
Un outil flexible pour l'analyse de traces

Gestion, synchronisation, visualisation, analyse et partage de corpus d'interactions médiatisées par ordinateur avec Tatiana

Gregory Dyke*, Kristine Lund°, Jean-Jacques Girardot*

*Département RIM, Ecole Nationale Supérieure des Mines de St. Étienne, 158, cours Fauriel F-42023 Saint-Étienne, France

°Laboratoire de Recherche ICAR, CNRS, University of Lyon, 15 parvis René Descartes, 69700 Lyon, France

Email {Dyke, Girardot}@emse.fr

Email Kristine.Lund@univ-lyon2.fr

RÉSUMÉ. L'analyse de données d'interactions médiatisées par ordinateur est difficile. En effet, la création d'un objet d'étude qui fournisse une vision la plus complète, fidèle et pertinente possible de ces données dans l'optique d'une question de recherche posée constitue un véritable défi. Dans cet article, nous présentons Tatiana (Trace Analysis Tool for Interaction Analysts), un environnement destiné à aider les chercheurs à analyser leurs données en créant de manière itérative des artefacts qui expriment la compréhension que ces chercheurs ont acquis de ces données, ou qui s'appuient sur celles-ci pour créer de nouvelles représentations. Nous nous concentrons particulièrement sur les fonctionnalités de Tatiana qui lui permettent d'être étendu en termes d'applicabilité à de nouveaux types de données, en importation, exportation, transformation, formatage, visualisation et analyse.

ABSTRACT. The analysis of multimodal computer-mediated human interaction data is difficult. Indeed, creating an object of study that gives the most complete, faithful and pertinent view possible of observed data for the research questions being addressed is a challenge. In this paper, we present Tatiana (Trace Analysis Tool for Interaction Analysts) as an environment to assist researchers analysing their data by iteratively creating artefacts that exhibit researchers' current understanding of their data or otherwise build upon it. In particular, we describe the features of Tatiana which allow it to be extended for wider applicability in terms of data import and export, transformation and formatting, visualisation and analysis.

MOTS-CLÉS : traces, analyse d'interaction, cscl, cscw, conception, extensibilité

KEYWORDS: process data, interaction analysis, cscl, cscw, design, extendability

1. Introduction

L'étude sociocognitive des interactions humaines supportées par ordinateur peut s'effectuer au travers des enregistrements de ces activités, en particulier si ils ne sont pas limités aux seules traces produites par les outils, mais incluent également des captures audio et vidéo de ces activités (Avouris *et al.* 2007). Cox (2007) encourage les chercheurs à utiliser l'informatique et les diverses techniques qu'elle propose (visualisation, data mining, etc.) pour effectuer leurs analyses au travers de des interactions enregistrées dans les traces d'utilisation. Dans nos travaux précédents, nous avons présenté les difficultés liées à l'analyse de traces d'interactions médiatisées, particulièrement lorsqu'elles sont combinées à des interactions en face à face (Dyke *et al.* 2007, Dyke *et al.*, 2009). Ces difficultés nous ont menés à concevoir le logiciel Tatiana (Trace Analysis Tool for Interaction ANALysts), destiné aux chercheurs qui souhaitent analyser, dans une perspective d'interaction sociocognitive, des situations d'enseignement ou de travail collaboratif médiatisés par ordinateur.

En créant Tatiana, nous avons souhaité fournir au chercheur un véritable environnement d'analyse lui permettant de manipuler, annoter et partager des artefacts analytiques. Nous avons été particulièrement conscients du fait qu'il y a autant de questions de recherche que de chercheurs et que, par conséquent, un tel environnement ne saurait être complet sans proposer des moyens de s'adapter à de nouveaux besoins. Dans cet article, nous présentons notre vision du travail analytique effectué par le chercheur et montrons comment Tatiana peut soutenir ce travail, en focalisant sur les notions de généricité et d'extensibilité.

2. Besoins du chercheur pour l'analyse de traces

Harrer et al. (2007) ont modélisé le travail d'analyse dans l'idée de s'appuyer sur ce modèle pour concevoir des outils d'analyse et des formats de représentation interopérables. Dans ce modèle le processus est composé d'étapes séquentielles telles que la collecte de traces, leur segmentation, l'annotation et l'analyse quantitative et qualitative, pour finir par l'interprétation. Notre point de vue actuel sur l'analyse, basé sur notre expérience personnelle (Lund *et al.*, 2007; de Vries *et al.*, 2002; Prudhomme *et al.* 2007; Baker *et al.* 2007) et celle de nos partenaires dans plusieurs projets, est qu'une telle modélisation du processus d'analyse risque de mener à des logiciels trop contraignants. En particulier, le parcours des étapes n'est pas séquentiel mais itératif, le chercheur effectuant différentes opérations jusqu'à parvenir à un résultat satisfaisant, qui puisse être présenté à la communauté scientifique, en répétant ou effectuant les étapes dans différents ordres, en fonction de ses besoins.

Nous nous sommes appuyés, pour la conception de Tatiana, sur un modèle qui place plus d'emphasis sur la nature itérative de l'analyse (cf. Figure 1). Les analystes

évaluent en permanence si leur collection de données primaires et d'artefacts secondaires est suffisante. Si ce n'est pas le cas, ils créent un nouvel artefact qui leur permet d'aller plus loin dans leur analyse, ou de réifier leur compréhension des données.

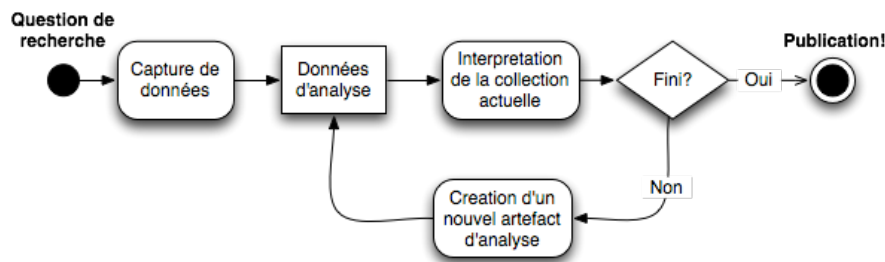


Figure 1. Représentation graphique du processus d'analyse soutenu par Tatiana

Ces artefacts sont souvent des représentations de données faisant intervenir une dimension temporelle : la séquence des événements enregistrés au cours d'une observation est présentée au chercheur soit par un lecteur de média (dans le cas d'une vidéo) soit encore sous la forme d'une représentation graphique (avec un axe temporel représenté horizontalement ou verticalement).

Nous proposons que les objets créés à partir de traces et qui conservent la dimension temporelle de l'ordonnancement des événements et des interactions, soient baptisés *rejouables*. Ils peuvent être rejoués, synchronisés et analysés. Une analyse consiste ainsi en une création itérative de jouables qui témoignent de la compréhension par le chercheur des données qu'il manipule, ou qui autorisent un approfondissement de cette compréhension. Chaque étape de cette itération comporte des transformations telles que : transcription, annotation, codage, visualisation, filtrage, synchronisation, fusion, etc.

3. Description de Tatiana

Nous pouvons ranger les caractéristiques de Tatiana dans quatre grandes catégories : la gestion des traces au travers de l'importation, l'exportation et la transformation de jouables, la synchronisation de jouables multiples, la visualisation de jouables et enfin leur analyse à travers l'annotation et la catégorisation. La création et transformation de jouables sont effectuées au travers de filtres qui s'appuient sur des scripts. Ces jouables peuvent être visualisés à l'aide de différents modules de visualisation, être enrichis par des analyses et être synchronisés entre eux et avec des données gérées par un lecteur externe (vidéo, par exemple). Ces dépendances sont illustrées par la Figure 2.

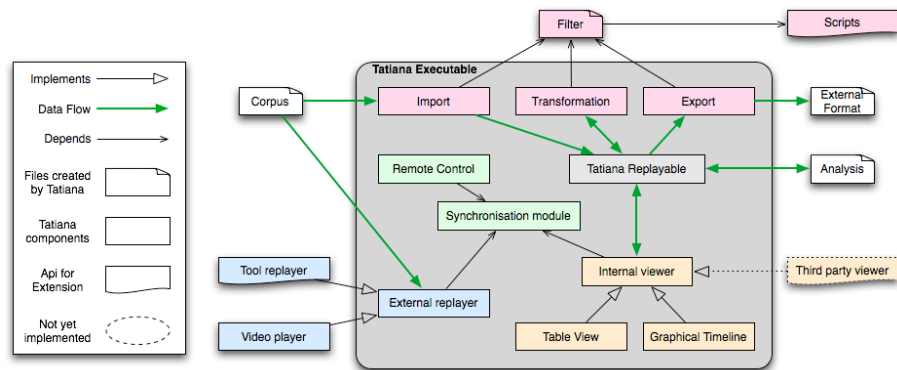


Figure 2. Schéma de l'architecture de Tatiana, montrant les dépendances entre composants, les composants conçus dans une optique d'extension, et les développements futurs.

3.1. Création et manipulation de rejouables

Tout outil d'aide à l'analyse doit manipuler deux types de données, celles qu'il est conçu pour analyser, et celles qu'il produit lui-même. Un premier type de donnée très commun est l'enregistrement multimédia. Les formats audio et vidéo sont relativement standardisés, et peuvent en général être gérés par les bibliothèques logicielles disponibles. En revanche, les outils utilisés dans les activités médiatisées par ordinateur génèrent des traces d'interactions dans des formats qui leur sont propres. Les concepteurs des outils utilisés pour l'analyse des interactions doivent dès lors faire des choix, afin que la représentation interne de leurs données soit à la fois assez complète pour répondre aux divers besoins des analyses, mais également assez souple pour permettre la prise en charge d'un grand nombre de formats de données.

Dans ActivityLens, (Fiotakis *et al.*, 2007) le choix a été fait de n'importer que des données représentées dans le format commun proposé par le groupe de travail CaviCola (Computer-based Analysis and Visualisation of Collaborative Learning Activities) au sein du SIG CSCL du réseau Kaleidoscope (Kahrimanis *et al.*, 2006). Ce format permet à ActivityLens de prendre en charge les notions d'utilisateur, de groupe, d'outil, d'étape, en imposant une certaine vision des données. Il demeure cependant assez générique, permettant l'ajout de nouveaux champs pour en élargir l'usage.

ABSTRACT (Analysis of Behaviour and Situation for menTal Representation Assessment and Cognitive acTivity modelling; Georgeon *et al.*, 2007; Georgeon *et al.*, 2006) décrit les données importées au travers d'ontologies rédigées en

OWL/RDF, et nécessite la programmation d'analyseurs pour permettre l'import des données (ce travail a été poursuivi par Settouti *et al.*, 2006).

Nous avons choisi d'aborder le problème d'une autre manière dans Tatiana, en partant du principe qu'un outil d'analyse doit pouvoir lire n'importe quel format de fichier de données (avec cette réserve que, hormis pour les fichiers audio et video, les fichiers devaient être au format XML et qu'une description de leur structure devait être disponible), et que le coût de l'intégration d'une nouvelle représentation doit être minimal. Nous supposons également qu'il n'est pas nécessaire que l'outil interprète la sémantique des informations contenues dans le fichier (sauf pour les informations de repérage et de synchronisation, c.f. ci-dessous), mais qu'il est suffisant de présenter l'information au chercheur pour que celui-ci puisse lui trouver un sens. Dans cette approche, nous avons choisi une représentation très générique. L'ensemble du corpus est conservé sous sa forme originale, mais Tatiana ajoute une représentation interne qui permet le repérage des événements, et, ultérieurement, leur structuration en graphes. La structure interne des données dans Tatiana constitue un format pivot, et consiste en une séquence d'événements, dont chacun comporte un certain nombre de facettes. Chaque facette associe un nom à une valeur typée. L'importation de données extérieures et la conversion de ces données vers la représentation interne sont effectuées par des scripts écrits en XQuery (W3C, 2007). De manière générale, un script d'importation a pour but de sélectionner un sous-ensemble du fichier à importer, tel que le "chat" au sein d'une trace DREW complète, ou la transcription des interactions (effectuée grâce au logiciel Elan), et à créer les facettes de chaque événement du format pivot au moyen des informations contenues dans le fichier à importer. A partir de ce format pivot, il est possible d'exporter relativement aisément vers d'autres représentations (Excel par exemple).

Afin d'effectuer des transformations automatisées dans Tatiana nous faisons à nouveau appel à des scripts XQuery, et à des filtres (ou *workflows*) pour combiner ces scripts. Ainsi, le script qui extrait la nouvelle position d'une boîte qui a été déplacée dans l'éditeur de graphes d'argumentation du logiciel DREW (Corbel, Girardot, & Jaillon, 2002 ; Corbel *et al.*, 2003), crée également une nouvelle facette pour cet événement, avec un contenu significatif pour le chercheur, tel que « Ann a placé la boîte n°3 à la position 456,324 » (en plus des facettes ayant pour contenu individuels « user : Ann », « boîte : 3 », « outil : DREW », etc.). Un autre script pourrait transformer cette phrase en "Ann a déplacé la boîte à gauche et vers le haut" si cette information était jugée plus pertinente pour le chercheur. Un filtre Tatiana pourrait combiner ces deux opérations, l'une qui extrait l'information, l'autre qui la réécrit, pour créer automatiquement un rejouable ajusté aux besoins du chercheur à partir des traces de DREW. De même, des scripts ont été utilisés pour filtrer les données (p.ex. extraire les informations relatives à un seul groupe d'élèves), créer de nouveaux événements, en regrouper, etc.

A titre d'exemple, dans un corpus relatif à la prise de notes dans un éditeur de texte partagé en réunion, nous souhaitons identifier, à partir des traces, les périodes de rédaction. Du fait de la représentation des interactions de l'éditeur de texte dans

le fichier de trace (l'état de l'éditeur de texte à chaque seconde s'il y a eu modification), cette information n'est pas immédiatement disponible. Pour résoudre le problème, nous créons un filtre qui identifie les périodes d'inactivité dans un outil, et construit un rejouable comportant des événements de début et de fin d'écriture d'un bloc de texte entre ces périodes d'inactivité. La combinaison de ces différents scripts produit un filtre réutilisable (cf. Figure 3) pour identifier les périodes d'activité et d'inactivité dans l'éditeur partagé. Le script lui-même peut dorénavant être utilisé sur d'autres rejouables pour détecter des périodes d'activité et d'inactivité. Nous développons actuellement une interface graphique de création et d'édition de filtres, permettant au chercheur de combiner des scripts prédéfinis, afin de créer les rejouables dont il a besoin pour ses travaux.

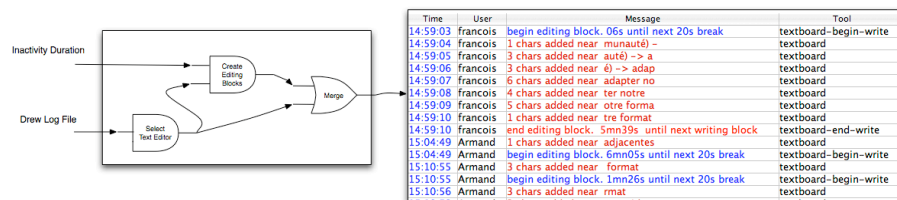


Figure 3. Un filtre Tatiana qui combine trois scripts, prenant comme paramètres un fichier de trace DREW et la durée d'une période d'inactivité, et produisant un rejouable enrichi de deux nouveaux types d'événements, textboard-begin-write, textboard-end-write, en plus des événements textboard. Le rejouable est ici affiché sous forme tabulaire.

DREW étant doté d'un rejoueur (un mode du logiciel qui relit les traces qu'il a produites et réaffiche à l'écran ce qui s'est passé), nous avons pu profiter des transformations de traces dans Tatiana pour prendre les traces de l'éditeur de graphe d'un autre logiciel (non doté de rejoueur), les importer dans le format pivot et les exporter vers le format de DREW afin de les visionner dans son rejoueur. Cet exemple confirme la généricité du format pivot de Tatiana ainsi que les possibilités variées offertes par les transformations automatisées.

3.2. Visualisation de rejouables

Il y a plusieurs raisons d'enregistrer les données d'une session pour une analyse ultérieure. D'abord, trop d'événements se produisent simultanément pour que l'analyse en temps réel soit possible (bien que la prise de notes par un observateur apporte souvent un éclairage qui autrement aurait été perdu). Surtout, le processus d'analyse nécessite que les données constituent à la fois un artefact de travail (p.ex. un document ou une vidéo annotables) et une preuve à laquelle on puisse se référer, afin que les hypothèses puissent être vérifiées séparément. De fait, rendre les corpus plus simples à partager et à analyser de manière indépendante dans les domaines du

CSCL et du CSCW relève d'une nécessité scientifique (Reffay *et al.*, 2008). La contrepartie de cette approche est que nous disposons dès lors d'un ensemble de documents, qui sont non seulement impressionnants du simple fait du volume de leurs données, mais aussi dont il est difficile à rétablir le lien avec ce qui s'est passé durant l'observation, en particulier dans le cas des traces brutes d'interaction (les vidéos posent d'autres problèmes, liés par exemple aux angles de prises de vue, à la qualité du son, etc.). Le premier réflexe du chercheur, lorsqu'il est confronté au problème, est de créer des rejouables qui éliminent nombre de données, réduisant ainsi la complexité de l'information et simplifiant l'établissement de liens avec les événements observés. L'un des dangers de cette approche est le risque d'aboutir à une fausse interprétation de ces données du fait de la perte de leur contexte. Le défi consiste à créer un ensemble de rejouables synchronisés, tel que chacun ne reflète qu'une portion réduite de l'information disponible, mais qui tous ensemble brossent un tableau global permettant au chercheur de disposer à tout instant du contexte. Nous allons présenter les différentes formes de visualisation des rejouables, et de quelle manière ces rejouables peuvent être construits, transformés et formatés avant leur affichage.

L'un des outils les plus simples de CSCL est le chat. Les traces qu'il produit peuvent aisément s'afficher sous une forme tabulaire, avec une ligne par événement, et des colonnes pour les différentes données de celui-ci : date, utilisateur, intervention. De fait, des outils comme ExcelTM sont souvent utilisés par les chercheurs pour visualiser ce type d'interaction.

Une autre forme de visualisation est la représentation graphique ponctuelle des événements sur une ligne horizontale. Ce mode de présentation se retrouve, par exemple, dans ABSTRACT (c.f. Figure 4). Le reproche principal que l'on peut adresser à cette représentation, qui fournit une information appréciable, est sa faible adaptabilité. Lorsque cette fonction est implémentée par un logiciel, il manque presque toujours la possibilité de modifier l'affichage, et en particulier la représentation graphique des événements (sauf intervention dans le code source de l'application).

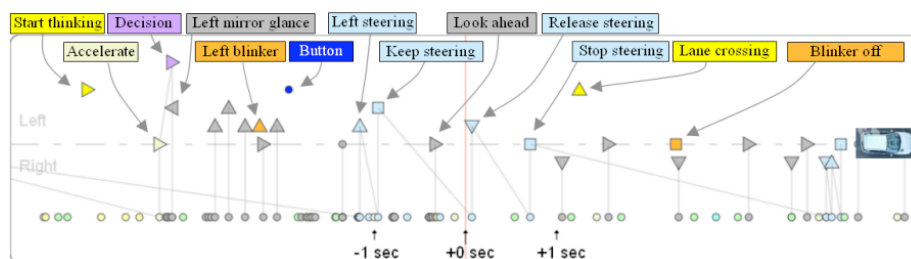


Figure 4. Représentation graphique symbolique dans ABSTRACT des événements enregistrés lors d'un changement de file sur une autoroute (<http://liris.cnrs.fr/abstract/>)

Finalement les traces informatiques peuvent être visionnées de manière très efficace pour la compréhension dans un rejoueur, tel que celui implémenté par DREW. Les traces multimédia peuvent bien sûr être lues par des lecteurs audio-vidéo.

Tatiana fournit deux types d’affichage des données représentées dans son format pivot : une vision tabulaire classique (avec une ligne pour chaque événement, une colonne pour chaque type d’information que l’on choisit d’afficher), et une vue graphique symbolique, avec un axe des temps horizontal (cf. figure 5). L’affichage tabulaire est très largement paramétrable au moyen d’une description indiquant pour chaque colonne les informations à afficher (de quelles facettes elles sont extraites) et sous quelle forme (formatage et présentation). Ceci permet d’éviter l’affichage de colonnes qui n’intéressent pas le chercheur, de regrouper entre elles des colonnes presque toujours vides, etc. Ainsi, une même colonne peut présenter les identifications des objets graphiques (boîtes, liens) bien que ces informations apparaissent sous des noms différents dans des facettes différentes.

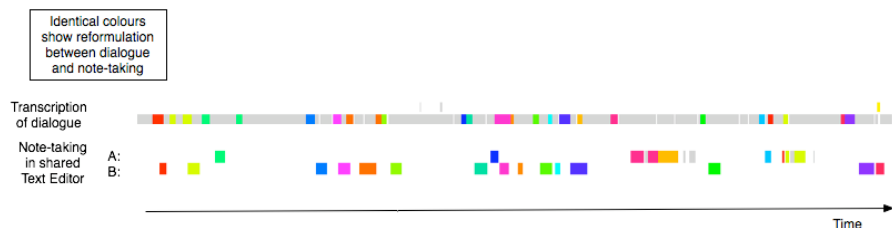


Figure 5. Visualisation dans Tatiana de la reformulation entre un dialogue et des notes prises dans un éditeur de texte partagé.

Dans la vue graphique, différentes propriétés (forme, taille, couleur, position verticale) sont associées à chaque événement. La largeur et la position sur l’axe horizontal sont calculées à partir de l’estampillage temporel et de la durée de l’événement. Les autres paramètres sont calculés d’après diverses facettes, en suivant des règles définies par le chercheur au sein de Tatiana. Ainsi, la Figure 5 est une représentation de la reformulation effectuée par des étudiants prenant des notes d’un éditeur de texte partagé du discours que leur tenait leur tuteur, lors d’une réunion autour d’un projet de programmation. Il nous semble important que le chercheur puisse, au travers d’essais successifs, choisir la représentation des rejouables (ici, leur visualisation graphique) la plus appropriée, tout en conservant la synchronisation de ces rejouables avec le reste du corpus (Lund, et al. 2008). Il est également possible d’afficher sous forme de graphe les liens que le chercheur a établi entre les événements.

3.3. Synchronisation de rejouables

Les outils dont s'inspire Tatiana (Greenhalgh *et al.*, 2007 ; Kipp *et al.*, 2001 ; Morrison *et al.*, 2006 ; Georgeon *et al.*, 2007 ; Fiotakis *et al.* 2007) permettent la synchronisation entre différents types de rejouables (cf. Figure 6), que ce soient de simples fichiers vidéo, des transcriptions ou des traces de divers types. Dans la section précédente, nous avons montré l'utilité et la nécessité de synchroniser différentes vues (p. ex. vidéo et transcription) du même corpus afin de resituer un rejouable dans son contexte. Tatiana permet de synchroniser entre eux les rejouables dans son format pivot qui comporte, pour chaque événement, une facette temporelle obligatoire. Les scripts d'importation tiennent compte de cette contrainte, ce qui permet, par exemple, de fusionner sans problème une transcription et des données issues de la trace d'interaction d'un outil.

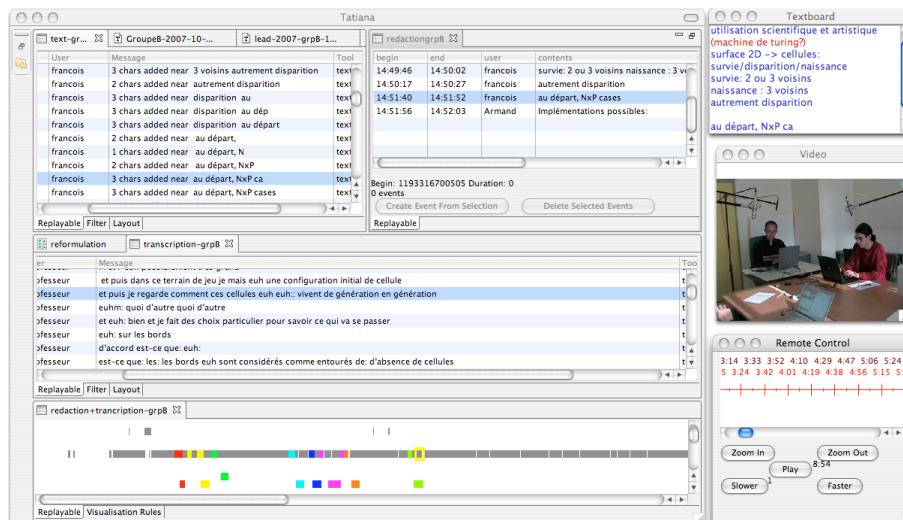


Figure 6. Visualisation synchronisée de divers rejouables sous formes tabulaire et graphique avec un rejoueur externe et une vidéo.

De plus, Tatiana permet la synchronisation avec des rejoueurs externes (tel VLC pour la vidéo ou le rejoueur de DREW) via une interface XML-RPC. Tatiana utilise également la stratégie “brush and link” (Becker & Cleveland, 1987) entre les divers rejouables qui sont visualisés à un moment donné : ainsi, lorsque le chercheur clique sur un élément dans un rejouable, cet élément est mis en évidence dans tous les autres rejouables correspondants et les éventuelles vidéos et rejoueurs sont également amenés à cet instant dans le temps.

3.4. Ajout d'informations analytiques sur des rejouables

Une partie du processus d'analyse consiste en la création de ce que nous appelons des *analyses*. Ce sont tous les objets qui enrichissent des rejouables existants par des informations créées par le chercheur (sens supplémentaire, abstraction, etc.) pour décrire ce qui s'est passé durant l'observation. Nous avons ainsi identifié trois larges aspects de l'activité d'analyse : l'annotation, les regroupements et segmentations ainsi que la création de liens.

Bien que les pratiques de catégorisation soient distinctes de celles des annotations, elles peuvent s'implémenter au travers de la même technologie : les annotations peuvent être restreintes à un ensemble (ouvert ou fermé) de mots clefs, ce qui, par essence, permet la catégorisation. Pour annoter des segments, il faut les créer au préalable. Une fois ces deux fonctionnalités mises en place, d'autres peuvent suivre. Par exemple, la transcription équivaut (sur le plan technologique) à une segmentation suivie d'une annotation. Il est également possible d'utiliser la segmentation pour effectuer des regroupements, permettant ainsi de former des rejouables composés d'événements de plus haut niveau que ceux présent dans les traces initiales.

L'une des activités d'analyse qu'aucun des outils que nous avons étudiés ne semble offrir est la description de relations entre événements. Dans nos études de cas sur des analystes, nous avons plusieurs fois pu observer des créations de représentations graphiques illustrant les liens entre les différents événements (par ex. Prudhomme *et al.*, 2007).

Tatiana offre ces trois possibilités au travers d'un mécanisme que nous avons baptisé *ancre* et *lien*. Une ancre est une désignation spécifique dans un corpus (un "point du corpus"). Un lien regroupe des ancres et/ou des annotations. Un fichier d'analyse décrit chaque action d'analyse comme une série de liens entre différentes ancres (Corbel *et al.*, 2006). Ainsi :

— Une annotation est un lien associé avec un texte, tandis qu'une catégorisation est un lien entre deux ancres (l'une désigne une catégorie, l'autre un point du corpus). Par exemple, l'événement de création d'une boîte dans le grapheur pourrait être annoté "C'est la seconde fois que l'utilisateur crée cette boîte". L'ancre du lien désignerait l'événement correspondant du fichier trace (très précisément, l'élément XML décrivant cet événement). De même, l'application de la catégorie "critère d'affirmation" à la création d'une flèche dans le grapheur d'argumentation aurait pour effet de créer un lien avec une ancre sur l'événement du fichier trace, et une autre vers le fichier XML contenant l'énumération des catégories.

— La segmentation ou le regroupement d'événements est une création de lien : le regroupement de cinq événements dans l'éditeur de texte (au motif qu'ils représentent une "unité d'édition") consiste à créer un lien regroupant les cinq ancres sur ces événements.

— La création d'une relation entre deux événements est également une création de lien — ainsi, affirmer que l'envoi d'un message dans le chat est une réponse à la création d'une boîte dans le grapheur consiste à créer un lien associant les ancres de ces deux événements ; l'identification de ce lien comme représentant une relation de "réponse" se fait en ajoutant une annotation ou une ancre vers le fichier décrivant les catégories de liens.

Grâce à ces informations (une ancre désignant de manière non ambiguë un point du corpus, tout comme un événement dans le format pivot), un fichier d'analyse peut être fusionné à un rejouable de la même façon qu'une nouvelle couche de données peut être superposée à une carte topographique. Le résultat de cette opération est d'ajouter des facettes supplémentaires aux événements annotés, facettes qui peuvent à leur tour être visualisées.

Les représentations dans le format pivot sont susceptibles d'évoluer au cours de l'analyse (l'application de filtres permet d'éliminer, de modifier, de rajouter des facettes ou des événements), mais il est important qu'une liaison existe entre un élément d'un tel fichier et le point du corpus qui est à l'origine de cet élément. Pour ce faire, les scripts d'importation associent à chaque événement une facette "ancree source" qui fait référence au point du corpus d'où l'événement a été extrait. Différents scripts peuvent importer (en créant des facettes différentes) le même événement, mais l'ancree source sera identique, et toutes les transformations (filtrage, fusion, etc) conservent cette ancre source. En d'autres termes, celle-ci constitue une "référence universelle" sur le point du corpus.

4. Extensibilité et généricité

En dotant Tatiana de ces fonctions de base, nous espérons les avoir rendues suffisamment génériques pour satisfaire de nombreux besoins analytiques, dont certains que nous ignorons encore. En effet, l'architecture d'importation et de transformation permet d'étudier facilement des traces provenant de nouveaux logiciels, de créer de nouveaux scripts de transformation ainsi que de recombinaison de scripts dans de nouveaux filtres. Les visualisations, particulièrement en mode graphique, sont aussi d'une grande richesse, permettant facilement au chercheur de modifier les règles d'affichage pour voir apparaître les informations sous un autre angle. Finalement, la possibilité d'enregistrer séparément corpus, rejouables et analyses permet leur partage avec d'autres chercheurs. Nous sommes particulièrement conscients du besoin des analystes de s'affranchir de leur dépendance sur les informaticiens. Pour cela, nous travaillons activement sur les interfaces des règles de visualisations ainsi que sur celles de recombinaison de scripts en filtres.

Tatiana a été développé sur la plateforme Eclipse (www.eclipse.org), afin de tirer parti de l'architecture des plugins Eclipse (Bolour, 2003). Cela permettra à terme à des informaticiens de créer de nouvelles formes de visualisation et d'analyse qui

pourront profiter de la synchronisation et des transformations existant déjà dans Tatiana. De plus, Tatiana fournit une API¹ pour les rejoueurs externes. Un tel rejoueur doit, après son lancement, créer un serveur de XML-RPC qui accepte les requêtes d'ouverture de fichier, de récupération des événements dont le rejoueur souhaite être notifié et de navigation à la date spécifiée. Avec ce protocole de communication, Tatiana est capable de piloter un rejoueur externe et de le synchroniser avec les autres rejouables, en envoyant des requêtes de type "Goto" lorsque nécessaire (en se basant sur la liste des événements reconnus par le rejoueur).

En fournissant ces trois possibilités d'extensions (i.e. les scripts, l'interface avec des rejoueurs externes, et les plug-ins de visualisation, cf. Figure 2), nous espérons élargir la variété de questions de recherches auxquelles Tatiana est susceptible de s'appliquer, tout en gardant une infrastructure de développement raisonnable, qui permette aux chercheurs de résoudre nombre de questions de recherche au prix d'un investissement minimal.

5. Conclusions et développements futurs

Dans ce document, nous avons présenté la démarche de chercheurs désirant analyser des traces d'interactions et introduit la notion de *rejouable* : un artefact créé à des fins analytiques à partir de traces et qui conserve la notion d'ordonnancement dans le temps. Nous avons évoqué les diverses difficultés liées à la création et la manipulation de ces rejouables et montré comment Tatiana permet de pallier ces difficultés. Nous avons décrit les principales caractéristiques de Tatiana, en termes de création, de transformation, de synchronisation, de visualisation et d'analyse de ces rejouables et montré comment ce logiciel peut s'adapter à des situations pour lesquelles il n'a pas été conçu.

Tatiana répond aux besoins intrinsèquement itératifs et divers d'analyses relatives au CSCL et CSCW, en fournissant des méthodes très flexibles de transformation et d'affichage de données, en particulier grâce à des passerelles permettant d'étendre le logiciel pour répondre à de nouveaux besoins. La difficulté de faire revivre, à partir des données enregistrées, ce que les participants éprouvaient lors du recueil du corpus trouve des réponses au travers des visualisations multiples de données synchronisées, en particulier grâce à l'utilisation de rejoueurs externes. Enfin, du fait de la possibilité de sauvegarder et de partager des analyses, Tatiana permet aux chercheurs en sciences humaines et sociales de travailler en équipe, et d'intégrer et de comparer leurs analyses.

Nos travaux futurs devraient porter d'une part sur l'amélioration de notre compréhension des besoins et des méthodologies d'analyse des chercheurs, d'autre part sur le développement de Tatiana en tant qu'environnement de gestion de

¹ Application Programming Interface

rejouables (et des autres artefacts des analyses que nous n'avons pas présentés dans cet article). Nous espérons ainsi leur offrir une meilleure approche de leur corpus, tout en leur permettant de partager cette création de sens avec d'autres collègues. Nous souhaitons également aboutir à une meilleure compréhension du processus d'analyse des interactions sociocognitives, ce qui rendrait plus facile l'évaluation et le partage des méthodologies au sein de la communauté du CSCL.

Références

- Avouris N., Fiotakis G., Kahrmanis 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).
- Becker R. A., & Cleveland W. S. (1987). Brushing scatterplots. *Technometrics*, 29:127-142.
- Bolour A. (2003). Notes on the Eclipse Plug-in Architecture. Retrieved April 10, 2008 from http://www.eclipse.org/articles/Article-Plug-in-architecture/plugin_architecture.html
- 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.
- Corbel, A, Girardot, J.J., & Lund, K. (2006). A method for capitalizing upon and synthesizing analyses of human interactions, In (eds) W. van Diggelen & V. Scarano, Workshop proceedings Exploring the potentials of networked-computing support for face-to-face collaborative learning. *EC-TEL 2006 First European Conference on Technology Enhanced Learning*, Crete, pp. 38-4
- 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.
- Dyke, G., Girardot, J.-J., Lund, K., & Corbel, A. (2007). Analysing Face to Face Computer-mediated Interactions. *EARLI '07*. Budapest, Hungary.
- Dyke, G., Girardot, J.-J., Lund, K., (2009). Tatiana: an environment to support the CSCL analysis process. *CSCL 2009*, Rhodes, Greece.
- 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. *Ingénierie*

- des Connaissances. Semaine de la connaissance.* Rémi Lehn, Mounira Harzallah, Nathalie Aussenac-Gilles, Jean Charlet ed. Nantes, France.
- Georgeson, 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
- 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.
- Kipp, M. (2001) Anvil - A Generic Annotation Tool for Multimodal Dialogue. Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech), pp. 1367-1370.
- 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 16-21, New Brunswick, New Jersey, USA.
- 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, Andriessen, Prudhomme, van Amelsvoort (in preparation). The varying influences of collective debating tasks on individual knowledge construction.
- 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.
- 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, *journal STICEF*, Volume 15, ISSN : 1764-7223. Consulté le 14 octobre 2008 <http://sticef.org>.
- Settouti, L.-S., Prié, Y., Mille, A., Marty, J.-C. (2006) Système à base de traces pour l'apprentissage humain. [Trace-based system for human learning] *Colloque international*

TICE 2006 «Technologies de l'Information et de la Communication dans l'Enseignement Supérieur et l'Entreprise».

W3C (2007). *XQuery 1.0: An XML Query Language*, Retrieved April 10, 2008 from <http://www.w3.org/TR/xquery/>