message répond à un autre, un message suit un autre, un message reprend une phrase utilisée dans un autre message, etc.). Au niveau représentationnel, un graphe de contingences est un graphe avec des nœuds (représentant les évènements) et des arêtes (représentant les contingences). Au

niveau conceptuel, ce graphe représente des coordinations médiatiques entre les participants, exprimant la frontière entre le domaine privé et le domaine public, et certaines des contingences peuvent contribuer à justifier une relation *d'uptake* entre ces coordinations, définie par la considération de certains aspects contenus dans d'autres coordinations en tant que pertinentes à la poursuite de l'activité. Il y a donc *uptake*, par exemple, lorsqu'un participant reprend une idée d'un autre participant et la réitère, l'adopte ou la transforme. Suthers *et al.* (2009) proposent d'une part que l'apprentissage collaboratif peut être analysé via les notions conceptuelles de coordination médiatique et d'*uptake* (basé sur leur conception épistémologique de l'apprentissage collaboratif) et, d'autre part, que les aspects empiriques et représentationnels pourraient former une base commune d'analyse d'interactions médiatisées. L'idée est séduisante mais, malgré une solide base théorique, ce cadre analytique n'a encore été appliqué qu'à un nombre limité de cas. De plus, Suthers et ses collègues notent la controverse soulevée par la notion que ce cadre pourrait être d'utilité universelle (*ibid*, p. 33).

La séparation en trois niveaux nous paraît particulièrement pertinente et applicable à notre problématique. En effet, on imagine aisément que d'autres niveaux conceptuels, basés sur d'autres épistémologies, pourraient être calqués sur les niveaux représentationnels et empiriques des graphes de contingences (par exemple le graphe pourrait, cf. §5.6). A leur tour, d'autres niveaux représentationnels pourraient être calqués sur le niveau empirique constitué d'évènements et de contingences. On pourrait imaginer une analyse structurelle des contingences et évènements produisant, par exemple à niveau représentationnel correspondant à une typologie structurelle et des fréquences d'observation.

Notre désir étant de fournir une plateforme générique pour la création d'artéfacts analytiques, il nous semble naturel de proposer qu'il serait intéressant de fournir cette généricité au niveau représentationnel, en adoptant une représentation empirique unique et en laissant aux chercheurs le soin de calquer leurs niveaux conceptuels sur l'espace de représentations que nous serons en mesure de fournir. Nous pouvons maintenant examiner, par le biais de quelques thèmes analytiques, la variété des niveaux représentationnels utilisés par différents chercheurs.

2.2. Quelques thèmes d'analyse

Dans cette section, nous examinons l'état de l'art des types d'artéfacts créés pendant l'analyse ainsi que des outils et techniques pour les créer et manipuler. Ces artefacts décrivent l'ensemble représentationnel (pour reprendre la terminologie de Suthers) dont nous souhaitons pouvoir soutenir le processus de création et d'exploitation.

2.2.1. Codage et Comptage

Plusieurs types d'analyses utilisent le codage et le comptage, le plus souvent pour des analyses de contenu (Strijbos *et al.*, 2006). Cette pratique est tellement commune qu'il existe une importante littérature autour des techniques variées d'intelligence artificielle destinées à aider à automatiser cette tâche de comptage (Rosé *et al.*, 2008; Erkens *et al.*, 2008). Dans l'application de ce type de méthodes, une question importante est celle de l'unité d'analyse (De Wever *et al.*, 2006), à savoir la granularité pertinente : est-ce le tour de dialogue, une proposition au sein d'un tour de dialogue, une succession de tours de dialogue ? Une autre question est celle du codage à appliquer, qui peut être tiré de la littérature (e.g. Baker *et al.*, 2007) ou être spécifiquement créé pour une analyse particulière. Une fois ce schéma appliqué se pose la question de sa validation (De Wever *et al.*, 2006). Enfin, il convient d'effectuer les calculs statistiques désirés.

Quel que soit le choix de l'unité d'analyse, le corpus doit être découpé en de telles unités. Selon la nature du médium, différentes segmentations sont possibles. La transcription des dialogues et des gestes humains à partir d'un support audio ou vidéo s'effectue typiquement avec un outil tel que Elan (<http://www.lat-mpi.eu/tools/elan/>). Les traces des outils informatiques sont transformées en séquences d'actions, qui peuvent alors être regroupées, re-segmentées ou filtrées. Dans tous les cas, le résultat se présente sous formes d'évènements (ou d'actions), dotés de propriétés telle que la date, l'outil, l'utilisateur, le contenu échangé, etc. Plus concrètement, ces données peuvent se représenter dans un outil tel qu'ExcelTM, avec une ligne pour chaque évènement et une colonne pour chaque propriété. Parmi les logiciels de support à l'analyse, Elan, Videograph® (<http://www.ipn.uni-kiel.de/aktuell/videograph/enhtmlStart.htm>) et DRS (Digital Replay System, Greenhalgh *et al.*, 2007) permettent directement l'annotation et le codage des vidéos.

Les schémas de codage et d'évaluation sont de natures très diverses. Ils sont souvent le résultat d'un processus itératif impliquant un codage partiel du corpus (par exemple en catégories A, B et C), puis une redéfinition du schéma de codage (par exemple pour fusionner les catégories A et B et séparer la catégorie C en deux catégories distinctes), voire un travail collaboratif de codage (De Vries *et al.*, 2002). Il est rare qu'un schéma de codage préexistant soit appliqué sans aucune modification. Cet aspect est important pour la conception du logiciel, puisqu'il implique de devoir prendre en compte les codages selon des catégorisations évolutives. Le codage peut dès lors être considéré comme l'ajout, par la personne effectuant l'analyse, d'une ou de plusieurs propriétés à chaque évènement. La nature subjective de cette activité, ainsi que la nécessité d'effectuer ultérieurement des analyses statistiques sur les données codées impliquent une validation de ce codage. Il est fréquent de voir appliquer la méthode des juges (De Wever *et al.*, 2006), qui, indépendamment de la métrique utilisée (le kappa de Cohen, l'alpha de Krippendorff, etc.), implique de comparer côte à côte au moins deux codages, qu'ils soient complets ou partiels, afin de vérifier leur compatibilité et des les valider. Les

données sont alors regroupées en fonction des codes attribués et de certaines autres caractéristiques (comme l'utilisateur, le groupe, l'outil, etc.). Par l'observation des distributions obtenues, et par l'application d'outils statistiques, certaines hypothèses peuvent être confirmées ou infirmées.

2.2.2. Signets, collections et annotations

D'autres formes d'analyse, notamment les études de cas (ex. Rummel *et al.*, 2008), l'analyse conversationnelle (Sacks *et al.*, 1974) et les analyses basées sur la théorie de l'activité (ex. Avouris *et al.*, 2007), consistent à décrire le corpus par l'ajout d'annotations, la création de collections ou de groupes d'évènements. Ces opérations peuvent s'effectuer manuellement (dans un fichier texte, ou avec un tableur), ou encore au moyen d'outils dédiés, comme ActivityLens (Fiotakis *et al.*, 2007), Videograph ou Transana (<http://www.transana.org/>).

2.2.3 Synchronisation temporelle

La nécessité de synchroniser les différentes sources de données ainsi que les artefacts construits par le chercheur est évidente : les évènements observés — au moment de l'expérimentation — sont situés temporellement, et doivent être replacés dans leur contexte afin de pouvoir être compris. De plus, les différentes vues des mêmes évènements (p.ex. vidéo et transcription) se complètent bien lorsqu'elles sont synchronisées. En fait, c'est également une nécessité scientifique que de disposer de cette synchronisation (ou au minimum de pouvoir se reporter aux données primaires) : lorsque l'on cherche à établir une hypothèse, la question qui se pose systématiquement est : « dispose-t-on dans les données primaires d'éléments soutenant cette hypothèse ? ». Les chercheurs ont besoin d'artefacts qui présentent le corpus sous une forme mieux adaptée à la compréhension, l'affichage ou l'analyse (comparons par exemple des données audio ou vidéo et leur transcription). Cependant, quand une hypothèse est avancée à partir d'informations découvertes dans de telles données secondaires, il est nécessaire de vérifier que l'on retrouve effectivement dans les données originelles la confirmation de cette hypothèse. La synchronisation offre un moyen de revenir aisément aux données primaires chaque fois qu'une telle confirmation est nécessaire.



Figure 1. Dans cet exemple, les évènements sélectionnés dans la vue de gauche sont mis en évidence dans la vue aérienne (centre) et synchronisés avec les vidéos (à droite). (www.dcs.gla.ac.uk/~morrissaj/Replayer.html)

La synchronisation des différents média sources est prise en compte, à des degrés divers, par nombre d'outils : Elan, ActivityLens, DRS, Replayer (Morrison *et al.*, 2006) et ABSTRACT (Analysis of Behaviour and Situation for menTal Representation Assessment and Cognitive acTivity modeling; Georgeon *et al.*, 2007; Georgeon *et al.*, 2006) permettent tous la synchronisation des artéfacts qu'ils construisent (c.f. Figure 1).

2.2.4. Visualisations graphiques

Les visualisations graphiques permettent aux chercheurs d'avoir des visions différentes des données, et aident souvent à isoler des phénomènes intéressants, ou fournissent une vision intuitive de ce qui s'est passé. Une forme typique de visualisation graphique est la représentation symbolique des événements sur un axe temporel horizontal. Ce type de représentation est bien ancré dans les habitudes, et se retrouve dans SAW (Synchronized Analysis Workspace, Goodman *et al.*, 2006), ABSTRACT (cf. Figure 2), Replayer, DRS et d'autres encore. Cette représentation permet d'explorer la dimension temporelle des données. Le graphe de contingences de Suthers *et al.* (2009) augmente ce type de visualisation d'annotations relationnelles. Teplov *et al.* (2007) utilisent les données des fichiers de traces d'un forum structuré pour générer une visualisation interactive des messages aux propriétés similaires (même auteur, même fil de discussion, proximité sémantique). Il existe de nombreux types de visualisations dont le but est de représenter les connaissances analytiques à propos de données, telles que les représentations de conceptions sous formes de diagrammes argumentatifs (Prudhomme *et al.*, 2007) ou encore les visualisations liées aux outils d'awareness (van Diggelen *et al.*, 2008).

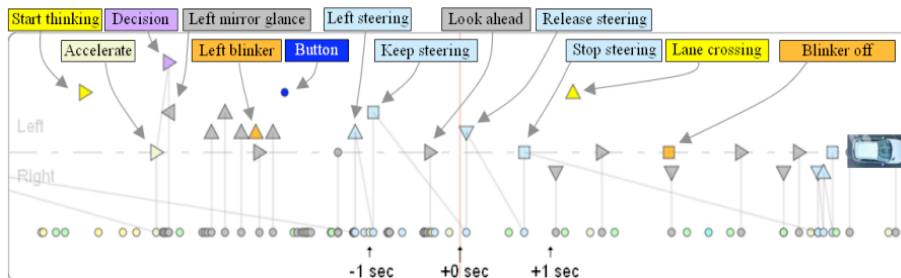


Figure 2. Représentation graphique symbolique des événements enregistrés lors d'un changement de file sur une autoroute, visualisés avec ABSTRACT (<http://liris.cnrs.fr/abstract/>).

3. Un modèle simple d'analyse

Dans la section précédente, nous avons décrit l'espace des représentations que nous souhaitons pouvoir générer. Dans cette section, nous structurons cet espace en termes de types d'artéfacts et de manipulations et examinons le niveau empirique

sous-jacent nécessaire à la modélisation de ces artefacts en vue de la création d'un environnement informatique qui assiste ces analyses.

Harrer *et al.*, (2007) ont modélisé le processus d'analyse dans l'idée de s'appuyer sur ce modèle pour concevoir la structure d'un outil d'analyse et rendre les représentations interopérables. Ils décrivent les différentes phases de traitement : capture de donnée, segmentation, annotation, analyse, visualisation, interprétation ainsi qu'une boucle d'itération permettant de retourner à l'annotation.

Dans l'état actuel de notre compréhension de l'analyse, basée sur nos études de cas et nos expériences passées (de Vries *et al.*, 2002; Prudhomme *et al.*, 2007; Baker *et al.*, 2007), ce qui nous semble le plus important dans ce modèle est la boucle itérative : les chercheurs qui en arrivent à la phase d'interprétation et ne sont pas satisfaits des résultats vont affiner leur analyse en réitérant le processus, jusqu'à parvenir à un résultat susceptible d'être transmis à la communauté scientifique. Il ne semble pas non plus nécessaire de parcourir toutes les étapes dans un ordre particulier, ni d'attendre la phase d'interprétation pour décider si une nouvelle itération est nécessaire. Par exemple, la segmentation des données en unités d'analyses est intrinsèquement liée au type de codage auquel le chercheur souhaite procéder ; si ce codage s'avère impossible pour certaines unités, il peut être nécessaire de revoir immédiatement la segmentation sans procéder à la suite de l'analyse.

Nous nous sommes appuyés, pour la conception de Tatiana, sur un modèle similaire (quoique simplifié), qui place plus d'emphasis sur la nature itérative de l'analyse (cf. Figure 3). Les analystes évaluent en permanence si la combinaison de leurs données primaires et des artefacts secondaires est suffisante pour obtenir un résultat. Si ce n'est pas le cas, ils sont amenés à créer un nouvel artefact, qui permet de mieux comprendre leurs données, ou de réifier leur vision de ces données.

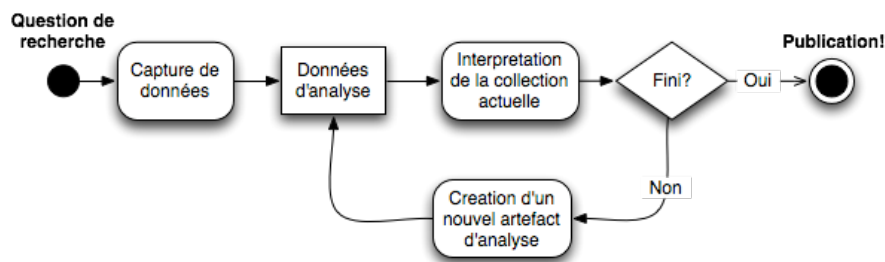


Figure 3. Représentation graphique du modèle de traitement de Tatiana

Dans les schémas d'analyse présentés ci-dessus, nous avons montré la variété des artefacts que les chercheurs sont amenés à construire. Dans certains cas, cette création peut être automatisée (p.ex. la transformation d'une représentation vers une

autre, ou des calculs statistiques). Dans d'autres cas, cette transformation est entièrement manuelle (telle la construction de certaines visualisations), ou peut être aidée par un outil (p.ex. la transcription, l'annotation, le codage). Ces artefacts sont souvent des représentations des données faisant appel à la dimension temporelle, au travers de différentes approches : la séquence des événements enregistrés est présentée au serveur par un outil de rejouage (auquel est associée une télécommande permettant la navigation dans les données), ou sous une représentation graphique ou tabulaire, le temps étant associé à l'axe horizontal ou vertical.

Ces artefacts peuvent être classifiés en trois espèces. Les *agrégations* résument les données correspondant à une certaine période de temps, afin de permettre la production d'indicateurs ou de statistiques. Les *enrichissements* sont des artefacts qui ajoutent une connaissance créée par le chercheur après examen des données (ces analyses consistent en codages, annotations, liens, etc.). Enfin, nous proposons que les artefacts qui conservent la notion d'ordre des événements et d'interactions temporellement situées soient nommés *rejouables*. Ce sont des objets qui peuvent être rejoués, synchronisés et enrichis. Plusieurs exemples de tels objets seront présentés en §5. Nous proposons également de considérer le processus d'analyse comme une création itérative de nouveaux artefacts (tels que les rejouables et les enrichissements), qui mettent en lumière la compréhension des données par le chercheur, ou lui permettent d'affiner cette compréhension. Nous nous intéressons en particulier aux rejouables, car ils servent fréquemment de sources pour la création de nouveaux artefacts, et en particulier d'autres rejouables. Parmi les transformations de rejouables en d'autres artefacts on trouve les transcriptions, annotations, codages, visualisations, filtrages, synchronisations, fusions ou agrégations. Il est aussi possible de considérer que la création d'un enrichissement se ramène à celle d'un rejouable, dans lequel certains événements sont dotés d'une nouvelle propriété. Dans tous les cas, considérer les enrichissements comme des entités indépendantes permet de les réifier et de les transmettre à d'autres chercheurs en possession du même corpus, qui peuvent dès lors les intégrer à leurs données.

4. Tatiana : un environnement générique d'analyse

Dans les sections précédentes, nous avons montré certaines des représentations analytiques créées pendant l'analyse, ainsi que leur cycle de vie. Bien qu'il existe nombre d'outils destinés à assister certaines de ces tâches, même les plus généraux de ces outils, comme ActivityLens ou Digital Replay System ont des carences autour d'aspects tels que l'automatisation de transformations, la facilité d'introduire de nouvelles données, ou la possibilité de s'adapter à de nouveaux types d'analyses. Tatiana (Trace Analysis Tool for Interaction ANALysts) est un environnement conçu pour la manipulation des divers types d'artefacts décrits ci-dessus, les rejouables en particulier. Nous présentons dans cette section les diverses caractéristiques du logiciel qui permettent de gérer ces possibilités.

Le logiciel Tatiana repose sur un certain nombre de concepts et composants fondamentaux (cf. Figure 4) et particulièrement autour de la notion de rejouable. Ces rejouables peuvent être créés automatiquement (par exemple, par importation de données extérieures ou par des transformations) ou manuellement. Une fois créés, les rejouables bénéficient de quatre fonctionnalités : ils peuvent être transformés, enrichis, visualisés ou synchronisés. Afin de coordonner ces fonctionnalités, les rejouables sont enregistrés sous un format pivot interne à Tatiana, basé sur une représentation sous forme d'évènements ayant un certain nombre de propriétés.

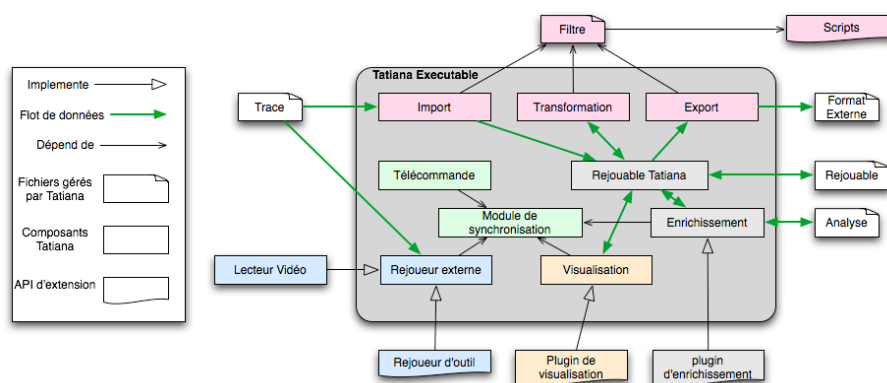


Figure 4. L'architecture de Tatiana faisant apparaître les les composants et les possibilités d'extension de ces composants.

4.1. Transformations

Les rejouables peuvent être transformés (automatiquement ou interactivement) et exportés. L'importation, la transformation et l'exportation s'effectuent par l'applications de filtres. Un filtre combine des scripts sous forme d'un flux de traitement de données. Les scripts sont de petits programmes (écrits dans le langage XQuery), conçus pour exécuter une opération spécifique : transformer un fichier d'un corpus en la représentation pivot de Tatiana, sélectionner certains évènements d'un rejouables, combiner des rejouables, etc. Comme nous n'attendons pas des chercheurs (qui, pour la plupart, ne connaissent pas XQuery) qu'ils réalisent eux-mêmes leurs scripts, nous avons conçu un éditeur interactif graphique, qui permet de construire un filtre générant le rejouable désiré par le chercheur à partir d'un ensemble de scripts et d'autres filtres existant. Un filtre peut ainsi combiner un script réalisant l'importation de données créées par un logiciel externe et d'autres scripts appliquant divers traitements (regroupement automatique, sélection, extraction, fusion, etc.). Les transformations manuelles interactives permettent d'autres opérations sur les évènements, telles que la destruction, le réordonnancement, le regroupement ou la dissociation en de nouveaux évènements.

De nouvelles instances de transformations peuvent être ajoutées à Tatiana par la création de nouveaux scripts, et/ou filtres.

4.2. Enrichissement

Dans Tatiana, tous les rejouables peuvent être enrichis par des données analytiques créées par le chercheur. Trois types d'enrichissements sont actuellement supportées par le logiciel : les annotations, les catégorisations et l'enrichissement par graphe. La catégorisation consiste simplement à ajouter des annotations extraites d'une liste de termes, et peut servir au codage, à l'étiquetage ou à la pose de mots-clefs. La liste des catégories disponibles peut en fait être modifiée à n'importe quel moment de l'analyse, ce qui permet de faire évoluer le schéma pour l'adapter aux besoins du chercheur. L'enrichissement par graphe permet de poser des liens ou contingences entre des événements.

Ces analyses sont des instances particulières d'un concept générique d'enrichissement qui permet d'attacher à n'importe quel événement de rejouable une série de données analytiques supplémentaire. Un mécanisme de plugin est en place pour créer de nouvelles formes d'enrichissement (par exemple en créant la possibilité d'assigner une valeur scalaire plutôt qu'une catégorie).

4.3. Visualisation

Tous les rejouables peuvent être visualisés dans différents afficheurs. Il existe actuellement deux types d'afficheurs :

- Une forme tabulaire qui présente les données de manière similaire à celle d'Excel, avec une ligne par événement et une colonne pour chacune des propriétés associées aux événements, l'axe des temps étant vertical.

- Une visualisation temporelle horizontale qui est la première version d'un outil générique d'aide à la création automatique de visualisation. Chaque événement y est présenté sous forme d'un objet graphique dont les propriétés (couleur, forme, taille, position verticale, etc.) peuvent être choisies en fonction des propriétés associées à l'événement (utilisateur, outil, date, catégories d'analyse, etc.). Les contingences d'un enrichissement par graphe peuvent aussi être visualisés.

Un mécanisme de plugin permet d'intégrer de nouvelles formes de visualisation à Tatiana.

4.4. Synchronisation

Enfin, tous les rejouables dans Tatiana peuvent être synchronisés entre eux, et avec des données visualisées dans des afficheurs externes, tels que des lecteurs vidéo ou des outils de rejouage (ce terme de « rejouage » désignant un mode de fonctionnement particulier de certains outils de collaboration, qui peuvent relire les traces qu'ils ont produites lors de leur utilisation, et reproduire sur l'écran le déroulement d'une session antérieure). Tatiana propose un mécanisme permettant de

piloter ces outils externes. Le rejouage synchronisé implique que, lorsqu'une date est choisie dans la « télécommande », l'afficheur vidéo et les rejoueurs externes se trouvent immédiatement positionnés à la date correspondante dans leurs média respectifs, les évènements associés étant également mis en évidence dans Tatiana. (cf. Figure 5).

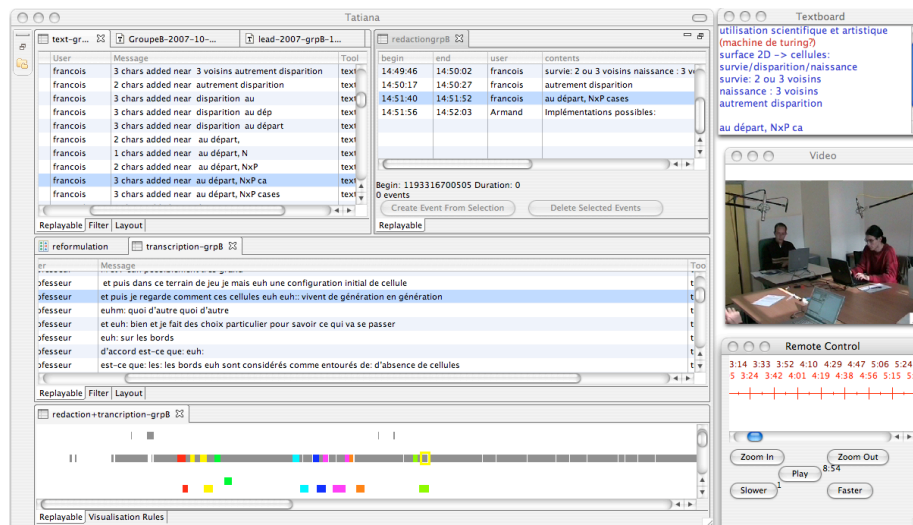


Figure 5. Affichage de différents rejouables dans Tatiana : trace d'un éditeur de texte partagé (haut gauche), transcription des dialogues (milieu gauche), unités de rédaction (haut centre), visualisation des reformulations (bas gauche), ainsi que la « télécommande » (bas, droit) permettant la synchronisation avec des outils extérieurs tels que le rejoueur DREW (haut droit) ou un afficheur vidéo (milieu droit). Cet exemple sera développé en §5.1.

De manière générale, la sélection d'un évènement dans une visualisation de rejouable positionne à cet instant précis tous les autres médias affichés. Par exemple, dans la Figure 5, lorsque le chercheur clique, dans la vue tabulaire, sur l'évènement « 3 chars added near au départ, NxP ca », le rejoueur montre l'état de l'éditeur de texte partagé juste après avoir inséré ces caractères. La dynamique de l'interaction devient ainsi immédiatement accessible au chercheur, alors qu'elle est difficile à découvrir dans les traces brutes de l'outil. Les opérations de « zoom » permettent d'afficher les épisodes à différents niveaux de granularité. L'existence de ces liens entre les rejouables permet au chercheur de se concentrer sur une visualisation particulière présentant des informations limitées, sachant qu'il lui sera possible, si nécessaire, de se référer instantanément à toutes les autres formes de représentation pour tout complément d'information.

5. Etudes de cas

Pendant le développement de Tatiana, nous avons effectué de nombreuses analyses dont nous rapportons ici une partie sous forme d'études de cas. Dans le cadre d'un cycle de conception itératif, ces études de cas nous ont permis d'affiner la notion de rejouable, d'améliorer l'implémentation de Tatiana (et d'en établir les priorités de développement) et d'évaluer le cadre analytique proposé. Dans cette section nous présentons dix études de cas. Nos quatre premières études de cas proviennent du projet Lead³ (Lead, 2006) lors duquel nous avons eu l'opportunité d'observer les activités de recherche de quatre équipes lors de la conception et de l'analyse d'expérimentations liées à ce projet. Toutes ces expérimentations ont pour objet l'utilisation en présentiel de communications médiatisées par ordinateur (les communications sont verbales ou médiatisées par ordinateur). Deux études de cas supplémentaires ont été effectuées lors de d'analyses en collaboration avec d'autres chercheurs. Enfin, les trois dernières études de cas proviennent d'analyses effectuées par d'autres chercheurs de manière indépendante et utilisant Tatiana. En présentant ces études de cas, nous ne pouvons entrer en détail sur les hypothèses et les suppositions théoriques, qui sont souvent liées à des études non encore publiées.

5.1. Etude de cas, Lyon/Saint-Étienne

Nous profitons de cette première étude de cas pour détailler un exemple d'utilisation de Tatiana. L'usage typique de Tatiana est la création itérative de rejouables et d'enrichissements, qui permettent au chercheur d'approfondir et de matérialiser sa compréhension du corpus.

Nous avons été les premiers utilisateurs de Tatiana durant les phases initiales du développement (avec, à certains moments, la casquette du développeur, à d'autres celle du chercheur en CSCL ou du testeur de l'ergonomie de l'outil). Afin, entre autres, de récolter un premier corpus pour Tatiana, nous avons observé neuf dyades sur une série de trois à cinq rendez-vous avec leur encadrant, dans le contexte du projet de programmation d'un cours d'introduction à l'informatique. Ces réunions se déroulaient en face à face, avec utilisation d'un « chat » et d'un éditeur de texte partagé, auxquels les intervenants avaient accès sur des ordinateurs portables. Les étudiants étaient encouragés à prendre leurs notes dans l'éditeur de texte partagé.

Lors de l'une des analyses, nous nous sommes intéressés à la provenance de ces notes, qui étaient très souvent des reformulations d'interventions orales. Nous voulions aboutir à une meilleure compréhension de ces reformulations d'un point de vue linguistique, et en explorer les conséquences pédagogiques éventuelles.

3. Le projet Européen LEAD (*Technology-enhanced learning and problem-solving discussions: Networked learning environments in the classroom*, <http://www.lead2learning.org/>) est financé par le 6^{ème} framework *Information Society Technology* LEAD IST-028027.

Pour parvenir à ces fins, nous avons d'abord effectué une transcription des dialogues dans l'outil Elan, que nous avons ensuite importés en tant que rejouables dans Tatiana. Nous avons également importé et transformé les traces de l'éditeur de texte partagé pour que ces traces fassent sens pour le chercheur⁴, créant ainsi un nouveau rejouable pour chaque séance. L'implémentation de cet éditeur de texte partagé est telle que des « événements » sont enregistrés chaque seconde. Nous voulions procéder à des analyses de plus haut niveau, et donc disposer d'informations contenant plus de sémantique. Ces événements ont donc été regroupés à la main dans un nouveau rejouable où chaque événement est une *unité de rédaction* cohérente (par exemple la rédaction d'une phrase). Nous avons ensuite créé deux enrichissements pour réifier les liens de reformulation entre la transcription et les unités de rédaction. La première, une analyse par catégories, définit une catégorie unique pour chaque instance de reformulation, cette catégorie étant associée à la fois à l'énoncé reformulé et à sa reformulations sous forme d'unité de rédaction. Par ailleurs (quelque temps plus tard en raison des aléas d'implémentation de Tatiana) nous avons créé un enrichissement par graphe montrant ces mêmes relations de reformulation. En fusionnant la transcription et les unités de rédaction ensemble, nous avons créé un rejouable unique doté de ces mêmes enrichissements et que nous avons visualisé sous forme de ligne temporelle horizontale (cf. Figure 6).

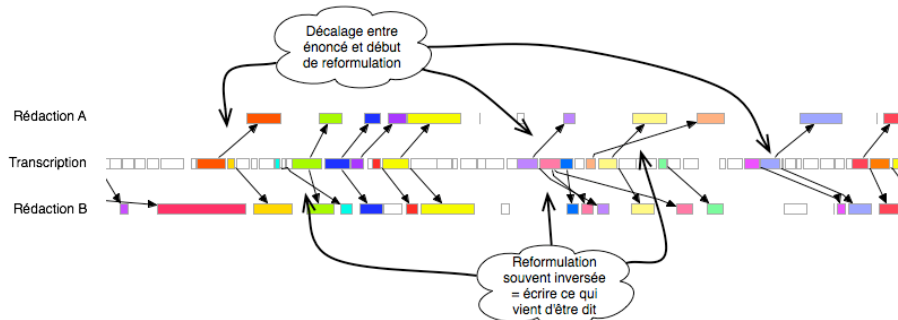


Figure 6. Un exemple de visualisation de reformulation. Les couleurs identiques ainsi que les flèches indiquent la reformulation.

Cette visualisation a été paramétrée pour montrer sur des lignes différentes les énoncés de l'enseignant et les reformulations des étudiants sous forme d'unités de rédaction. Les liens de reformulation sont montrés à la fois par des arcs entre les événements (provenant de l'enrichissement par graphe) et par des couleurs identiques (provenant de la catégorisation). Ces visualisations nous ont permis de

⁴ La forme particulière des traces de cet éditeur de texte partagé reproduit le texte complet de l'éditeur de texte après chaque modification, rendant difficile d'identifier exactement quelle modification a été effectuée. Nous avons tenté de fournir une information plus précise sur la nature de la modification.

mettre en évidence certains phénomènes intéressants, tels que l'inversion, lors de reformulations particulièrement longues entre l'ordre d'élocution de l'enseignement et de prise de note par les étudiants, et d'acquérir une compréhension intuitive du processus de reformulation. Plusieurs des rejouables présentés dans ce paragraphe sont affichés de manière synchronisée dans la Figure 5. Notons en particulier que malgré que les unités de rédaction et les liens de reformulation sont créés à la main, leur exploitation subséquente est automatisée.

5.2. Etude de cas, Paris

Dans ce second cas, nous nous sommes rendus sur place pendant deux jours, pour comprendre les pratiques de l'équipe et examiner comment Tatiana pouvait s'avérer utile dans le cadre de ces pratiques. Le résultat fut matérialisé par l'enregistrement audio et vidéo des discussions et d'une session de test de Tatiana, que nous ne détaillerons pas ici. L'une de leurs analyses impliquait l'étude, sur le long terme, de l'utilisation du logiciel CoFFEE (De Chiara *et al.*, 2007), développé durant le projet LEAD. Il s'agissait d'un suivi de paires de dyades travaillant en commun dans un espace partagé. Chaque dyade disposait d'une machine, et chaque paire de dyade travaillait dans l'espace commun, les deux machines des deux dyades étant placées côte à côte. L'équipe s'intéressait, dans cette analyse, aux types d'interactions (Bernard, 2008) se produisant au sein d'une dyade (focus sur l'outil ou sur le sujet du travail, communication orale ou médiatisée, etc.), en souhaitant découvrir le lien entre ces interactions et la collaboration entre les dyades.

Pour procéder à cette étude, ils désiraient jouer la vidéo en synchronisation avec les traces médiatisées, en opérant un marquage des blocs tandis que les différents types d'interaction se produisaient au sein de chaque dyade. Ces interactions seraient alors catégorisées selon un schéma de codification pré-établie, décrivant les différents types d'interaction, et permettant la comparaison en parallèle des interactions de chaque paire de dyade avec les interactions médiatisées. L'utilisation de Tatiana a dû être différée de plusieurs semaines en raison de l'impossibilité de créer un rejouable à partir d'une vidéo dans Tatiana à cette époque.

5.3. Etude de cas, Utrecht

Nous avons également passé deux jours in situ pour ce troisième cas, analysant les pratiques du groupe pour déterminer si Tatiana était adapté dans ce cas. Ce séjour se traduisit également par l'enregistrement vidéo et audio des discussions. Leur travail s'intéresse principalement à la production de diagrammes argumentatifs (c.f. Overdijk *et al.*, 2008). Une de leurs méthodes d'analyse consiste à conserver les seuls événements consacrés à la création de boîtes (quasi équivalent à la création d'un message), puis à réordonner ces événements pour faire apparaître la structure du diagramme (à la manière des fils de discussion d'un forum) plutôt que leur enchaînement chronologique. A partir de cette production, ils peuvent appliquer sur les données les méthodes d'analyse de discours ou de contenu.

Un autre processus d'analyse utilisé est la création de visualisations qui montrent comment l'activité des étudiants se déploie dans les différents fils de discussion au cours du temps. Ces vues combinées (chronologiques, par fil de discussion, par étudiant) offrent une bonne visibilité et permettent une meilleure compréhension de la situation (van Diggelen et al., 2008).

Au moment de cette visite, les enrichissements par graphe n'existaient pas encore et il n'était pas possible de réordonner les événements d'un rejouable dans le cadre d'une visualisation. Ces deux choses sont maintenant possibles.

5.4. Etude de cas, Nottingham

Dans ce dernier cas du projet LEAD, nous n'avons pas eu l'opportunité d'effectuer de visite sur le site, mais, comme avec les autres partenaires, nous avons pu avoir de nombreuses discussions au cours des réunions du projet LEAD ou par email. Une de leurs pratiques consiste à utiliser des méthodes statistiques (telles que celles qui sont disponibles dans SPSS ou Excel) au cours de leurs analyses. Dans une de leurs études (faisant suite à Gelmini Hornsby *et al.*, 2008), ils étaient intéressés à comprendre comment différentes circonstances influent sur les changements d'opinions exprimées dans le module de vote de CoFFEE. Le scénario comportait un premier vote, une discussion, puis un second vote. Avant la fermeture de chaque scrutin, les étudiants pouvaient changer d'opinion autant de fois qu'ils le désiraient, ce qui était enregistré dans les traces du logiciel. Les chercheurs devaient chercher manuellement le dernier changement d'opinion pour chaque étudiant et chaque vote dans les traces, afin de disposer des données nécessaires à leur étude.

Ils ont aussi exprimé un intérêt particulier pour le calcul automatique d'indicateurs statistiques assez généraux en relation avec les discussions, tels que le nombre de mots par tour de parole, par tour et par étudiant, le nombre d'interventions par étudiant et par sujet, etc. Ils ont également confirmé l'importance, pour tout nouvel outil, de pouvoir s'intégrer à leurs pratiques habituelles (utilisation de Excel, SPSS, mais aussi d'autres outils de transcription ou d'analyse). Nous avons créé des scripts pour répondre à ces besoins, dont une partie (filtres par utilisateur, par groupe, etc.) sont réutilisables dans d'autres cas.

5.5. Etude de cas, Conception collaborative en ingénierie

Dans ce cas, nous avons collaboré avec des collègues dans un projet région qui étudie les réunions de conception collaborative. Lors de l'analyse du corpus, consistant en une vidéo de réunion lors de laquelle douze points de décision sont abordés, la transcription a été importée dans Tatiana, puis codée selon deux schémas de catégorisation, le premier décrivant le sujet de l'élocution, le second sa fonction illocutoire (e.g. question, contre-argument, etc.). Tatiana a permis la conciliation de codages effectués séparément par des experts de différents domaines (conception collaborative, argumentation, ingénierie mécanique) et diverses corrélations statistiques. Un enrichissement par graphe a permis de réifier les articulations entre

solutions et critères, de manière à ce que cette articulation, créée à la main, puisse aussi être automatiquement visualisée sous forme de ligne de temps horizontale, montrant, par exemple, comment l'évaluation de la solution est répartie entre les intervenants et au fil du temps (cf. Figure 7).

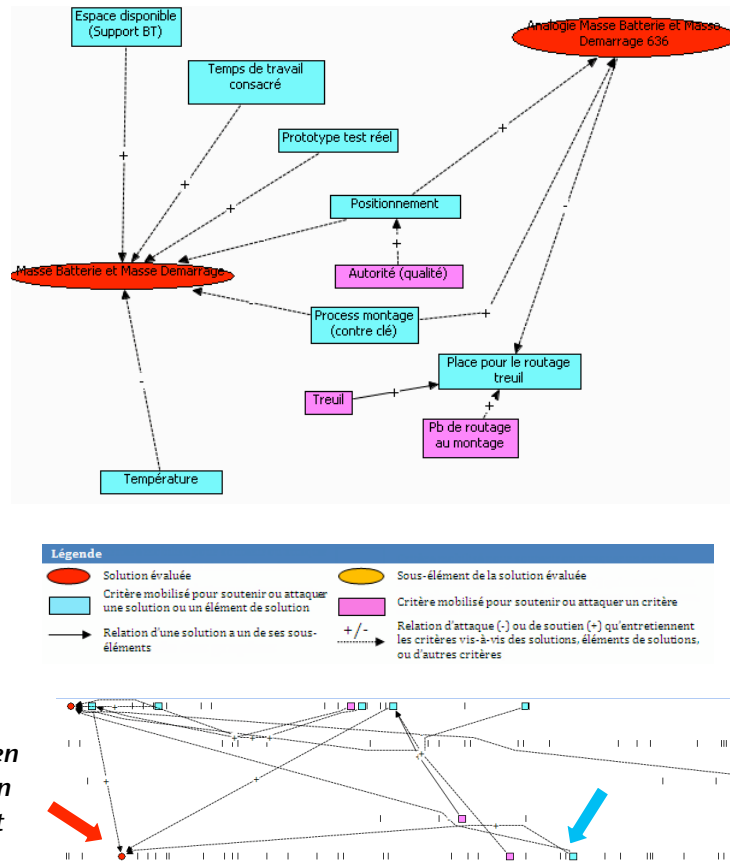


Figure 7. Deux solutions (en rouge) sont comparées selon une variété de critères (en bleu) dans un enrichissement par graphe (en haut). Ce graphe est ensuite visualisé sous forme de ligne temporelle horizontale montrant la répartition par intervenant et l'évolution au fil du temps (Cassier, en préparation).

5.6. Etude de cas, Description de la structure d'un dialogue de chat

Lors d'un projet Européen précédent nommé SCALE⁵, une étude fut menée comparant des situations d'argumentation uniquement par chat avec des situations

⁵ Internet-based intelligent tool Support Collaborative Argumentation-based LEarning in secondary schools, IST-1999-10664, finance sous le plan IST, framework-VI.

utilisant à la fois du chat et un éditeur partagé de graphes argumentatifs (Amelsvoort *et al.*, 2008). De manière similaire à l'étude de cas précédente, de nouvelles analyses ont été entamées grâce à Tatiana, visant d'une part à mettre en évidence la structure argumentative du chat afin de la comparer avec celle des graphes, et d'autre part à représenter l'évolution des structures argumentatives au fil du temps dans les deux médiums.

5.7. Etude de cas, Tension et relaxation

Cette étude de cas est la première pour laquelle nos collaborateurs n'étaient pas sous obligation d'utiliser Tatiana (dans le sens que nous n'étions pas impliqués ensemble dans un projet formalisé). L'analyse a porté sur des situations de travail en groupes par dialogue oral. Cette analyse est basée sur la notion théorique selon laquelle une collaboration efficace doit être constituée d'un équilibre entre des moments de tension (désaccord, critiques) et de moments de relaxation (interactions sociales, accord). Stéréotypiquement, lorsqu'il y a trop de tension les partenaires ne voudront plus collaborer et lorsqu'il y a trop de relaxation, les problèmes clés ne seront pas identifiés et résolus (Baker *et al.*, 2009).

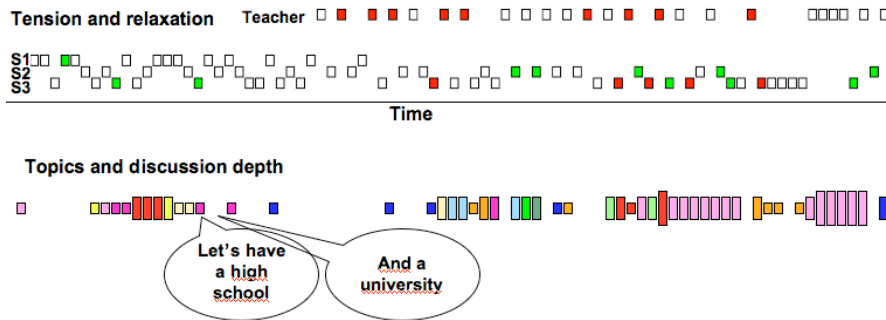


Figure 8. Deux visualisations examinées de manière synchronisée au fil du temps. En haut : les rangées correspondent aux intervenants, le vert à la relaxation, le rouge à la tension. En bas : Les topics sont montrés de différentes couleurs, la hauteur des éléments correspond au niveau de profondeur (Baker *et al.*, 2009).

Nos collègues ont souhaité utiliser Tatiana pour visualiser cet équilibre de manière intuitive et le comparer avec le degré de profondeur des discussions afin de montrer que les discussions déséquilibrées avaient aussi un niveau plutôt superficiel. Pour ce faire, ils ont enregistré les dialogues sous forme de vidéo et ont importé la transcription dans Tatiana. Les dialogues ont ensuite été codés selon trois analyses par catégorisation : le premier distingue les élocutions sources de tension des élocutions sources de relaxation et des élocutions neutres, le second décrit le concept dont il est question dans chaque élocution, et le troisième décrit qualitativement la « profondeur » de chaque contribution sur une échelle de 1 à 3. Deux visualisations en graphe temporel horizontal de chacun des jouables résultants ont

été créées, montrant respectivement la tension et la relaxation par intervenant et la profondeur de discussion par concept. Ces visualisations ont permis de caractériser différents types de discussion (par exemple un enseignant qui entrave la progression d'un groupe en montant trop le niveau de tension) et d'illustrer certaines relations entre tension-relaxation et profondeur (cf. Figure 8).

5.8. Etude de cas, MULCE

Les chercheurs du projet MULCE travaillent sur le partage de corpus d'apprentissage en vue de réplique et de validation de résultats (Reffay *et al.*, 2008). Ils ont choisi d'utiliser le format Tatiana, d'une part pour partager certaines de leurs données et d'autre part pour partager certains de leurs analyses sur ces données. Sans assistance particulière de notre part (hormis la documentation existante), ces chercheurs ont non seulement transformé des données au format Tatiana mais ont aussi recréé plusieurs de leurs analyses existantes, consistant en des codages et en des visualisations inédites sous forme de graphe temporel de ces codages⁶. Cette étude de cas montre l'utilité des notions établies dans Tatiana pour le partage de données et d'analyse, ainsi que les nouvelles possibilités ouvertes par la réification d'analyses sous forme re-exploitable.

5.9. Etude de cas, Blogs et Twitter

Dans cette étude de cas, un chercheur du Danemark nous a sollicité pour l'assister dans l'utilisation de Tatiana pour l'analyse de données issues de blogs et de Twitter afin de comprendre la structure des communautés soutenues par ces médias (Lomborg, 2009). L'analyse a ensuite essentiellement utilisé les fonctions d'annotation et de calcul statistique intégrés sous forme de transformation dans Tatiana. Différents rejouables ont été créés représentant des collections de posts et de commentaires de nature similaires.

5.10. Etude de cas, Hong Kong

Dans cette étude de cas, nous avons collaboré par email avec des collègues de Hong Kong souhaitant visualiser des données issues d'un forum et préalablement codées dans Excel. L'idée était d'utiliser ces visualisations pour identifier des hypothèses de rapports entre différents les différents codages. Ces codages étaient cependant plutôt des *tags*, pouvant se cumuler à plusieurs de même nature par message. De plus, la visualisation désirée aurait du produire un élément graphique par code utilisé. Or, la visualisation par ligne graphique temporelle existante de Tatiana ne produit qu'un élément par événement. Une solution a cependant pu être trouvée au travers de laquelle, lors de l'importation depuis Excel, les événements comportant des codages multiples sont démultipliés (cf. Figure 9).

⁶ Ce corpus est disponible en ligne et peut être consulté sous http://mulce.univ-fcomte.fr/metadata/doc/visu-corpus/copeas-T5_contexte.xml

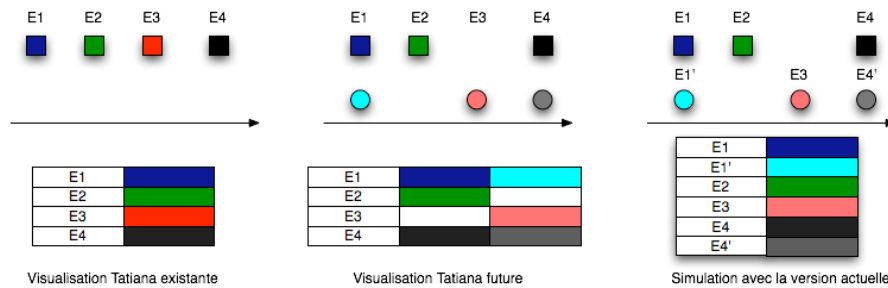


Figure 9. A gauche, la relation entre les données visualisées sous forme tabulaire et sous forme de ligne de temps (Un élément graphique par évènement). Au milieu, la visualisation désirée (un élément graphique par code appliqué). A droite, la solution retenue sans développement de plugins supplémentaires (démultiplication d'autant d'éléments qu'il y a de codes à appliquer à cet élément).

Cette étude de cas semble montrer un défaut de Tatiana. D'un côté, il est vrai que Tatiana ne comporte pas la visualisation désirée, ni la possibilité de coder de cette façon. En revanche, il est possible de répondre à ce besoin d'étendre Tatiana en créant un nouveau plugin de visualisation ainsi qu'un nouveau plugin d'enrichissement, sans modifier quoi que ce soit d'autre. Il s'agit donc d'une confirmation supplémentaire des concepts sur lesquels Tatiana est basé.

5.11. L'apport des études de cas

Ces diverses études de cas ne peuvent prouver l'universalité de Tatiana. Elles apportent cependant plusieurs éléments confirmant l'intérêt du système proposé. D'une part, elles couvrent une variété de types de données (vidéo, chat, blog, forum), de sujets d'étude (interactions langagières, conception collaborative, argumentation) et d'épistémologies et méthodologies d'analyse (allant du quantitatif au qualitatif et de l'analyse conversationnelle aux sciences cognitives). D'autre part, en examinant de plus près l'évolution des études de cas dans le temps, on remarque que le type de modification à faire change. Une hiérarchie des changements peut être établie dans laquelle un changement à un niveau donné nécessite une réévaluation tous les niveaux supérieurs (cf. Figure 10). Ainsi, les problèmes d'ergonomie, tout en haut de cette hiérarchie, n'ont pas d'autres conséquences. En deuxième, la création de nouvelles transformations ou de nouveaux plugins de visualisation et d'enrichissement provoquera, au pire, de nouveaux problèmes d'ergonomie. En revanche, une modification de l'implémentation de Tatiana ou du modèle de données derrière (3^{ème} niveau) peut invalider les plugins et transformations existantes. Au dernier niveau, on trouve la définition de la notion de rejouable et des opérations sur les rejouables. Lorsque des problèmes sont identifiés à ce niveau, tout est à revoir.

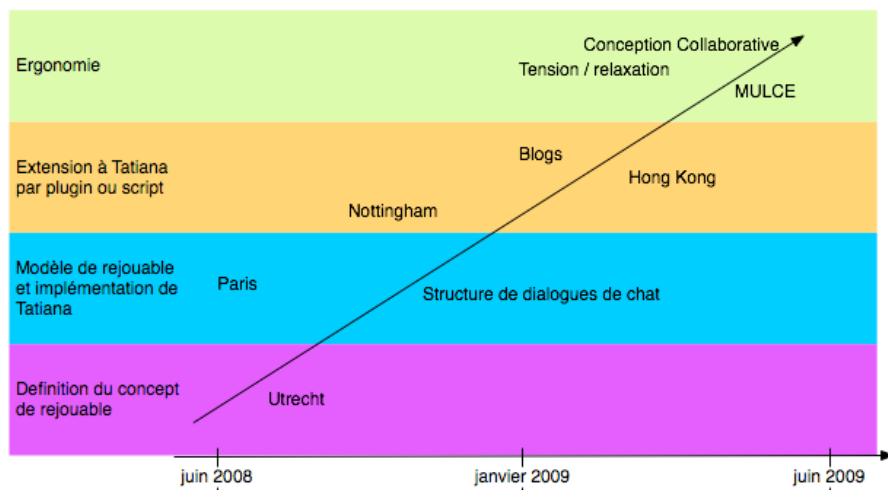


Figure 10. *Évolution des études de cas au fil du temps. Les premières études de cas ont nécessité des modifications importantes de notre définition de la notion de rejouable et de l'implémentation de cette notion dans Tatiana. Les études de cas plus récentes soulèvent plutôt des problèmes d'ergonomie et demandent le développement de nouvelles instances de transformation, de visualisation et d'enrichissement*

L'étude de cas Saint-Etienne/Lyon a influencé notre réflexion dès le début et à tous les niveaux. Les études de cas de Paris, Utrecht et de la structuration de dialogues de chat nous ont obligés à revoir la notion de rejouable et l'implémentation de Tatiana (possibilité de d'ordonnancer les événements d'un rejouable autrement que par le temps, possibilité d'enrichissement par graphe, possibilité de création de rejouables à la main à partir de sélections sur une ligne de temps). Les études de cas de Nottingham, des blogs et de Hong Kong ont montré le besoin de création de nouvelles instances de transformation, de visualisation et d'enrichissement, mais n'ont pas remis en cause l'implémentation de Tatiana, ni notre définition de rejouable. Enfin, les études de cas de conception collaborative, de MULCE et de tension/relaxation ont uniquement soulevé des problèmes d'ergonomie.

En l'état actuel de l'implémentation, les problèmes identifiés dans les études de cas et encore non résolus ne demandent plus que la création de plugins et de scripts ou une amélioration de l'ergonomie. Ces problèmes, néanmoins importants, ne remettent pas en cause la notion de rejouable, ni l'implémentation logicielle dans Tatiana. On peut d'ailleurs remarquer que, au fil du temps, les problèmes soulevés appartiennent de plus en plus à cette catégorie de problème. Bien que ce ne soit qu'une tendance sur un petit nombre de cas, ce résultat est très encourageant. A long

terme, nous nous attendons en particulier à ce qu'il subsiste des besoins de créer de nouvelles instances d'opérations sur rejouables – c'est bien pour cela que nous avons voulu créer une plateforme générique et à ce niveau là que nous espérons que le temps nous donnera raison.

6. Discussion et conclusion

Dans cet article, nous avons évoqué nombre des représentations créées par les chercheurs lorsqu'ils analysent des corpus de CSCL ou similaires. A partir de ces exemples, nous avons proposé un modèle simple décrivant l'activité d'analyse. Dans ce modèle, les chercheurs créent itérativement de nouveaux artéfacts, qui leur permettent d'améliorer ou de concrétiser leur compréhension du corpus. Nous avons identifié un type particulier d'artéfact, temporellement orienté, que nous appelons rejouable. Nous avons ensuite présenté Tatiana, un outil qui permet la création itérative de ces rejouables, et nous avons décrit les principales caractéristiques de ce logiciel en termes de création, exportation, transformation, analyse, visualisation et synchronisation de ces rejouables. Enfin, nous avons présenté dix études de cas d'analyses effectuées à l'aide de Tatiana.

Tatiana répond aux besoins, itératifs par essence, des analyses sociocognitives des interactions, en fournissant un ensemble très souple de transformations et visualisations de données, et en offrant de nombreuses possibilités d'extensions pour répondre aux besoins nouveaux, qu'il s'agisse de créer de nouveaux filtres pour des transformations, ou de créer de nouvelles vues pour l'affichage et la modification de rejouables. Ces visualisations pourraient s'apparenter à celles qui sont proposées dans Suthers *et al.* (2009), ou encore Teplovs *et al.* (2007). L'utilisation de visualisations multiples et synchronisées des données, et la possibilité de faire appel à des afficheurs ou rejoueurs extérieurs permettent en partie de pallier les difficultés que peut rencontrer le chercheur qui veut comprendre ce qu'ont pu faire les participants observés. Enfin, grâce à sa possibilité d'enregistrer et de partager des analyses sous forme d'éléments distincts du corpus de base et pouvant être exploités à d'autres artéfacts basés sur le même corpus, Tatiana permet aux chercheurs en sciences humaines et sociales de travailler en équipe afin d'intégrer et de comparer leurs analyses.

Tatiana est actuellement limité aux analyses qui s'effectuent essentiellement par le biais de rejouables. En particulier il n'existe qu'un support limité des agrégations, ces artéfacts qui représentent le résultat d'un calcul et qui ne sont plus ordonnés dans le temps. Ces objets ne peuvent donc être synchronisés, analysés et visualisés de la même façon que les rejouables. Par ailleurs, d'importants problèmes d'ergonomie subsistent et la barrière à la création de nouvelles formes de visualisation, transformation et enrichissement reste élevée.

Au travers des quatre types d'opérations sur les rejouables, Tatiana forme une base génératrice d'un espace de représentations analytiques sur laquelle nous

espérons qu'une variété d'épistémologies pourra se calquer. En créant de nouvelles instances de visualisation, d'enrichissement et de transformation, il est possible de créer de nouvelles formes de représentations analytiques, tout en bénéficiant de manière intégrée des instances des autres opérations. C'est à ce niveau que Tatiana se différencie d'autres outils du même type tels DRS, ABSTRACT, ou ActivityLens (cf. §2.2). Ces outils ne définissant pas un cadre conceptuel dans lesquelles de nouvelles fonctionnalités peuvent être implémentées, chaque nouvelle fonctionnalité augmentera de manière linéaire les possibilités de l'outil. Dans Tatiana, la structuration en types d'opérations démultiplie les possibilités offertes, chaque nouvelle instance d'opération offrant, en combinaison avec les opérations existantes des autres types, un nouvel hyperplan dans l'espace de représentations possibles dans Tatiana.

D'autres modèles pour représenter des traces ont été proposés (Settoui *et al.*, 2009, Choquet & Iksal, 2007) visant notamment à représenter les connaissances contenues dans ces traces de manière exploitable et réutilisable. Nous voyons ces travaux comme étant complémentaires aux nôtres. En effet, si nous reprenons les niveaux (conceptuel, représentationnel et empirique) proposés par Suthers *et al.* (2009), nous nous sommes consacrés au niveau représentationnel tout en choisissant un niveau empirique assez classique (similaire, d'ailleurs aux événements et contingences, mais aussi aux modèles UTL et SBT). Ces deux autres travaux en particulier ont développé la modélisation informatique du niveau empirique afin théoriser l'exploitation des traces. Comparé à ces travaux, le niveau empirique utilisé dans Tatiana ne permet pas une forte description sémantique et ontologique des traces. Ce manque se fera sentir lorsqu'il s'agira d'exploiter automatiquement des données contenues dans des rejouables en dehors de Tatiana. Il serait enrichissant maintenant d'exploiter ces travaux complémentaires, afin de baser l'espace représentationnel défini par les opérations sur les rejouables sur un modèle empirique plus évolué et permettre une plus forte capitalisation sur les analyses effectuées.

Nos travaux futurs ont pour objectifs simultanés de mieux comprendre les méthodologies utilisées par les chercheurs en CSCL/CSCW, tout en développant Tatiana en tant qu'environnement d'étude et de gestion de rejouables et autres artefacts de l'analyse. Nous visons à rendre particulièrement cohérente l'infrastructure du logiciel, et à nous assurer que des chercheurs qui n'ont pas de connaissance particulière en programmation puissent utiliser toute la puissance de cet environnement.

Nous espérons que la diffusion de Tatiana⁷ (<http://code.google.com/p/tatiana>) et d'outils similaires permettra aux chercheurs de tirer un meilleur parti de leurs corpus, tout en partageant ce savoir avec les autres chercheurs. Nous souhaitons également, au travers des usages de Tatiana, acquérir une meilleure connaissance

7. Notons que Tatiana est diffusé gratuitement et sous une licence libre, permettant à chacun de disposer du code source.

des processus d'analyse sociocognitive des interactions, et rendre plus aisés au sein de la communauté du CSCL l'évaluation et le partage des corpus et des analyses.

6. Bibliographie

- Amelsoort, M.A.A., Andriessen, J.E.B., & Kanselaar, G. (2008). How students structure and relate argumentative knowledge. *Computers in human behavior*, 24, 1293-1313.
- Avouris N., Fiotakis G., Kahrimanis G., Margaritis M., Komis V., « Beyond logging of fingertip actions: analysis of collaborative learning using multiple sources of data », *Journal of Interactive Learning Research JILR*. vol. 18(2), 2007, pp. 231-250.
- Baker M., Andriessen J., Lund K., van Amelsoort M., Quignard M., « Rainbow: A framework for analyzing computer-mediated pedagogical debates », *ijcscl* 2 (2-3), 2007.
- Baker, M., Andriessen, J., & Lund, K. (2009). Socio-relational, affective and cognitive dimensions of CSCL interactions: integrating theoretical-methodological perspectives. In *Symposium held at CSCL'09*. Rhodes, Greece.
- Bernard F.-X., « Un modèle d'analyse des interactions médiatisées par les technologies éducatives : le carré médiatique », *Congrès national de la Société Française de Psychologie (SFP)*, 2008, France.
- Carletta J., Evert S., Heid U., Kilgour J., Robertson J., Voormann H., « The NITE XML Toolkit: flexible annotation for multi-modal language data », *Behavior Research Methods, Instruments, and Computers*, special issue on Measuring Behavior, 35(3), 2003, 353-363.
- Cassier, J.-L. (en préparation). *Modélisation de l'Activité de Conception Collaborative Synchronique de Produits Industriels*. Thèse présentée à l'Université de Lyon.
- Cox R., « Technology-enhanced research: educational ICT systems as research instruments », *Technology, Pedagogy and Education*, 16 (3), 2007, 337-356.
- Choquet, C., & Iksal, S. (2007). Modélisation et construction de traces d'utilisation d'une activité d'apprentissage : une approche langage pour la réingénierie d'un EIAH. *Revue Sticef*, 14.
- De Chiara R., Di Matteo A., Manno I., Scarano V., « 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., « Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions », *The Journal of the Learning Sciences*, 11(1), 2002, p. 63-103.

- De Wever B., Schellens T., Valcke M., Van Keer H., « Content analysis schemes to analyse transcripts of online asynchronous discussion groups: A review ». *Computers & Education*, 46(1), 2003, p. 6-28.
- Dyke G., Girardot J.-J., Lund K., & Corbel A., « Analysing Face to Face Computer-mediated Interactions », *EARLI 2007*. Budapest, Hungary.
- Erkens G., Janssen J. « Automatic coding of communication in collaboration protocols », *ijcscl* 3 (4), 2008.
- Gelmini Hornsby G., Ainsworth S., Buda M., Crook C., O'Malley C., « Making your views known: The importance of anonymity before and after classroom debates ». *Proceedings of ICLS 2008*, Utrecht, The Netherlands.
- Georgeon O., Mille A., Bellet B.T., « ABSTRACT: un outil et une méthodologie pour analyser une activité humaine médiée par un artefact technique complexe », *Ingénierie des Connaissances. Semaine de la connaissance*, Rémi Lehn, Mounira Harzallah, Nathalie Aussenac-Gilles, Jean Charlet ed., 2006, Nantes, France.
- Georgeon O., Mathern B., Mille A., Bellet T., Bonnard A., Henning N., Trémaux, J.-M., *ABSTRACT Analysis of Behavior and Situation for menTal Representation Assessment and Cognitive acTivity modelling*, 2007. Retrieved April 10, 2008 from <http://liris.cnrs.fr/abstract/abstract.html>.
- Fiotakis G., Fidas C., Avouris N., « 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., *Applying User Models to Improve Team Decision Making*, 2006, Retrieved April 10, 2008 from http://mitre.org/work/tech_papers/tech_papers_07/06_1351/
- Greenhalgh C., French A., Humble J., Tennent P., « Engineering a replay application based on RDF and OWL », *Online Proceedings of the UK e-Science All Hands Meeting 2007*, September 10-13 2007, Nottingham: NeSC/JISC.
- Harrer A., Zeini S., Kahrimanis G., Avouris N., Marcos J. A., Martinez-Mones A., Meier A., Rummel N., Spada, H., « 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., « 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.
- Law, N., Lu, J., Leng, J., Yuen, J., & Lai, M. (2009). Exploring Knowledge Building Discourse using argumentation markers. In *Workshop on Analysis of computer-supported collaborative learning (CSCL) discourse: advances in theory and analysis tools, CITE Research Symposium 2009*. Hong Kong.

- Lead. *Problem solving through face to face networked interaction in the classroom*, 2006, Retrieved April 10, 2008 from <http://www.lead2learning.org>
- Lomborg, S. (2009). Navigating the blogosphere: towards a genre-based typology of weblogs. *First Monday*, 14(5).
- Lund K., Prudhomme G., Cassier J.-L., « 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.
- Morrison A., Tennent P., Chalmers M., « Coordinated Visualisation of Video and System Log Data ». *Proceedings of 4th International Conference on Coordinated & Multiple Views in Exploratory Visualisation*. London, 2006, 91-102.
- Overdijk M., van Diggelen W., « Appropriation of a shared workspace: Organizing principles and their application », *ijcscl* 3 (2), 2008.
- Prudhomme G., Pourroy F., Lund K., « An empirical study of engineering knowledge dynamics in a design situation », *Journal of Design Research*, 3, 2007, 333-358.
- Reffay C., Chanier T., Noras M., Betbeder M.-L., « Contribution à la structuration de corpus d'apprentissage pour un meilleur partage en recherche », *journal STICEF*, Volume 15, ISSN : 1764-7223, mis en ligne le 14 octobre 2008, <http://sticef.org/>
- Rosé C., Wang Y.-C., Cui Y., Arguello J., Stegmann K., Weinberger A., Fischer F., « Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in CSCL », *ijcscl* 3 (3), 2008.
- Rummel N., Hmelo-Silver C., « Using contrasting cases to relate collaborative processes and outcomes in CSCL », *ICLS 2008*, Utrecht, The Netherlands.
- Sacks H., Schegloff E. A., Jefferson G. « A simplest systematics for the organization of turn-taking for conversation », *Language*, 50, 1974, p. 696-735.
- Settoui, L. S., Prié, Y., Champin, P. A., Marty, J.-C., & Mille, A. (2009). A Trace-Based Systems Framework: Models, Languages and Semantics. Published on HAL-CCSD.
- Strijbos J., Martens R. L., Prins F. J., Jochems W. M., « Content analysis: what are they talking about ? ». *Comput. Educ.* 46, 1, Jan. 2006, p. 29-48.
- Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *ijcscl* 5 (1), pp. 5-42
- Teplovs C., Scardamalia M., « Visualizations for Knowledge Building Assessment ». *AgileViz workshop, CSCL 2007* July 16-July 21 2007, New Brunswick, New Jersey, USA.
- Van Diggelen W., Jansen J., Overdijk, M., « Analyzing and Presenting Interaction Data: A Teacher, Student and Researcher Perspective », *ICLS 2008*, Utrecht, The Netherlands.