

ViCoDiLi, un outil de Visualisation de Contenus de Discussions en Ligne

Philippe Teutsch, Francis Bangou, Charlotte Dejean-Thircuir

Lium, Université du Maine, Le Mans

Résumé : Les corpus de données issues de formations en ligne (FEL) intéressent les chercheurs dans leur démarche d'observation et de compréhension des interactions humaines provoquées par les dispositifs d'activités médiatisées. Cependant, les plateformes de FEL ne sont pas conçues pour que les données (notamment de type échanges) issues des formations soient conservées à long terme et puissent ainsi constituer un matériel d'étude. Le projet Odil a rapproché plusieurs équipes de recherche qui se sont donné pour objectif de faire en sorte de préserver des données issues de formations en ligne et d'échanger ces données afin d'en faire des analyses croisées. Deux conditions semblent essentielles pour faciliter l'analyse de corpus de FEL : pouvoir accéder aux contenus des échanges indépendamment de la plateforme hôte et disposer de fonctionnalités de visualisation de ces contenus.

Cet article porte sur la double question de l'accès et de la manipulation des interactions ayant eu lieu dans le cadre d'une formation en ligne. Il interroge les techniques et outils utilisés dans le cadre de l'Analyse de Données Qualitatives Assistée par Ordinateur (ADQAO). Cet article présente également ViCoDiLi, un outil de Visualisation de Contenus de Discussions en Ligne qui permet aux chercheurs d'accéder aux interactions d'une session de formation et de les examiner selon différentes modalités d'affichage. ViCoDiLi s'appuie sur un modèle de description des échanges qui permet d'harmoniser leur représentation à la fois indépendamment de la plateforme et indépendamment de l'outil de communication utilisé.

Mots-clé : échange de corpus, interaction en ligne, visualisation, analyse

Introduction

Les environnements numériques utilisés pour la mise en œuvre de dispositifs de formation ouverte et à distance (FOAD) gèrent un grand nombre de données produites par l'activité des participants. Les chercheurs peuvent ainsi accéder aux contributions issues des activités d'apprentissage, ou encore aux messages échangés à travers les forums de discussion. Ces corpus de données conservés sur les ordinateurs intéressent les chercheurs dans leur démarche d'observation et de compréhension des situations d'enseignement-apprentissage provoquées par les dispositifs médiatisés. L'analyse des traces d'activité permet par exemple d'améliorer l'appropriation des dispositifs par les utilisateurs (Settouti et al., 2007), d'étudier l'usage des dispositifs dans une démarche de ré-ingénierie (Choquet et al, 2007), de les interroger dans le cadre d'une analyse didactique (Dejean-Thircuir & Mangenot, 2006) ou encore de les caractériser dans le cadre d'une formation de formateur.

L'étude des discussions et des échanges verbaux qui se créent entre les acteurs de ces environnements médiatisés de formation constitue un champ de recherche particulièrement porteur du point de vue de l'analyse des interactions en ligne (Épal, 2007). Ce domaine recouvre plusieurs types de recherches (Bruillard, 2006 ; Henri et al., 2007) : l'étude des communications médiatisées par ordinateur (CMO) en tant que telles et des dynamiques sociales qui s'y développent (Preece & Maloney-Krichmar, 2003 ; Herring, 2004 ; Reffay & Lancieri, 2006), l'étude des interactions en ligne comme contribution aux apprentissages (Peraya, 2005 ; Mangenot, 2006), et l'étude des pratiques en relation avec les outils qui les instrumentent (George & Bothorel, 2006). Les principales disciplines concernées par ces recherches sont les sciences du langage, les didactiques, la sociologie, les sciences de l'information et de la communication et l'informatique.

Quelle que soit la situation de formation, l'apprentissage s'effectue toujours dans un contexte riche, multiple, voire complexe. Pour analyser des phénomènes liés à cet apprentissage en situation, il est nécessaire de travailler avec un ou plusieurs corpus issus de dispositifs de formation réels, avec de vrais enjeux et responsabilités pour les participants (Bruillard, 2006). La nature écologique des situations de formation ayant permis de constituer les corpus est caractéristique des recherches sur les interactions en ligne. Elle en traduit même l'enjeu : l'analyse des interactions n'est possible que si les corpus issus de formations complètes sont existants, préservés et diffusés (Reffay et al., 2007).

Toutes ces recherches sont guidées par l'attrait que représentent les technologies de communication comme solution aux problèmes d'isolement de la formation à distance et comme moyen de favoriser la réflexion et la structuration des échanges (Henri & Lundgren-Cayrol, 2001). Néanmoins, (Henri et al, 2007) remarquent que certains résultats d'expérimentation de FEL sont plutôt mitigés, voire décevants, quant à l'appropriation de ces environnements par les usagers. Pour expliquer ces disparités dans les résultats d'analyse, les auteurs soulignent que les recherches concernées sont menées dans des contextes très variés, peu comparables, avec des méthodes d'investigations et d'analyse également variées. Pour pouvoir partager des résultats issus d'analyses de corpus d'interactions, il semble utile de rendre ces corpus accessibles à d'autres chercheurs que ceux qui les ont créés ainsi que de proposer des instruments méthodologiques spécifiques à ce travail d'analyse et susceptibles d'être appliqués à différents corpus afin d'étendre la validité des résultats obtenus.

Cette proposition a été retenue dans le cadre du projet Odil (Odil, 2004) qui s'est intéressé aux questions de mutualisation de résultats de recherches portant sur des corpus issus de formation en ligne hétérogènes. Les recherches ont principalement porté sur l'analyse des interactions¹ produites entre acteurs d'une formation en ligne. La démarche de diffusion, de mutualisation et d'interopérabilisation fait partie des recommandations méthodologiques de la communauté (Henri & Charlier, 2005) et concerne autant les corpus eux-mêmes que les outils d'analyse et les résultats obtenus (Bruillard, 2006). Ces recommandations ont été prises en compte au sein du projet Odil par la mise en place de deux actions de recherche parallèles : le projet Mulce (Mulce, 2006) piloté par le LIFC et cherchant à définir un protocole de diffusion de corpus et le projet ViCoDiLi (ViCoDiLi, 2006) piloté par le Lium et cherchant à élaborer un outil de Visualisation de Contenus de Discussions en Ligne qui permette aux chercheurs d'accéder aux échanges d'une session de formation et de les examiner selon différentes modalités d'affichage et de manipulation.

Cet article présente les travaux réalisés dans le cadre du projet ViCoDiLi. Il aborde la question de l'instrumentation des activités d'analyse de corpus d'interactions verbales issues de dispositifs médiatisés de formation. L'article présente successivement le contexte général de la recherche, le cadrage théorique lié à l'analyse des interactions en ligne et l'outil Vicodili. Il s'achève par une discussion sur les choix épistémologiques qui guident, explicitement ou non, la conception et la mise en œuvre des outils d'assistance à l'analyse qualitative.

1 Contexte de recherche

Chercher à diffuser des corpus d'apprentissage en vue de leur mutualisation nécessite de définir les données concernées de départ, le processus de traitement à appliquer à ces données pour obtenir un ensemble structuré, ainsi que les modalités d'accès au corpus ainsi défini. Concernant ce dernier point, nous défendons l'idée qu'une relative harmonisation des modalités de structuration, de visualisation et de manipulation de ces données peut faciliter le travail d'analyse sur les corpus. L'intérêt a priori est, par exemple, de pouvoir dérouler un même protocole d'analyse sur plusieurs corpus, ou de comparer des corpus issus d'un même scénario pédagogique et joué dans des contextes différents : participants, calendrier ou environnement technologique par exemple.

Cette partie précise l'origine, la problématique, les objectifs et la méthodologie du projet de recherche mené autour des questions d'accès aux données d'un corpus d'apprentissage.

1.1 Origine du projet

La question de la réutilisation des corpus s'est posée dans le cadre du projet Odil à l'occasion de la "restauration" d'un des corpus du projet. Les différents corpus concernés par le projet Odil étaient tous dépendants de l'environnement informatique support à la formation les ayant produits. Pour consulter un corpus et effectuer leur tâche d'analyse, les chercheurs utilisaient leur identifiant d'enseignant et les modalités de consultation de la plateforme, ce qui suppose que la session de formation concernée par l'analyse soit toujours accessible et que la plateforme soit toujours opérationnelle. Un de ces corpus est devenu complètement inaccessible lorsque l'organisme de formation a arrêté l'exploitation de la plateforme hôte. Il s'agit du corpus Simuligne, issu d'une formation à distance pour l'apprentissage du FLE (Français Langue Étrangère) mise en œuvre en 2001 sur la plateforme WebCT dans le cadre du projet ICOGAD (Reffay et al. 2002). Le scénario invitait les apprenants à une production collaborative dans le cadre d'une activité de simulation globale. Le corpus concerne 40 adultes anglophones en formation continue, 10 étudiants natifs francophones et 4 tuteurs (1 par groupe). La formation s'est déroulée sur 10 semaines et a produit plus de 12 000 interventions à travers les outils de forum, courriel et clavardage.

Lorsque l'organisme responsable de la formation n'a pas renouvelé la licence d'exploitation de la plateforme hôte, les chercheurs disposaient d'une sauvegarde numérique de l'ensemble des contributions produites par la formation Simuligne mais n'avaient aucun moyen technique

de consulter ou de manipuler ces données. L'enjeu technique et scientifique lié à la restauration du corpus Simuligne était de permettre aux propriétaires d'un ensemble de données gérées par une plateforme "disparue" de retrouver ces données sous forme de corpus structuré et homogène présentant les échanges de façon assez proche de leur structure de communication d'origine.

Différents entretiens ont été menés auprès des chercheurs impliqués dans le projet Odil afin de les interroger sur la question de la réutilisation des traces d'activités produites et mémorisées sur les espaces numériques de travail. Ces entretiens ont abouti à plusieurs constats : un manque général d'accessibilité aux données, un défaut de pérennité des corpus et un manque de traces normalisées.

1.1.1 Accessibilité aux données

Les plateformes de FEL ne sont pas conçues pour que les données (notamment de type échanges) soient conservées à long terme et puissent ainsi constituer un matériel d'étude. Les droits d'accès sont limités aux participants à la formation et sur une période de temps définie pour la session de formation.

Même si ces données peuvent être exploitées en tant que corpus pendant un temps par les chercheurs qui ont eux-mêmes mis en place des formations ou des modules d'enseignement en ligne, les données risquent donc de disparaître, comme dans le cas de Simuligne. Leur accès devient dépendant des techniciens de l'informatique, des possibilités d'extraction normalisées des plateformes, des formats de structuration interne des données (bases de données interrogeables ou exportables) et de la possibilité de retrouver une interface de consultation de ces données.

1.1.2 Pérennité des données

Les traces d'activités et les contributions produites par une session de formation peuvent intéresser les chercheurs sur une durée relativement longue, ne serait-ce que pour pouvoir comparer plusieurs sessions d'un même dispositif de formation, y compris des sessions jouées dans des environnements logiciels différents. Par conséquent, il y a nécessité de garantir une assez grande pérennité aux corpus d'interactions en ligne mis à disposition des chercheurs.

Il ressort également des entretiens que, pour certaines recherches, il est nécessaire de pouvoir rejouer, ou du moins revoir, toute une session de formation dans ses conditions réelles de déroulement : environnement technique, navigation et interfaces de consultation. Dans ce cas,

la disponibilité de l'environnement support est indispensable. Pour d'autres recherches par contre, il peut être intéressant, voire nécessaire, de réduire la dépendance des contenus à analyser vis à vis de l'instrument qui les a produit. On peut alors considérer le corpus d'interactions comme un ensemble de données qui peuvent être extraites de leur environnement de production afin d'être mise à disposition des chercheurs. C'est dans cette deuxième perspective qu'a été conçu l'outil ViCoDiLi.

1.1.3 Normalisation des données

De nombreuses recherches s'intéressent à l'analyse de traces d'activité issues de dispositifs médiatisés de formation. Pour les travaux les plus proches de l'analyse de corpus, signalons notamment l'analyse d'usage en vue de l'amélioration de la conception des scénarios pédagogiques (Barré et al., 2005 ; Pernin, 2006). Ces travaux préconisent l'utilisation et l'extension de la norme IMS-LD (IMS-LD, 2003) pour rendre les dispositifs de formation interopérables et plus facilement analysables. IMS-LD semble adapté à la description du déroulement du scénario effectif pour une tâche individuelle mais semble plus difficile à utiliser pour décrire finement chacune des interventions qui peuvent être produites lors d'un échange en ligne. La démarche de structuration des corpus (Noras et al., 2007) participe à ces recherches sur la normalisation des traces d'interactions issues d'activités d'apprentissage en ligne. Pour l'instant, les plateformes actuellement utilisées dans les situations concrètes de formation en ligne ne respectent pas les normes de description de formation et bloquent tout espoir d'accès simplifié à ces données.

En bilan de cette section de présentation des origines du projet, on peut retenir que les plateformes de FEL ne sont pas prévues pour conserver les données à long terme, ni pour les diffuser à d'autres chercheurs que ceux qui ont participé à la formation. De plus il existe peu de moyens de garantir un accès normalisé et pérenne à ces données. Ainsi, la mise à disposition de corpus doit pouvoir répondre aux attentes des analystes en termes d'accessibilité, de visualisation et de manipulation des données. Cette problématique est présentée dans la section suivante.

1.2 Problématique et objectifs

Nous considérons que les résultats d'analyses portant sur des corpus hétérogènes peuvent être comparés si les chercheurs disposent de données, de méthodes, et d'outils de traitement également comparables, voire identiques. L'objectif de recherche est alors de montrer qu'il est possible et utile de permettre un accès homogène à des corpus hétérogènes.

Chercher à faciliter l'accès aux échanges en ligne nécessite de définir la notion même d'accessibilité. Une première question est de savoir décrire ces échanges et leurs contextes de production afin de permettre leur accès et leur interprétation en termes d'interactions. Une seconde question est de savoir décrire les modalités de visualisation et de manipulation de ces échanges en tenant compte du contexte d'usage du chercheur analyste. Ces deux problématiques de l'interaction sont certainement interdépendantes mais peuvent être abordées séparément.

La première question envisage l'interaction (humaine, médiatisée et pédagogique) comme objet d'étude. Les objets d'analyse sont des interactions produites dans un espace à plusieurs dimensions : dispositif de formation, situations personnelles d'apprentissage, environnement de travail médiatisé. La participation aux activités médiatisées constitue "*une activité située, finalisée et cadrée par un contexte*" (Henri & Charlier, 2005). La connaissance de ce contexte et des différentes dimensions concernées est nécessaire à la compréhension des interactions. La variété des situations d'apprentissage et des outils qui les instrumentent est à prendre en considération, par exemple le caractère collaboratif, ou non, des activités proposées, ou encore le caractère synchrone, ou non, de ces interactions. Un premier objectif est de savoir proposer un modèle de description des composants d'un corpus d'interactions qui permette de prendre en compte cette variété de scénarios et d'environnements qui les médiatisent.

La seconde question envisage l'interaction (entre usager et système informatique) comme une situation à construire pour permettre l'analyse des corpus. Comme dans tous les domaines, la tâche d'analyse est contextualisée : par les objectifs d'analyse, par les habitudes et l'expertise de l'analyste, ainsi que par l'environnement de présentation du corpus lui-même. Nous sommes ici dans une situation d'ingénierie et dans une démarche d'instrumentation. Un second objectif est de savoir proposer des outils et services d'aide à l'analyse des corpus qui tiennent compte de la variété des situations et des besoins d'analyse.

Pouvoir gérer les contenus d'interactions en ligne indépendamment de leur structure et de leur plateforme d'origine se décline donc en deux objectifs principaux : définir une structure

générique des interactions en ligne et définir une interface et des modalités de manipulation de ces interactions en relation avec les pratiques des analystes.

1.3 Démarche de conception

Le double objectif de réaliser un environnement informatique supportant l'activité de l'analyste tout en élaborant le modèle des données à analyser pose en soi une difficulté méthodologique. Notre proposition d'interface d'exploration homogène pour différents corpus issus de différentes plateformes risque de modifier le point de vue des analystes sur la notion même de corpus. Cette situation d'évolution technologique est caractéristique des recherches en Interaction Humain-Machine (IHM) qui, nécessairement pluridisciplinaires, cherchent à augmenter la performance du couple système-utilisateur dans la cadre de la conception d'environnements interactifs.

Le concept le plus important en IHM est probablement celui de conception centrée sur l'utilisateur (Norman, 1988), qui prend en compte la perspective de l'utilisateur dans le processus de conception de la nouvelle technologie. Dans ce cadre, la triangulation entre disciplines permet de multiplier les perspectives et d'avoir une représentation la plus large possible de l'activité considérée pour la création d'un artefact (Mackay & Fayard, 1997). L'utilisateur étant au centre du processus de conception, Rabardel (1995) définit la co-adaptation comme un phénomène croisé d'instrumentation de l'utilisateur par l'artefact proposé et d'instrumentalisation de ce même artefact à travers la mise en place de schèmes d'utilisation que l'utilisateur développe dans la situation d'usage réel.

La démarche participative (Mackay & Fayard, 1997) consiste alors à intégrer les utilisateurs (analystes) au processus de conception et d'évaluation du système en devenir. Le principe méthodologique est de leur permettre d'exprimer leurs besoins et leurs points de vue, et d'explicitier leur pratique. Les auteurs expliquent que les démarches de recherche prennent en compte l'introduction de l'artefact dans la situation d'usage. La conception d'un artefact se nourrit (Figure 1), d'une part d'une approche théorique vérifiée par expérimentation, et d'autre part d'une observation de terrain se structurant en modèle théorique. Le principe est de compléter l'analyse de la tâche a priori par l'analyse des fonctions pragmatiques et des besoins des acteurs qui effectuent la tâche sur le terrain. Chaque cycle permet de faire évoluer le prototype, d'affiner le modèle tout en créant de nouveaux usages. Les trois champs d'étude concernés sont donc le modèle théorique, l'artefact informatique et l'usage réel. Cette

démarche de conception itérative et participative permet d'identifier les besoins essentiels et de définir les outils correspondants ainsi que les fonctionnalités sous-jacentes, sans pour autant disposer au départ d'un modèle précis de l'activité de l'analyste.

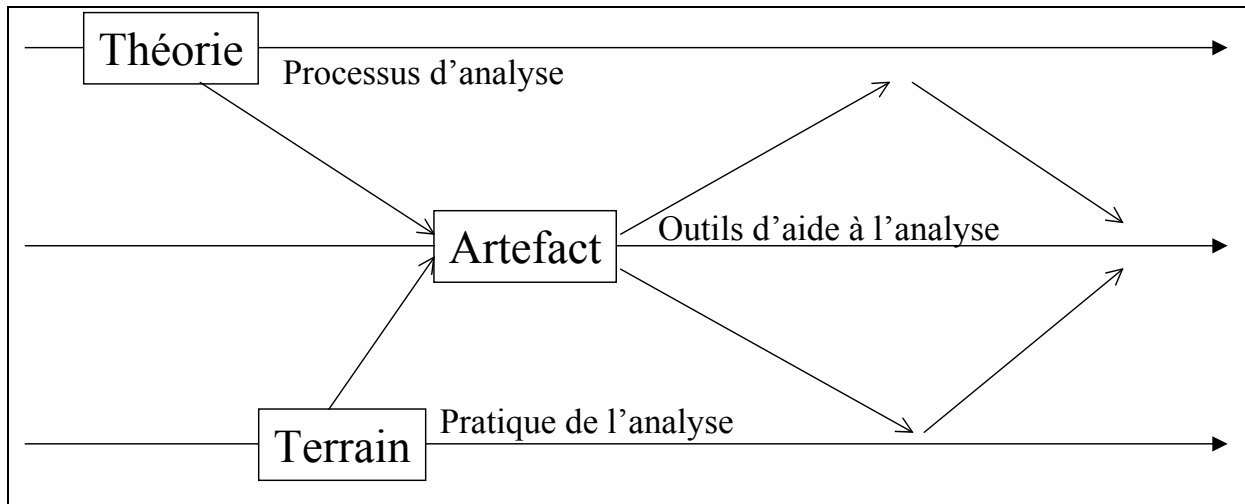


Figure 1 - Principe d'instrumentation appliqué à notre problématique.

Étant donné que la mise en œuvre d'un logiciel crée un nouveau milieu et de nouveaux usages, il est nécessaire d'intégrer cette dimension dès sa conception. Toute la difficulté du concepteur est de créer les conditions permettant le rapprochement de deux points de vue autour d'un même objet d'étude : l'artefact informatique dédié à la situation d'analyse. Afin de tenir compte des tâches et des démarches des analystes, l'élaboration de l'outil ViCoDiLi a suivi une méthodologie de conception itérative et participative (un informaticien et deux chercheurs en didactique analysant des interactions en ligne). Les aspects théoriques de l'analyse des interactions en ligne sont présentés dans la prochaine partie. Les aspects pratiques, le modèle de corpus et les interfaces de l'outil ViCoDiLi le seront dans la partie suivante.

2 Cadrage théorique

L'analyse des interactions médiatisées s'inscrit dans la démarche plus générale d'analyse de données qualitatives. Cette partie propose un rapide état de l'art sur l'analyse de données qualitatives et les outils qui l'instrumentent puis tente de caractériser les corpus et les pratiques liés à l'analyse des interactions en ligne.

2.1 Caractéristiques de l'analyse qualitative

Selon Tesch (1990), cité par (Savoie-Zajc, 2000), l'analyse de données qualitatives est “*un processus qui implique un effort explicite d'identifier des thèmes, de construire des hypothèses (idées) telles qu'elles émergent des données ainsi que de clarifier le lien entre les données, les thèmes et les hypothèses conséquentes*” (p. 113). L'analyse de données qualitatives suit donc une logique inductive nécessitant une démarche de pensée riche, complexe et intuitive.

La tâche centrale du chercheur en analyse qualitative revient à construire une structure, à manipuler des extraits de corpus, à les comparer, à les associer dans le but de leur donner du sens. On constate par ailleurs qu'il existe une proximité entre le chercheur et ses données. Les corpus sont généralement, mais pas exclusivement, constitués de relevés d'observations ou de transcriptions d'entrevues effectués par le chercheur lui-même. De ce fait, la construction des données est fortement guidée par le chercheur qui observe ou guide l'entretien (Tesch, 1990).

2.2 Pratiques de l'analyse qualitative

Comme de nombreux domaines d'activité humaine, l'analyse de données qualitatives correspond à une pratique qui peut être assistée par des logiciels spécifiques.

Selon Tesch, le processus d'analyse de données qualitatives comprend deux phases : premièrement une organisation des données, qui implique une segmentation et entraîne une décontextualisation, deuxièmement une interprétation, un regroupement et une thématization qui correspond à une recontextualisation. Savoie-Zajc (2000) témoigne de sa pratique d'analyse de données qualitatives sans l'aide d'outils informatiques en la décomposant en quatre étapes. La première étape consiste à lire plusieurs fois les documents afin de bien saisir leur message. Puis, elle reprend des épisodes d'observation afin d'identifier des thèmes. Au cours de la troisième étape, elle regroupe des thèmes apparentés. “*La tâche de l'analyste consiste alors à photocopier des extraits d'entrevues, à les découper, à les coller sur des fiches afin de rassembler ceux qui expriment une similitude thématique, à élaborer un système de codes (couleurs, signes divers) pour créer des repères visuels*” (p. 102). Enfin, la dernière étape consiste à créer et à organiser des catégories. Pour le chercheur, ces tâches de manipulation et d'organisation des données sans l'aide d'outils informatiques demandent de la rigueur et sont fastidieuses (p. 103). Il lui est difficile de rassembler l'ensemble des données dans un même espace (visuel) afin d'en faciliter les comparaisons et de clarifier le

sens qui s'en dégage. Il nous semble donc important de permettre au chercheur de disposer des vues variables sur un corpus donné. Par exemple le chercheur pourrait avoir accès à une vue d'ensemble du corpus. Il pourrait aussi disposer de vues intermédiaires montrant les différents niveaux de construction des interactions (e.g. scénario, activités, échanges), jusqu'à la vue détaillée de chaque intervention, dans son contexte de production.

Dans le cadre de l'Analyse de Données Qualitatives Assistée par Ordinateur (ADQAO), de nombreux logiciels ont été élaborés afin de faciliter l'analyse de telles données. Les plus connus sont : The Ethnographer, Atlas/TI, NUD*IST, et Xsight. Les logiciels d'assistance à l'analyse qualitative facilitent et structurent la tâche d'analyse en regroupant dans un même espace des documents textes variés et volumineux. Grâce à ces logiciels, l'analyste peut passer aisément d'un document à un autre afin d'effectuer des opérations de décontextualisation et de recontextualisation. Ces logiciels permettent aussi de gérer des annotations et de garder un lien constant entre chaque extrait codifié et son origine dans le corpus transcrit. De ce fait, la décontextualisation n'est plus incompatible avec la préservation des liens avec le corpus d'origine. Grâce à ces logiciels, la démarche d'analyse n'est pas imposée en soit : on peut débiter l'analyse avec une construction théorique pré-définie et un système d'index pleinement construit, ou, au contraire, débiter l'analyse en pleine induction en créant des codes au fur et à mesure qu'ils sont inspirés par la lecture des documents (Bourdon, 2000). Cependant, les analyses de données qualitatives utilisant des logiciels d'assistance portent essentiellement sur des corpus issus d'observations de terrain ou d'entrevues effectuées par le chercheur. Les textes analysés ont une structure linéaire et sont généralement issus d'interactions synchrones. Ces logiciels sont très peu adaptés à l'analyse de textes non linéaires provenant d'interactions asynchrones.

Des logiciels ont été aussi créés pour représenter visuellement les interactions en ligne (Heer & Boyd, 2005 ; Gibbs et al., 2006). Par exemple, le logiciel PeopleGarden représente visuellement les échanges qui s'effectuent dans un forum, le logiciel Babble permet de visualiser les activités des participants de clavardage. Bien que ces outils permettent de représenter graphiquement des échanges asynchrones et synchrones en dehors du contexte de la formation, ils ne permettent pas d'effectuer des opérations de décontextualisation et de recontextualisation.

Il nous semble donc important de proposer un outil de visualisation qui permettrait d'effectuer des opérations de décontextualisation et de recontextualisation sur des textes linéaires et non linéaires provenant d'interactions synchrones et asynchrones.

2.3 Analyse des interactions en ligne

Dans le cadre particulier de l'analyse des interactions en ligne, il n'existe pas de bibliographie décrivant précisément les caractéristiques des corpus analysés, de la tâche et des besoins de l'analyste. Cependant, l'étude approfondie de plusieurs publications et communications récentes dans ce domaine (Sidir et al., 2005 ; Bruillard, 2006 ; Dejean-Thircuir & Mangenot, 2006a) permettent d'émettre quelques suppositions sur la nature des corpus et sur les pratiques d'analyse de ces corpus.

Les corpus intéressant les analystes sont principalement constitués des traces d'activité, i.e. des échanges en ligne et des productions. Par ailleurs, les recherches sur les interactions en contexte éducatif tentent de rendre compte de phénomènes complexes qui nécessitent de croiser plusieurs éclairages complémentaires : *“rendre compte de la dynamique interactionnelle dans la conversation médiée par ordinateur dans ses relations avec les processus sociocognitifs mis en jeu par les interactants”* (Pudelko et al., 2006). De ce fait, la connaissance du contexte est essentielle à l'activité d'analyse.

Les méthodologies de recherche et les techniques de transcription, d'annotation, de codage et d'analyse sont discutées dans la communauté scientifique (Rourke et al., 2001), par exemple l'identification des unités d'analyse (fil de discussion, message, partie du message, ...). La question de la construction d'instruments valides semble donc préoccuper les chercheurs intéressés à développer des méthodes éprouvées pour l'analyse du contenu des interactions médiatisées (Bruillard, 2006).

2.3.1 Définition de corpus d'interaction en ligne

Les interactions en ligne se font généralement par écrit, à l'aide d'outils informatiques, et ils peuvent être synchrones (clavardage) ou asynchrones (forum, courriel). Il existe par ailleurs une disparité entre ces corpus, disparité liée aux dispositifs et aux plateformes utilisés (Mangenot, 2007). Les types d'outils utilisés. la structuration des données (base de données propriétaire, standardisée, format d'échange normalisé de type XML) et les interfaces de présentation et de manipulation de ces données (affichage, création, reprise du message auquel on répond, indentations) sont souvent complexes et hétérogènes.

Les porteurs du projet Mulce situent ces traces d'interactions comme une des composantes d'un corpus d'apprentissage défini plus globalement comme *“un ensemble de données et de traces issues d'une expérimentation, enrichies par des informations techniques, humaines, pédagogiques et scientifiques permettant leur analyse en contexte.”* (Noras et al., 2007). Les formations candidates à la diffusion de leurs corpus d'interactions en ligne ne bénéficient pas toujours de contexte de recherche a priori. Nous réduisons en conséquence la définition d'un corpus d'interactions en ligne à l'intégration des données suivantes : dispositif de formation, session de formation, contributions à la formation.

- Le dispositif de formation est décrit en termes de composants de la formation : objectifs, scénario pédagogique, séquences d'activités, consignes, ressources et matériels pédagogiques, outils de production et de communication, type et forme d'accompagnement (Collectif, 2001). Cette description est relativement générique puisqu'un dispositif a pour vocation à être mis en œuvre plusieurs fois.
- La session de formation précise le contexte de mise en œuvre de la session concernée : participants, identités, profils, trajets de formation (Teutsch et al., 2004).
- Les contributions désignent les différentes productions des participants aux activités (individuelles ou collectives, synchrones ou asynchrones, produits des tâches programmées dans le dispositif) et interactions entre ces participants (produits des activités de communication proposées par le dispositif) (Laperrousaz, 2006).

Ces contributions aux activités constituent pour nous le cœur des données à analyser. Les descriptions du dispositif et de la session de formation constituent quant à elles le contexte des interactions étudiées.

2.3.2 Critères de qualité des corpus

L'accessibilité à un corpus peut s'évaluer à travers les critères de qualité proposés par le projet Mulce (Reffay, 2007) : pérennité, lisibilité, homogénéité et validité. Nous illustrons ces critères à travers les caractéristiques des corpus disponibles au sein du projet Odil.

- Lisibilité : les données sont-elles lisibles à travers (au moins) un moyen de visualisation, dépendant ou non du dispositif d'origine ? Concernant le corpus Simuligne, l'accès aux données étant dépendant de la plateforme hôte, les acteurs de la formation eux-mêmes se sont vus dans l'impossibilité de retrouver les traces de leur activité.

- Pérennité : les données sont-elles accessibles en dehors de la plateforme d'origine, pour une durée de vie indépendante du temps de la formation concernée ? Concernant les corpus du projet Odil, de nombreuses données sont dispersées, incomplètes. Leur accès est limité aux participants à la formation, et donc interdit aux chercheurs extérieurs.
- Homogénéité : la structure des corpus permet-elle une manipulation homogène des données pour l'indexation ou la recherche ? Dans le cadre du projet Odil, les données "brutes" sont toutes dépendantes du système d'information d'origine. Les structures de bases de données et les formats d'échange, quand ils existent, sont hétérogènes.
- Validité (écologique, durée, volume) : le corpus est-il issu d'une formation réellement "écologique", i.e. non expérimentale, en termes de conception de dispositif pédagogique, de durée de la formation et de nombre de participants ?

2.3.3 Caractéristiques de l'analyse des interactions en ligne

La mise en perspectives des activités d'analyse des interactions dans le cadre plus général de l'analyse de données qualitatives montre plusieurs points de convergence et quelques différences.

Le premier point de convergence concerne les opérations de décontextualisation et de contextualisation. Par exemple, dans leur étude sur les interactions entre pairs, Dejean-Thircuir & Mangenot (2006, p. 81) ont regroupé sous un même thème "*Usage d'un registre familier*" des extraits provenant de diverses interactions :

"Waouh! C'est génial, tu nous as envoyé pleins de message et tu as carrément bien avancé dans les activités ..." [Julie]

"Tout baigne Judith !" [Cristina]

Nous pouvons supposer que pour construire ce regroupement de citations, les auteurs ont : 1° parcouru le corpus, 2° segmenté le corpus, 3° codifié le corpus, 4° regroupé les extraits codifiés 5° et thématisé les groupes ainsi constitué.

Nous pouvons aussi remarquer que les chercheurs sont souvent amenés à codifier des informations concernant le contexte des interventions (auteur, fil de discussion, thème, date&heure, rang). Par exemple, des informations de ce type accompagnent souvent l'extrait d'une intervention en ligne : "*17/03/2004 C1, salon rouge, entre Vasco, Portugais, et Liliana Argentine en espagnol [E179C]*" (Degache, 2006 : 63).

Nous pouvons donc dire que la démarche générale d'analyse d'interactions en ligne est comparable à celle de l'analyse de corpus textuels. Dans les deux cas, les chercheurs sont conduits à effectuer des opérations de décontextualisation et de recontextualisation. Cependant, comparativement aux corpus textuels, linéaires et homogènes, les corpus d'interactions en ligne sont complexes de part leur variété de structures et disparates de part l'hétérogénéité de leurs emplacements.

3 Spécifications de l'outil ViCoDiLi

Notre réflexion sur l'accessibilité de corpus en vue de leur analyse a été guidée par la nécessité de restituer l'accès au corpus Simuligne. Ce "défi" a été relevé à travers la conception de ViCoDiLi, outil destiné à la Visualisation de Contenus de Discussions en Ligne. Cette partie présente les caractéristiques générales de l'outil ainsi que le format de description et les interfaces de manipulation des échanges.

Les échanges manipulables par ViCoDiLi sont des échanges textuels, structurés par des outils de communication de type forum, clavardage, courriel ou blog. La première version de ViCoDiLi fonctionne sur le corpus Simuligne issu d'une plateforme WebCT ainsi que sur des corpus issus de la plateforme Moodle. L'outil a pour vocation à être étendu à d'autres corpus issus de différentes plateformes de FEL. Les travaux visent également à diversifier les analyses pouvant être facilitées par l'outil. Cette intention nécessite de définir une structuration générique des échanges à traiter.

3.1 Format de description des interactions

L'outil ViCoDiLi s'appuie sur un format unique de description des corpus d'échanges en ligne qui permet d'harmoniser leur représentation à la fois indépendamment de la plateforme et indépendamment de l'outil de communication utilisé. L'objectif est de garantir une exploitation homogène de ces corpus à travers les interfaces de visualisation et de manipulation.

L'unité de base retenue par les chercheurs analystes impliqués dans la conception de l'outil est l'intervention, soit le "message", au sens informatique du terme, posté dans l'outil de communication concerné : intervention en forum, intervention en clavardage, courrier électronique,... Chaque intervention est définie par son auteur, sa date de dépôt, le corps du texte et éventuellement l'intervention "mère" dont elle dépend dans le cas d'échanges

structurés. Les interventions composant un corpus d'origine subissent une transformation (automatique ou manuelle si nécessaire) permettant de respecter un format de description commun à l'ensemble des corpus. Ce format de référence s'appuie sur une structure arborescente qui supporte un nombre variable de niveaux intermédiaires entre le niveau supérieur de corpus et le niveau terminal de l'intervention. Techniquement, chacun de ces niveaux de description a un rôle équivalent. Le principe est de décrire récursivement l'ensemble du corpus à travers ce modèle et de disposer d'un fichier d'échange normalisé (format XML) interprétable par l'interface de visualisation (Figure 2). Par exemple, un corpus de forum extrait de Simuligne est progressivement décrit à travers les trois niveaux intermédiaires suivants : groupe de participants, activité de discussion prévue par le scénario pédagogique, et fil de discussion produit par l'activité. Par contre, deux niveaux intermédiaires suffisent à décrire un corpus de clavardage extrait de Moodle : thématique de discussion et session d'interactions synchrones. Ce modèle générique permet d'unifier l'accès à des interventions de différents types produites par différents scénarios sur différentes plateformes de FEL.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<groupes>
<groupe nom="Gallia">
  <forum nom="Prem : 02-29/04/01">
    <discussion nom="Salutations">
      <message date="3 4 2001" heure="23h10" idmess="i3" jour="Mardi" mess_pere="inull" nom_auteur="Sabrina GLISTON">Salut tout le monde. J'espere que tout va bien marcher, on ne sait jamais avec les ordinateurs!! Je serais heureuse de vous tous connaitre et de commencer un dialogue avec vous. Amitiés, Sabrina
    </message>
    <message date="4 4 2001" heure="15h56" idmess="i5" jour="Mercredi" mess_pere="i3" nom_auteur="Sylvie TESSIER">Bonjour Sabrina ! Nous vous souhaitons la bienvenue sur WebCT ! Nous sommes heureux de voir que vous êtes entrée dans votre groupe sans trop de difficultés. Nous attendons à présent la venue de vos partenaires. Cordialement. Sylvie.
    </message>
    <message date="5 4 2001" heure="22h9" idmess="i7" jour="Jeudi" mess_pere="inull" nom_auteur="Tony JENKINS">Salut tout le monde! je fais annonce de mon arrivee sur WebCT. J'espere que l'experience sera tres utile pour tous et toutes et j'attends avec impatience le plaisir de rencontrer les autres membres de la groupe. Amitiés Tony JENKINS
    </message>
  </discussion>
  ...
```

Figure 2 - Extrait de fichier de structuration informatique des interventions

Si la plateforme permet d'exporter des fichiers respectant un format d'échange à balises de type XML, l'adaptation au format ViCoDiLi est simple et automatisable. Pour le corpus Simuligne, par contre, le fichier de description XML a été entièrement reconstruit à partir de la nomenclature des répertoires et fichiers WebCT fournie par les porteurs du projet ICOGAD (Reffay et al. 2002).

3.2 Interfaces de visualisation et de manipulation

La méthode de conception participative utilisée pour le projet a permis de définir les modalités d'interaction avec les données en partenariat avec les analystes impliqués dans le projet. Cette section présente les spécifications fonctionnelles définies pour ViCoDiLi. Les exemples d'interfaces sont fournis à partir du corpus anonymisé (Reffay & Teutsch, 2007) de Simuligne.

ViCoDiLi répond aux besoins de visualisation et de manipulation des interactions à différents niveaux de granularité : par corpus, par groupes d'échanges, par intervention. L'utilisateur dispose de moyens simples et rapides de parcours et d'interrogation de l'ensemble des contenus. Il peut également composer lui-même des extraits de corpus spécifiques.

Principes	Spécifications
Unité d'analyse élémentaire : l'intervention	Permettre une manipulation homogène des différents types d'intervention (Forum, courriel ou clavardage pour Simuligne)
Retrouver les structures d'origine	Afficher les interventions dans leur contexte de production (fil de discussion sur forum, séquence de courriels)
Naviguer rapidement du niveau le plus général au niveau le plus détaillé	Organiser l'interface en zones de navigation cohérentes avec la structure du corpus. Organiser la présentation du corpus en vues d'ensemble, vues intermédiaires et vues détaillées
Décontextualisation	Permettre de segmenter le corpus
Recontextualisation	Permettre de regrouper et de thématiser les extraits segmentés

Tableau 1 - Synthèse des principes issus des ateliers de conception de ViCoDiLi

Les principales demandes des utilisateurs ont porté sur les conditions d'affichage des échanges et sur les conditions de navigation dans les extraits de corpus. Les conditions d'affichage sont restées proches de celles existantes dans les environnements de discussion en ligne : structuration arborescente des messages dans les forums par exemple. Ce principe correspond à la nécessité de fournir avec le nouveau logiciel les mêmes modalités de présentation que dans la version "habituelle". Les conditions de navigation permettent de

passer rapidement d'une vue d'ensemble des interactions à une vue détaillée de chaque intervention. En cas de recherche, chaque intervention est présentée dans le contexte de l'activité d'apprentissage l'ayant produite. Enfin, les principes de décontextualisation et de recontextualisation sont mis en œuvre. Les spécifications issues de ces principes d'interaction entre le corpus et son utilisateur sont résumées dans le tableau ci-dessus (Tableau 1).

Les moyens de visualisation fournis par ViCoDiLi permettent de naviguer rapidement dans les différents niveaux de représentation, d'afficher le contenu d'un échange et de déployer à la demande les interventions observées (Figure 3). Les fonctionnalités d'assistance à l'analyse fournies par l'outil concernent la recherche par auteur ou par mot clé, la présentation des occurrences trouvées dans leur contexte de production (échange et intervention) et le regroupement des interventions extraites du corpus dans le "panier" (Figure 4). Cette sélection d'interventions peut être manipulée et triée par l'analyste avant d'être extraite de ViCoDiLi sous forme de fichier texte, pour effectuer un traitement complémentaire, ou sous forme de fichier diffusable .pdf.

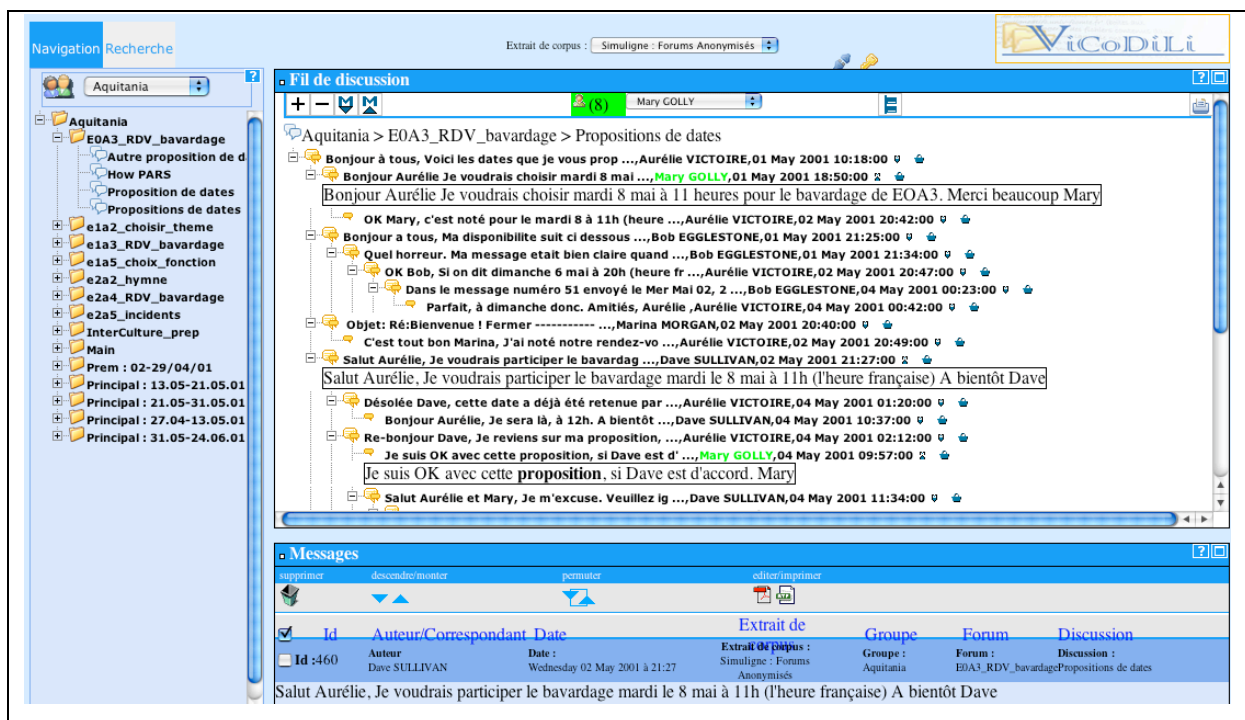


Figure 3 - Navigation dans un corpus

Messages						
supprimer		descendre/monter		permuter		editer/imprimer
<input checked="" type="checkbox"/>	Id	Auteur/Correspondant	Date	Extrait de	Groupe	Forum
<input type="checkbox"/>	Id :460	Auteur Dave SULLIVAN	Date : Wednesday 02 May 2001 à 21:27	Extrait de corpus : Simuligne : Forums Anonymisés	Groupe : Aquitania	Forum : E0A3_RDV_bavardagePropositions de dates
Salut Aurélie, Je voudrais participer le bavardage mardi le 8 mai à 11h (l'heure française) A bientôt Dave						
<input type="checkbox"/>	Id :467	Auteur Mary GOLLY	Date : Friday 04 May 2001 à 09:57	Extrait de corpus : Simuligne : Forums Anonymisés	Groupe : Aquitania	Forum : E0A3_RDV_bavardagePropositions de dates
Je suis OK avec cette proposition, si Dave est d'accord. Mary						

Figure 4 - Regroupement d'énoncés d'origines diverses dans le "panier"

3.3 Bilan ViCoDiLi

Le projet ViCoDiLi a tout d'abord permis de restaurer entièrement le corpus Simuligne où une cinquantaine de personnes ont échangé plus de 12 000 interventions. Ce corpus est donc à nouveau visible, exploitable et analysable tout en étant indépendant de sa plateforme d'origine.

La structuration des interventions et leur intégration à l'outil ViCoDiLi ont ensuite été appliquées à d'autres corpus de différentes tailles et issus de plateformes variées, en particulier deux corpus d'échanges (quelques centaines de messages sur forums dédiés à des activités d'échanges de points de vue et de réflexion) produits lors d'une formation de master en didactique des langues et extraits de la plateforme Moodle via un fichier de format XML.

Ces différentes expérimentations d'intégration de corpus permettent de valider le principe de distinction entre contenus d'interactions (interventions, échanges) et supports d'interaction (plateforme d'origine) et d'ainsi garantir une indépendance entre ces deux éléments de la formation en ligne. Les données peuvent ainsi être accessibles, visualisées et manipulées, indépendamment des environnements les ayant produites. Une fois intégrés dans l'outil, ces corpus vérifient trois des critères de qualité préconisés par la communauté de recherche sur les interactions en ligne : pérennité, lisibilité, homogénéité. La validité reste dépendante de la formation étudiée.

L'outil ViCoDiLi a cependant besoin d'être enrichi de nouvelles fonctionnalités avant d'être expérimenté en réelle situation d'analyse. Les retours sur l'interface ont souligné la faiblesse des modalités d'expression des recherches et l'impossibilité de gestion des extraits regroupés dans le panier comme un corpus en soi, qui bénéficierait à son tour des modalités de décontextualisation. La méthode de conception entamée en partenariat avec des praticiens doit

donc prolonger ses itérations pour faciliter l'évolution de l'artefact vers un instrument réellement modelé par ses utilisateurs avec leur propres schèmes d'utilisation (Rabardel, 1995). Au-delà de la vérification de l'utilisabilité du logiciel, les questions de l'utilité et de l'acceptabilité (Tricot et al., 2003) pourront alors se poser.

4 Discussion

La définition de critères de qualité d'une recherche a toujours fait l'objet de discussions au sein de la communauté scientifique (Henri et al., 2007). Plus précisément, la question de la construction d'instruments valides semble préoccuper les chercheurs intéressés à développer des méthodes éprouvées pour l'analyse du contenu des interactions médiatisées (Bruillard, 2006). Ainsi, on a pu constater au sein du groupe Odil une grande proximité entre les enseignants porteurs de formation en ligne et les chercheurs analystes des situations provoquées par ces formations. Le travail d'analyse se fait alors sur des données parfaitement connues, explicitement ou non, par les chercheurs comme c'est souvent le cas pour l'analyse de données qualitatives (Savoie-Zajc, 2000). Henri et al. (2007) ont tenté de problématiser les recherches sur les forums de discussion dans un cadre éducatif grâce à une méta-analyse de différentes études. Les auteurs proposent de différencier les travaux dans ce domaine selon quatre perspectives de recherche : les recherches 'sur' les forums comme dispositif médiatique, les recherches 'sur' les forums comme lieu d'interaction verbales, les recherches 'avec' le forum comme lieu d'apprentissage et les recherches 'pour' le forum" comme instrument. Cette typologie est liée à quatre perspectives de recherche : la communication médiatique, l'approche interactionnelle, les perspectives pédagogiques, et l'approche instrumentale. Selon eux, "*une claire explicitation de l'objet de recherche, et du cadre qui le définit aiderait sûrement les chercheurs à mieux positionner leurs travaux dans le champ*" (p. 27). De fait, les critères de qualité d'une recherche ne peuvent être universels. Au contraire, ils varient en fonction de l'orientation épistémologique du chercheur. C'est pour cette raison que tout comme Henri et al. (2007) nous pensons qu'il est important pour le chercheur d'explicitier ces choix concernant ses objectifs, ses méthodes, et son approche épistémologique.

Dans cette perspective, les outils d'assistance à l'analyse tels que ViCoDiLi devraient encourager une diversité épistémologique et méthodologique en analyse qualitative. Cependant, certains chercheurs comme Savoie-Zajc (2000) ont souligné que la structure de certains logiciels "influence la logique du chercheur" (p. 116) et cite le cas du logiciel

NU*Dist qui a été conçu selon la logique de la théorie enracinée (“grounded theory“ de Glaser & Strauss, 1967) et qui a comme unité d'analyse minimale le mot. Selon l’auteur, la structure arborescente et hiérarchique du logiciel "amène le chercheur a conceptualiser ses données dans une logique qui va du général au spécifique" (p. 115). Un tel logiciel serait difficile à utiliser si un chercheur considère le paragraphe comme son unité minimale d'analyse ou s’il suit une logique qui va du spécifique au général. Nous constatons que les concepteurs d'un tel logiciel ont fait des choix épistémologiques et méthodologiques qui peuvent influencer la pratique d'analyse du chercheur. Dans le cadre de l'élaboration de l'outil ViCoDiLi, notre intention était de concevoir un outil générique d’assistance à l’analyse, mais de fait, nous avons également fait des choix ayant pour conséquence de particulariser les approches méthodologiques induites par l’utilisation de l’outil, notamment en permettant aux chercheurs d'avoir accès a des vues d'ensemble, intermédiaires et détaillées des échanges en ligne ou en choisissant l'intervention en tant qu'unité d'analyse minimale. Positionner ainsi nos travaux dans le champ de l'ingénierie d'outils d'assistance à l'analyse des échanges en ligne permet aux chercheurs de choisir l'outil qui leur convient le mieux.

Conclusion

Cet article a présenté les différentes facettes d’une recherche en ingénierie d'outils d'assistance à l'analyse des interactions en ligne. Le projet ViCoDiLi a permis de restaurer le corpus Simuligne, de proposer un prototype qui facilite et harmonise l’accès à des corpus d’interactions issus de contextes techniques et pédagogiques hétérogènes, et d’ainsi participer à la démarche générale de diffusion et de partage des corpus d’apprentissage. Les fonctionnalités de ViCoDili se limitent actuellement à la manipulation des interactions dans leur structure d’origine. Elles devront s’enrichir de moyens de navigation dans les autres composants d’un corpus “complet“ tels que le scénario pédagogique, le contexte de la session de formation ou les retours d’enquête auprès des participants. L’étude n’a pas non plus pris en considération le cas des corpus multimodaux qui commencent à apparaître au travers de plateformes de formation spécifiques.

Les questionnements soulevés par cette recherche concernent principalement la pratique de l’analyse d’interactions en ligne et la diffusion de corpus. L’activité d’analyse est en phase d’instrumentation grâce à des outils logiciels de mieux en mieux adaptés aux données numériques issues des environnements de communication médiatisée et grâce à la prise en

compte des pratiques des usagers. Une première question reste alors posée : quelles sont les conditions d'appropriation de ces environnements par les chercheurs ? Une seconde question concerne plus généralement le principe de diffusion des corpus : quelles sont les conditions nécessaires et suffisantes à l'échange de corpus entre chercheurs en termes de préparation, de normalisation et de méthodologie d'étude de ces corpus ?

Références

Bibliographie

- BARRÉ V., EL-KÉCHAI H., CHOQUET C. (2005). Re-engineering of collaborative e-learning systems: evaluation of system, collaboration and acquired knowledge qualities, In: AIED'05 Workshop: Usage analysis in learning systems, juillet 2005, Amsterdam, Pays-Bas, p. 9-16.
- BOURDON S. (2000). L'analyse qualitative informatisée : logique des puces et quête de sens, *Recherches Qualitatives*, 21, pp.21-44 http://www.recherche-qualitative.qc.ca/Textes_PDF/21Bourdon.pdf
- BRUILLARD É. (2006). Editorial du Numéro Spécial Forum "Le forum de discussion : un cas d'école pour les recherches en EIAH", *Revue STICEF*, Volume 13, 2006, ISSN : 1764-7223, <http://sticef.org>
- CHARLIER B., DAELE A. (2006). Comprendre les communautés virtuelles d'enseignants : pratiques et recherches. Paris : L'Harmattan, 306 p.
- CHOQUET C., LUENGO V., YACEF K. (2007). Usage Analysis of Learning Systems: Existing Approaches and Scientific Issues. *Journal of Interactive Learning Research special issue - Vol 18, n° 2*, AACE Publications , http://www.editlib.org/index.cfm?fuseaction=Reader.TOC&sourceissue_id=363
- COLLECTIF de CHASSENEUIL (2001). *Accompagner les formations ouvertes*. Paris : L'Harmattan, 2001.
- DEJEAN-THIRCUIR C., MANGENOT F. (dir. ; 2006a). Les échanges en ligne dans l'apprentissage et la formation, Présentation du n° spécial *Le Français dans le monde*, *Recherches et applications* n° 40, Paris, CLE International, p. 5-13.
- DEJEAN-THIRCUIR C., MANGENOT F. (2006b). Pairs ou tutrices ? Pluralité des positionnements d'étudiants de maîtrise FLE lors d'interactions en ligne avec des apprenants australiens, *Le Français dans le monde*, *Recherches et applications* n° 40, Paris, CLE International.
- DEGACHE, C. (2006). Aspects du contrat didactique dans une formation plurilingue ouverte et à distance ». Dans Mangenot, F. & Dejean-Thircuir, C. (coord.) *Les échanges en ligne dans l'apprentissage et la formation*, numéro thématique *Le Français Dans Le Monde*, *Recherche & Applications* n° 40, juillet 2006, p. 58-74.
- DESJARDINS C. (2002). *Vers une méthodologie d'analyse des construits cognitifs collaboratifs produits dans les forums électroniques textuels asynchrones*. Thèse en communication, Université de Montréal.
- DIMITRACOPOULOU A., BRUILLARD E. (2007). Interfaces de forums enrichies par la visualisation d'analyses automatiques des interactions et du contenu, *revue STICEF*, vol. 13, 2006, en ligne sur www.sticef.org.
- GIBBS, W. J., OLEXA V., BERNAS R.S. (2006). A visualization tool for managing and studying online communication. *Educational Technology & Society*, 9 (3), 232-243.
- HEER, J., BOYD, D. (2005). Vizter: Visualizing Online Social Networks. in *InfoVis 2005*, IEEE Symposium on Information Visualization.
- HENRI F., PERAYA D., CHARLIER B. (2007). La recherche sur les forums de discussion en milieu éducatif : critères de qualité et qualité des pratiques, Conférence plénière au colloque *Épal Échanger pour apprendre en ligne*, Grenoble, juin 2007.
- HENRI F., CHARLIER B. (2005). L'analyse des forums de discussion pour sortir de l'impasse, *Symphonic : Symposium, formation et nouveaux instruments de communication*, Amiens, http://www.dep.u-picardie.fr/sidir/articles/henri_charlier.htm

- HENRI, F., LUNDGREN-CAYROL, K. (2001). Apprentissage collaboratif à distance. Pour comprendre et concevoir les environnements d'apprentissage virtuels. Sainte-Foy, Québec : Presses Universitaires du Québec.
- HENRI, F., PUDELKO, B. (2003). Understanding and analyzing activity and learning in virtual communities. *Journal of Computer Assisted Learning*, 19, 474-487
- HERRING S. C. (2004). Computer-mediated discourse analysis: An approach to researching online behavior. In: S. A. Barab, R. Kling, and J. H. Gray (Eds.), *Designing for Virtual Communities in the Service of Learning*, New York: Cambridge University Press, p. 338-376.
- LAPERROUSAZ Ch. (2006). Le suivi individuel d'apprenants engagés dans une activité collective à distance, Thèse de l'Université du Maine, spécialité informatique, Le Mans, France, décembre 2006.
- MACKAY, W., FAYARD, A.L (1997). HCI, Natural Science and Design : A Framework for Triangulation Across Disciplines, DIS'97 : Designing Interactive Systems. Amsterdam, Holland.
- MANGENOT F. (2002). Écriture collective par forum sur le Web : un nouveau genre d'écrit universitaire ? *Communication, Les défis du Web - Lyon 2002*. <http://www.interdisciplines.org/defispublicationweb/papers/3/version/original>
- MANGENOT F. (2006), Analyser les interactions pédagogiques en ligne, pourquoi, comment ? , Colloque « La langue de la communication médiatisée par les technologies de l'information et de la communication », Bordeaux, France, mai 2006
- MANGENOT F. (2007), Pratiques pédagogiques instrumentées et propriétés des outils : le cas des forums, In Lamy, Mangenot, Nissen (coord., 2007) Actes du colloque Échanger pour apprendre en ligne (ÉPAL), Grenoble, juin 2007. <http://w3.u-grenoble3.fr/epal/actes.html>
- NORAS, M., REFFAY, C., BETBEDER, M.-L. (2007). Structuration de corpus de formation en ligne en vue de leur échange. "Environnement Informatique pour l'Apprentissage Humain", Lausanne, Suisse, p. 59-64, <http://edutice.archives-ouvertes.fr/edutice-00154372>
- NORMAN, D.A. (1988). *The Design of Everyday Things*. New York, NY: Basic Books.
- PARAYA D. (2005). Axes de recherches sur les analyses de communication dans les forums Notes pour un texte, *Symphonic : Symposium, formation et nouveaux instruments de communication*, Amiens, <http://www.dep.u-picardie.fr/sidir/articles/peraya.htm>
- PERNIN J-P. (2006). Normes et standards pour la conception, la production et l'exploitation des EIAH, Chap. 9, in *Environnements Informatiques pour l'Apprentissage Humain*, Grandbastien M., Labat J.-M. (ed.), *Traité IC2 Information Commande Communication*, Hermès, Paris, avr 2006, p.201-222
- PUDELKO, B., DAELE, A., HENRI, F. (2006). Méthodes d'études des communautés virtuelles. Dans B. Charlier et A. Daele, *Comprendre les communautés virtuelles d'enseignants : pratiques et recherche*, L'Harmattan, Paris.
- PREECE J., MALONEY-KRICHMAR D. (2005). Online communities: Design, theory, and practice. *Journal of Computer-Mediated Communication*, 10(4), article 1. <http://jcmc.indiana.edu/vol10/issue4/preece.html>
- RABARDEL P., (1995), *Les hommes et les technologies, approche cognitive des instruments contemporains*, Colin, Paris, 1995.
- REFFAY Ch. (2007), Symposium "Corpus d'apprentissage en ligne : conception, réutilisation, échange", colloque ÉPAL Échanger pour apprendre en ligne, Grenoble, juin 2007, http://mulce.univ-fcomte.fr/epal_symposium/
- REFFAY, C., CHANIER, T., NICOLET, J., (2002). Produire ensemble pour apprendre : expérience d'une simulation globale en ligne. Actes du Colloque national Apprendre avec l'Ordinateur, Bordeaux, France.
- REFFAY C., LANCIERI L. (2006). Quand l'analyse quantitative fait parler les forums de discussion, *Revue STICEF*, Volume 13, 2006, en ligne sur www.sticef.org
- REFFAY, C., TEUTSCH, Ph. (2007) Anonymisation de corpus réutilisables., <http://edutice.archives-ouvertes.fr/edutice-00158877>
- REFFAY Ch., CHANIER Th., NORAS M., BETBEDER ML. (2007). Contribution à la structuration de corpus d'apprentissage pour un meilleur partage en recherche. In Lamy, Mangenot, Nissen (coord., 2007) Actes du colloque Échanger pour apprendre en ligne (ÉPAL), Grenoble, juin 2007. <http://w3.u-grenoble3.fr/epal/actes.html>

- ROURKE, L., ANDERSON, T., GARRISON, R., ARCHER, W. (2001). Methodological Issues in the Content Analysis of Computer Conference Transcripts. In International Journal of Artificial Intelligence in Education, 12, p.8-22, http://communitiesofinquiry.com/documents/2Rourke_et_al_Content_Analysis.pdf
- SETTOUTI LS., PRIÉ Y, MILLE A., MARTY JCh. (2006) Systèmes à base de traces pour l'apprentissage humain, in TICE 2006, oct 2006, Toulouse.
- SIDIR M., BRUILLARD É., BARON G.-L. (dir. ; 2006). Premières journées Communication et Apprentissage Instrumentés en Réseau (JOCAIR'2006), Université de Picardie Jules Verne, 543 p.
- SIDIR M., BRUILLARD É., BARON G.-L. (dir. ; 2005). Actes en ligne du Symposium Symfonic, Amiens, janvier 2005, <http://www.dep.u-picardie.fr/sidir/articles/index.php>
- SAVOIE-ZAJC, L. (2000). L'analyse des données qualitatives: Pratiques traditionnelles et assistées par le logiciel NUD*IST. Recherches Qualitatives, 21, pp.99-123. http://www.recherche-qualitative.qc.ca/Textes_PDF/21SavoieZajc.pdf
- TESCH R. (1990). Qualitative Research : Analysis Types and Software Tools, New-York, The Falmer Press.
- TEUTSCH Ph., BOURDET JF., GUEYE O. (2004), Perception de la situation d'apprentissage par le tuteur en ligne, Actes de TICE'2004, Compiègne (France), p. 59-66.
- TRICOT A, PLÉGAT-SOUTJIS F., CAMPS JF., AMIEL A., LUTZ G, MORCILLO A. (2003). Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH, Conférence EIAH 2003, Strasbourg, France.

Sites

- Projet Odil (2004). Outils et Didactique pour les Interactions en Ligne, Résumé du projet : <http://w3.u-grenoble3.fr/epal/image/resume-odil.pdf>
- Projet Mulce (2006). Échange de corpus d'apprentissage multimodaux, LIFC Université de Besançon, <http://mulce.univ-fcomte.fr>
- Projet Odil (2006). Outil de Visualisation de Discussion en Ligne, Lium Université du Maine, <http://demolium.univ-lemans.fr>
- Colloque ÉPAL (2007). “Échanger pour apprendre en ligne“, Grenoble, France, juin 2007, <http://w3.u-grenoble3.fr/epal/>
- Consortium IMS-LD (2003). IMS Learning Design v1.0, Final Specification, <http://www.imsglobal.org/learningdesign/>

ⁱ Pour répondre aux exigences d'un travail pluridisciplinaire et clarifier certains termes polysémiques, nous considérerons dans cet article que le terme “échanges“ désigne les contributions des participants à un dispositif de FEL tandis que le terme “interactions“ renvoie à l'acte d'analyse de ces échanges en tant qu'objet d'études. D'un point de vue IHM, le terme “interaction“ désigne également les actions et rétroactions qui s'installent entre l'utilisateur et le système interactif.