Environnement audio graphique synchrone : recueil et transcription pour l'analyse des interactions multimodales

Marie-Laure Betbeder, Christophe Reffay, Thierry Chanier

M.-L. Betbeder, C. Reffay, and T. Chanier., Environnement audio graphique synchrone : recueil et transcription pour l'analyse des interactions multimodales. In *Actes des Premières journées Communication et Apprentissage instrumentés en réseau, JOCAIR 2006*, Amiens, France, 6-7 juillet 2006, pp.406-420.

Environnement audio graphique synchrone : recueil et transcription pour l'analyse des interactions multimodales

Marie-Laure Betbeder* — Christophe Reffay* — Thierry Chanier *

** Laboratoire d'informatique de l'université de Franche-Comté*
16 route de Gray, 25030 BESANCON cedex
{prenom.nom}@univ-fcomte.fr

RÉSUMÉ. Cet article présente un travail sur les transcriptions d'actions multimodales en caractérisant ce qui est nouveau et spécifique des situations d'apprentissage collaboratif en environnement audio-graphique synchrone. Une expérimentation appelée Copéas ainsi que l'environnement Lyceum y sont décrits. Nous montrons l'intérêt de l'enregistrement vidéo pour le recueil des données d'interactions dans ces situations et le besoin de les transcrire pour les analyser. Nous avons conçu et développé un outil « Tasync » pour aider le transcripteur à construire la base de données des actions à partir de la vidéo enregistrée. Nous le positionnons par rapport aux outils existants avant de présenter sa conception et son implémentation. Ce résultat nous engage dans la qualité de la construction des corpus en vue de leur échange au sein de notre communauté scientifique.

MOTS-CLÉS : environnement audio graphique synchrone, transcription, multimodalité.

INTRODUCTION

Les environnements audio graphiques synchrones sont aujourd'hui bien au point et d'utilisation aisée, donc leur utilisation en situation de formation se répand rapidement. On sait cependant peu de choses sur la façon dont apprenants et tuteurs y travaillent (Lamy, 2006). Pour comprendre ce qui s'y passe avant même de songer à concevoir des outils d'aide à l'apprentissage ou à la fonction tutorale, il faut être capable d'analyser les interactions multimodales qui s'y déroulent. A l'inverse des situations en présentiel, la localisation (où ils sont) et la perception (ce qu'ils voient) des participants dans ces environnements synchrones sont à la fois critique et difficile à percevoir pour le participant et donc pour le chercheur.

Le défi apparaît à deux niveaux : il faut construire des cadres d'analyse des interactions multimodales et pour ce faire (deuxième niveau) il faut disposer des contenus des actions des acteurs dans chaque modalité.

Certains chercheurs en informatique qui ont développé leur propre environnement ont directement accès aux traces et peuvent donc se livrer à des analyses (Avouris *et al.,* 2005). Mais pour la majorité des environnements à partir desquels vont se construire des formations, ces données multimodales ne seront jamais disponibles. L'enregistrement vidéo de l'écran partagé est alors le moyen de conserver le déroulement des actions. Cependant, pour restituer toutes les dimensions de ces actions, les chercheurs doivent au préalable opérer une retranscription de ces actions à partir du vidéogramme. Des outils de retranscription d'actions multimodales n'existent pas (ex : Transana ou Anvil se restreignent principalement à une transcription de la parole). Précisons que ces environnements synchrones induisent des données hétérogènes de par leur format (fichier son, vidéo, texte, image) et multimodales (audio, textuelles, graphiques). Cette hétérogénéité rend très complexe une analyse. Il est alors nécessaire d'analyser les différents fichiers, les organiser, recréer le fil de la séance, c'est-à-dire restituer la dimension temporelle. Face à la diversité des médias les chercheurs ont besoin de lisibilité et donc besoin de réunir des informations hétérogènes dans un recueil centralisé et indexé.

Notre équipe de recherche travaille sur les deux niveaux de problème (analyse et retranscription), mais dans ce papier nous abordons uniquement celui de la retranscription. Pour cela nous décrivons notre expérimentation dans la section 1 puis la problématique (section 2) et les travaux existants (section 3). Avant de conclure nous présentons dans les sections 4 et 5 notre solution et l'outil de retranscription d'actions multimodales (Tasync) développé.

1. Contexte

1.1 *Dispositif pédagogique*

Le projet de recherche pluridisciplinaire Copéas (Communication Pédagogique en environnement orienté Audio Synchrone) mené par deux équipes (sciences du langage et informatique) a permis de réaliser une expérimentation écologique qui s'est déroulée sur 16 séances (8 par groupe) dans un environnement audio graphique synchrone. Il s'agit d'une formation

qui vise à développer des compétences d'expression orale dans un contexte professionnel en anglais langue seconde chez 14 apprenants en master professionnel FOAD (Université de Franche-Comté). Le scénario de la formation, conçu par les tuteurs anglophones de l'Open University, propose des activités collaboratives sur la négociation de critères d'évaluation de sites Web pédagogiques.

1.2 *Lyceum : plateforme audio graphique synchrone*

La plateforme audio graphique synchrone utilisée dans cette expérimentation est Lyceum : plateforme développée et utilisée au sein de l'Open University (GB). En tant qu'environnement d'apprentissage audio graphique synchrone, Lyceum permet à un enseignant/tuteur de retrouver, à distance, des apprenants en mode synchrone. Les différents participants connectés à l'environnement peuvent donc se parler en temps réel, intervenir dans un clavardage et voir/modifier simultanément des productions textuelles ou graphiques.

L'interface de Lyceum (cf. Figure 1) rassemble trois composants activables simultanément :

- un composant spatial (cadre 1) pour se déplacer du hall d'entrée dans les salles de travail ou techniques,
- un composant rassemblant les outils de communication synchrone (cadre 2) : module audio (avec outil de vote) à gauche et clavardage dans la partie inférieure droite,
- un composant permettant d'intégrer selon les besoins de l'activité différents modules de production collaborative : traitement de texte, carte conceptuelle et/ou tableau blanc (cadre 3).

Dans Lyceum, tous les acteurs (tuteur et apprenants) disposent de la même interface et des mêmes droits.

1.3 *Les interactions en environnement audio-graphique synchrone*

Pour rendre compte de leur richesse et de la complexité de leur analyse, nous donnons ici une liste des actions possibles et perceptibles dans un environnement tel que Lyceum en les illustrant lorsque c'est possible, sur la Figure 1.

L'acteur peut se situer dans l'espace grâce aux rectangles grisés dans le composant spatial, ici, l'acteur se trouve à l'étage 1 dans la salle 101. Il peut aussi voir qui se trouve dans le hall d'entrée (lobby). Lorsque d'autres

étages ou salles sont occupés, leurs numéros apparaissent en gras. Les acteurs ne peuvent percevoir les autres (audio, graphique, clavardage, productions) que s'ils sont réunis dans la même salle. Ils sont alors listés (par ordre d'arrivée) dans le composant de communication (cadre 2).

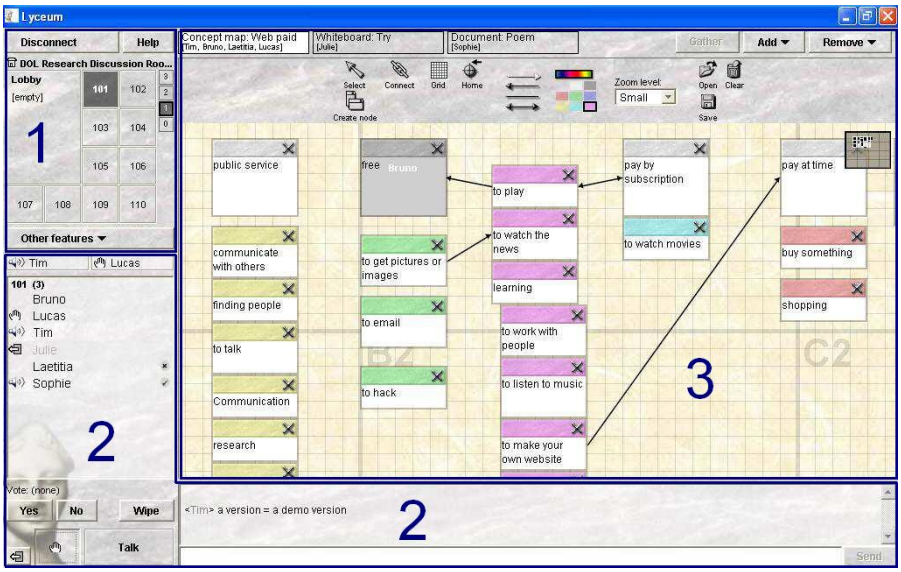

Figure 1. *Interface de Lyceum*

Chacun peut, à tout instant, parler en sélectionnant le bouton « Talk » (ex : Tim et Sophie), lever la main pour demander la parole (ex : Lucas), voter « Yes » (ex : Sophie) ou « No » (ex : Laetitia) pour répondre collectivement à une question ou prendre une décision. Il est possible de notifier aux autres une absence momentanément (ex : Julie). Le clavardage est un outil qui s'ajoute à cet ensemble. Il est souvent utilisé pour des conversations parallèles au flux oral pour le désambiguïser ou éviter de le perturber (ex : salutations à un nouvel arrivant).

A cet ensemble déjà riche, s'ajoute la possibilité, pour le groupe, d'ouvrir plusieurs modules de production collaborative de 3 types : traitement de texte, tableau blanc et carte conceptuelle (Figure 1). Chaque module est visualisable (indépendamment) par chacun grâce aux onglets de la frise supérieure du cadre 3. On peut y lire à chaque instant la liste (parfois incomplète) des acteurs visualisant tel ou tel module. Les acteurs réunis dans une même salle peuvent donc partager l'ensemble des communications (audio, iconique et clavardage) sans nécessairement visualiser le même document/module. Tous les acteurs peuvent ajouter ou supprimer un module, sauvegarder ou charger un document préparé avant dans le module,

et bien sûr, créer, éditer, ou supprimer les objets propres à chaque type de module (Texte : paragraphes ; Carte conceptuelle : concepts et relations ; Tableau blanc : traits, formes, textes, etc.). On voit (Figure 1) Bruno en train d'éditer le concept « free » de la carte conceptuelle « Web Paid » visualisée à cet instant par Tim, Bruno, Laetitia et Lucas. Cette action n'est donc pas perceptible pour Julie ou Sophie qui visualise respectivement le tableau blanc « Try » et le texte « Poem ».

La richesse et la souplesse d'utilisation de ce type d'environnements (Vetter, 2004), tant appréciées par les acteurs de telles situations constituent un défi dans la complexité d'analyse pour le chercheur. La notion de groupe y est subtilement conjuguée selon les modes d'interaction et de production pour permettre à l'acteur de communiquer avec le groupe tout en lui ménageant un espace suffisant pour participer à la production. Il devient alors impossible pour l'observateur de noter l'intégralité des mouvements, communications et actions d'un groupe au cours d'une séance. Les conserver pour une analyse ultérieure est une nécessité.

1.4 *Protocole de recueil*

Lors de la conception de l'expérimentation, une phase importante consiste à définir quelles traces, interactions, données devront être recueillies durant ou après l'expérimentation. Ces données identifiées permettent la mise en place d'un protocole de recueil de données précis. Ce protocole inclut les enregistrements audio et vidéo, les enregistrements des productions individuelles et collectives et l'organisation des données.

Pour l'expérimentation Copéas, nous avons choisi d'enregistrer par captures d'écrans vidéo l'interface de l'environnement Lyceum. Toutes les séances et toutes les salles utilisées ont donc été enregistrées. Nous avons également récupéré des logs (entrée/sortie) des serveurs de l'Open University et les traces de clavardage. Nous avons élaboré et administré un questionnaire à l'issue de l'expérimentation, mené des entretiens semi-dirigés et des auto-confrontations à partir d'extraits vidéo choisis (*Critical Event Recall*). Pour ne citer que quelques chiffres, l'ensemble des données compte 37 vidéogrammes d'une durée cumulée de 27h, 512 fichiers (productions, audiogrammes des entretiens, questionnaires, etc.) et occupe 35 Go.

2. Problématique

Actuellement, les plateformes d'enseignement à distance se multiplient et bon nombre de chercheurs en EIAH, issus de différentes disciplines, expérimentent ces plateformes dans le but d'observer des apprentissages, des phénomènes de groupe, de collaboration, etc.

Du côté des plateformes asynchrones maintenant très répandues, le recueil des données pose en général moins de problèmes puisque, par sa nature asynchrone, l'environnement doit conserver les productions et les communications de chaque acteur et les rendre visibles pour les autres. Le chercheur peut donc aisément récupérer ces informations minimales. Pour les compléter, il doit parfois extraire du logiciel serveur ou des logs http, des précisions telles que le fait qu'un acteur a ouvert un message ou non, ou les dates de connexion, etc. Les sauvegarder à l'état brut risque de poser le problème de leur pérennité si le logiciel serveur initial change de version ou n'est plus disponible.

A l'inverse, dans un environnement synchrone, l'enregistrement des actions côté serveur est secondaire. C'est la gestion (pour les différents groupes) des flux audio et/ou vidéo qui devient la priorité pour permettre la fluidité des conversations. L'ergonomie des interfaces clients s'apprécie aussi par la lisibilité (pour les autres) des interventions d'un acteur. Le recueil des données devient alors plus critique pour le chercheur. L'enregistrement vidéo de l'interface dans toutes les salles est coûteux, mais constitue un objet échangeable et indépendant de l'environnement.

L'hétérogénéité, la multimodalité et la simultanéité des actions nous approchent des situations en présentiel et justifie le recours à la vidéo. Mais la localisation et la perception nous en éloignent. Ainsi, nous avons besoin, dans nos modèles de représentation des actions, de préciser de manière spécifique, cette localisation et cette perception pour chaque acteur et à chaque instant.

Les vidéos enregistrées permettent de revoir le déroulement de la séance et donc toutes les actions des différents acteurs. On peut visualiser les entrées/sorties des acteurs dans une salle, voir et entendre les prises de paroles, lire le clavardage, voir les actions graphiques des acteurs dans les modules d'édition collaborative, etc. Cependant ce format est un « fondu à plat » des actions réalisées au sens ou aucune de ces actions n'est repérable. Aucun traitement informatique ne peut être effectué pour rechercher des occurrences, repérer des schèmes d'actions, retrouver une séquence particulière à partir d'une information, etc. Il est également impossible d'en

extraire des données synthétiques sur les outils, les espaces les plus utilisés par les différents types d'acteurs.

De ces difficultés est née le besoin d'un outil permettant à une personne de transcrire des actions à partir d'une vidéo d'expérimentation. La vidéo étant un moyen très souvent utilisé dans des expérimentations, un tel outil pourrait aussi convenir à des chercheurs de disciplines diverses pour d'autres transcriptions sans utiliser forcément les spécificités de localisation et de perception qui caractérisent les nôtres.

Les premières retranscriptions ont été effectuées à l'issue de l'expérimentation et en parallèle du travail de conception de l'outil. Elles ont été réalisées à l'aide d'un tableur Excel, et concernaient la retranscription de l'audio, du clavardage et des votes. Elles ont permis des analyses sur le temps de paroles des apprenants et tuteurs ou la prédominance d'un outil de communication (Chanier et Vetter, 2006). La Figure 2 propose un exemple de ces retranscriptions. La première colonne correspond au temps de la vidéo (min : sec), la deuxième identifie chaque action en la numérotant par canal de communication utilisé, la troisième spécifie l'acteur et la dernière le contenu de l'action (transcription audio, message de clavardage ou valeur du vote).

18:40	aud79	AR4	euh no + euh I don't know the + the style ++ in french it's a band + named + {les enfoirés} ++ you know euh + {enfoirés} ↑ +++
18:57	vot21	AR7	yes
18:58	aud80	TutR	anybody else know ↑
19:00	clav23	AR6	french's singers
19:02	vot22, vot23, vot24	AR3, AR2, AR6	yes
19:07	vot25, vot26	TutR, AR1	no
19:12	aud82	AR4	(you can translate by)
19:12	aud83	TutR	(everybody except) + everybody except the foreigners + that is AR1 and I + I've not heard about this euh we haven't + so did you enjoyed the concert AR4 ↑ + was it good ↑ +
19:16	clav24	AR3	famous in France
19:17	clav25	AR2	!!
19:18	clav26	AR4	restaurant of the heart
19:27	clav27	AR1	ok
19:28	vot27	AR4	yes

Figure 2. *Retranscription multimodale*

Cet exemple ne montre qu'une partie des informations décrivant les interactions. Un outil de transcription intégrant la complexité et la structuration des informations à enregistrer est nécessaire. On précise pour chaque action : le temps, le canal, l'acteur, le contenu mais aussi le document et/ou lieu dans lequel elle a été effectuée, l'enregistrement vidéo, la séance, le groupe, etc.

3. Positionnement par rapport aux travaux existants

Le domaine des SHS (Sciences Humaines et Sociales) travaille depuis longtemps à partir de corpus vidéo. Ayant été confronté aux mêmes besoins de retranscription, de nombreux outils de retranscription ont été conçus. Même si les situations retranscrites n'ont pas beaucoup de points de similarité avec les nôtres (il ne s'agit pas nécessairement de situations d'apprentissage, ce sont des humains qui sont filmés et non une interface d'un environnement d'apprentissage) ces outils semblent cependant correspondre à nos besoins.

Ces outils comme par exemple Anvil, EXMARaLDA, Transana, ELAN permettent la transcription de l'audio et parfois de la gestuelle (Rohlfing *et al.,* 2005). Plusieurs de ces outils utilisent la métaphore d'une partition musicale pour permettre la visualisation de l'alignement temporel des actions, un graphique représentant le signal sonore et un lecteur médias intégré. Certains paramétrages sont parfois possibles (types d'annotation, acteurs) mais ces conventions doivent souvent être modifiées manuellement dans le code source (*Xi*Tools). Cependant, les modèles d'actions sont adaptés à la nature des situations à transcrire : dialogue entre deux personnes ce qui ne permet pas de décrire les informations spécifiques à nos situations (ex : localité, perception).

Dans le domaine informatique, l'équipe HCI (Avouris *et al.,* 2005) a conçu l'outil ColAT (Collaboration Analysis Tool). ColAT est un outil d'analyse d'activités collaboratives qui prend comme données les traces récupérées, dans un format particulier, issues de Synergo (plate-forme développée par la même équipe) ou d'une autre plate-forme. L'objectif de cet outil est d'annoter un ensemble d'actions en vue d'analyser les interactions. Cet outil n'est pas non plus réutilisable dans notre contexte, car d'autres caractéristiques des actions (notamment la durée de l'action) ne pourraient être décrites.

4. Solution conceptuelle

Notre volonté n'est pas de construire une énième plateforme ad hoc qui permettrait de tracer toutes les actions souhaitées. Nous considérons qu'il est préférable d'envisager le problème des corpus quel que soit son environnement et non de travailler à partir d'une plateforme spécifique. Si nous avions la main sur le développement de la plateforme, nous pourrions tracer automatiquement certaines actions (e.g. le clavardage, les votes, actions de production), mais certaines données comme par exemple l'audio ne pourraient pas être retranscrites directement.

Dans notre expérimentation les données principales sont les vidéos. Ce format est très largement utilisé dans bon nombre de disciplines. Rappelons que notre outil n'est pas destiné uniquement à des chercheurs en EIAH mais à toutes les disciplines. Nous souhaitons donc concevoir un outil de transcription assez général, prenant en entrée des vidéos, et construisant en sortie une base de données des actions qui s'y déroulent.

Les vidéos ressources peuvent aussi bien concerner l'enregistrement d'une plateforme d'apprentissage que des individus. Pour supporter ces différentes situations, les types d'actions ne peuvent pas être déterminés au préalable, il faut permettre à l'utilisateur de les définir. Il serait ainsi possible de transcrire les actions des apprenants décrits dans (Smith et Gorsuch, 2004). Ces auteurs travaillent sur la prise en compte des transcriptions multi-dimensionnelles : intentions, gestuelle, paroles des apprenants. Leur cadre est l'utilisation d'un clavardage. Leur protocole expérimental comprend l'enregistrement vidéo par plusieurs caméras des apprenants et de leur écran.

Notre démarche de conception implique en parallèle de la phase de transcription la définition d'un modèle d'organisation des données (cf. Figure 3) lequel détermine la structure de la base de données. Cet ensemble constitue l'outil de transcription.

Le modèle que nous proposons grâce à la base de données est basé sur la notion de temps de sorte de ne pas perdre la dimension temporelle de la vidéo. En effet, les enregistrements vidéo ne permettent pas d'isoler des actions ni de les indexer mais ils permettent de « rejouer » la séquence en gardant la séquentialité des actions.

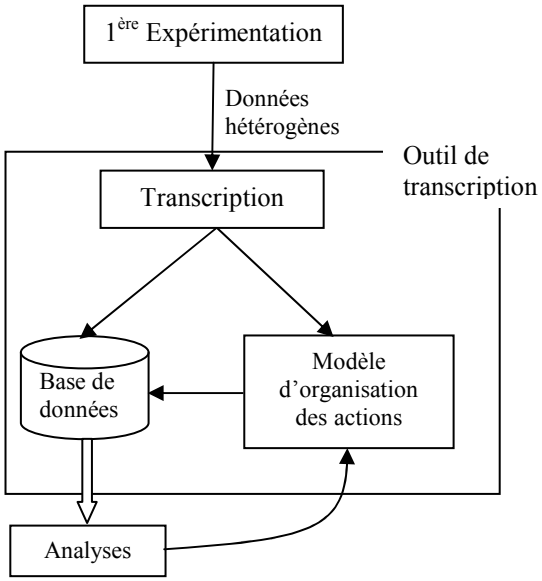

Figure 3. *Démarche de conception*

De plus, nous sommes conscients que manipuler une base de données n'est pas évident pour tout chercheur. Pour cela, nous proposons une fonctionnalité de visualisation des actions multimodales. Ceci permettrait au chercheur de visualiser la vidéo initiale « sous titrée » par les actions transcrites (par la première fonctionnalité de l'outil) de manière plus linéaire qu'une base de données, tout en gardant l'aspect essentiel d'indexation. Pour résumer, la base de données regroupe l'ensemble des actions de l'expérimentation et la fonctionnalité de visualisation permet d'isoler un extrait vidéo.

5. Implémentation

5.1 *Modélisation des données*

L'élément central de ce projet est le modèle des différentes actions multimodales réifié dans la structure de la base de données. Ce modèle définit les sessions (plages horaires pour chaque séance) associées à un groupe (classe) et ayant donné lieu à des enregistrements (vidéos,

clavardages, documents produits, etc.). Afin de repérer les espaces (dans la plateforme audio graphique synchrone), nous avons défini la notion d'espace-temps ET = (S, t_0, t_1) comme un lieu S (salle ou espace virtuel) où un groupe se retrouve effectivement dans un intervalle de temps [t_0, t_1] avec t_0 : la date d'entrée de la première personne dans l'espace et t_1 : la date de sortie de la dernière personne de cet espace. Cette notion permet de regrouper les actions ayant eu lieu dans un même espace-temps, c'est à dire, en général partagées par un groupe de personnes identifiées. Des actions ayant lieu dans un espace à un instant donné, ne sont pas lisibles/audibles par les personnes se trouvant au même moment dans un espace différent.

Chaque action est ainsi caractérisée (au minimum) par un canal (audio, vote, clavardage, outil de production, etc.), une valeur précisant le sens de l'action, une date de début et éventuellement une date de fin, l'acteur l'ayant réalisée et l'espace-temps dans lequel elle a eu lieu. Des tables spécifiques peuvent être ajoutées pour donner d'autres précisions sur des actions : ainsi, nous avons une table parole contenant le texte qui a été transcrit de l'audio ainsi que (si nécessaire) la transcription phonétique de certains extraits. Une autre table spécifique nous permet de relier les actions aux espaces de production appelés Espace-Document. Un espace-document correspond à un module de production collaborative (ex : Tableau blanc, Traitement de texte ou Carte conceptuelle dans Lyceum). Chaque type d'espace-document autorise certaines actions spécifiques (créer, éditer, supprimer, sélectionner, etc.) sur des objets spécifiques (lignes, rectangles, ellipses, textes, paragraphes, concepts ou relations). Il est ouvert à l'intérieur d'un espace. Il peut être utilisé dans plusieurs espaces-temps (successifs) correspondant à un même lieu. Il peut y avoir plusieurs espace-documents ouverts simultanément dans un même lieu, mais chaque acteur ne peut en visualiser qu'un seul à la fois.

Si l'on veut rendre compte des actions dans des situations collectives, il faut être capable de caractériser pour chaque action : qui la réalise, qui peut la voir, et donc où et quand elle a lieu. De façon duale, nous devons pouvoir extraire aisément de notre base de données, pour chaque acteur, les actions qu'il a réalisées et celles qui ont été réalisées par d'autres, mais qu'il a pu voir.

Partant de ces données exhaustives, nous espérons découvrir des séquences récurrentes d'actions permettant de rendre compte du comportement d'un groupe apprenant dans ces nouvelles situations. Mais nous souhaitons partager ces données avec d'autres chercheurs afin qu'ils puissent y appliquer leurs propres analyses et permettre enfin la comparaison des méthodes ayant opéré sur les mêmes données et des conclusions qui en résultent.

Mais ce catalogage exhaustif et minutieux de toutes les actions est une entreprise titanesque. Réalisée à la main, elle expose le catalogue à certaines incohérences dues à des erreurs de manipulation (estampillage de la date et heure précise par exemple). Ces manipulations peuvent être accompagnées par un outil d'aide à la transcription permettant à la fois d'accélérer le processus de transcription et d'améliorer la qualité du résultat ; c'est ce que se propose de faire Tasync : outil de Transcription d'Actions Synchrones.

5.2 *L'outil Tasync*

Tasync est un prototype développé pour aider la transcription de données issues d'expérimentations à partir de vidéos. Cet outil ne se limite pas à la transcription de vidéos de Lyceum ni même d'une autre plateforme d'apprentissage, il est en effet possible de typer les actions en fonction du contexte de l'expérimentation et des besoins des chercheurs. Pour un gain de temps de la phase de transcription, nous avons cependant personnalisé l'interface avec des icônes représentant les actions possibles dans Lyceum.

Tasync propose une autre fonctionnalité, celle de visualiser les transcriptions. Pour cela, l'outil utilise le formalisme SMIL (*Synchronized Multimedia Integration Language*), langage de représentation et de description des données, basé sur une syntaxe XML. Concrètement, une fois la transcription effectuée, le chercheur peut rejouer tout ou partie des vidéos augmentées de « sous-titrages » des transcriptions.

L'interface de Tasync (cf. Figure 4) est composée de 3 zones. La première (barre verticale gauche) concerne les différentes fonctions de Tasync : accès à l'enregistrement des acteurs, à la définition des types d'action, au code SMIL généré et à l'importation des logs de clavardage. La deuxième zone (centrale) est un lecteur vidéo avec deux barres de défilement de la vidéo de précisions différentes. La troisième zone concerne la transcription des actions. Une icône (horloge) permet de mémoriser l'heure de début d'une action, divers champs permettent de renseigner le type d'action, l'acteur, l'audio, la phonétique. L'utilisateur peut également visualiser les actions déjà transcrites (pour plus de détails, cf. (Djouad, 2005)).

Une fois la phase de transcription effectuée, la base de données est renseignée. La base de données permet alors d'effectuer des analyses automatiques comme par exemple : le décompte des tours de parole d'un acteur, la durée de ses interventions (pour une analyse quantitative sur la participation) ou la lecture de toutes ses interventions écrites et orales (retranscrites) (pour une analyse qualitative de son niveau d'anglais).

Tasync a déjà été utilisé sur des échantillons, testé par le transcripteur de l'expérimentation puis modifié. Nous travaillons actuellement sur la transcription des actions graphiques dans les modules d'édition collaborative. A partir de cette transcription finale nous pourrons nous livrer à des analyses longitudinales (analyser des séquences interactives tout au long du corpus) et synchroniques (développer notre modèle d'analyse du discours multimodal avec nos linguistes).

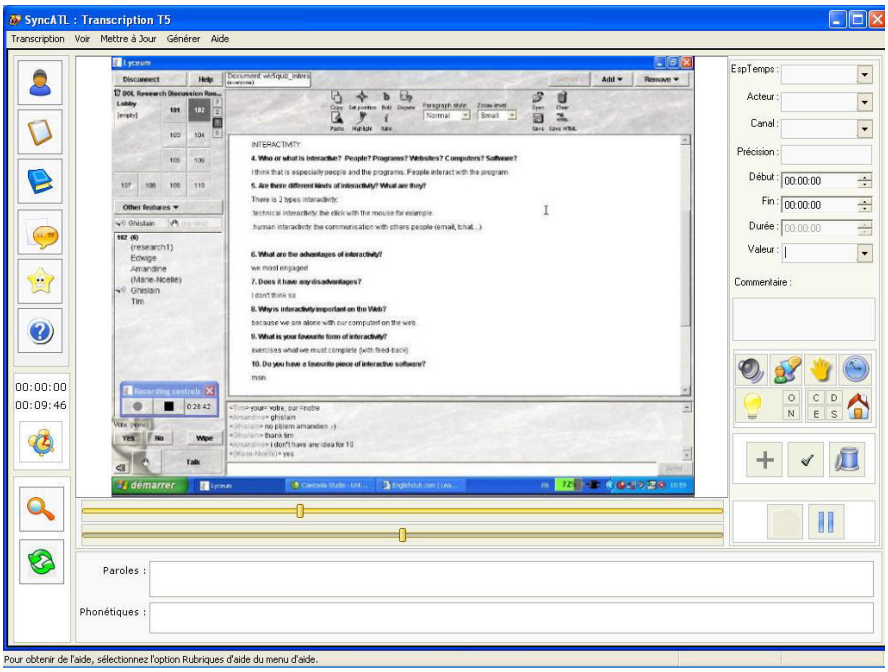

Figure 4. *Copie d'écran de TaSync*

5.3 *Problème de synchronisation*

La centralisation des données représentant les actions suppose une référence temporelle unique. Lorsqu'il y a plusieurs sources susceptibles d'enregistrer (et donc d'estampiller) les actions, il est extrêmement critique de synchroniser les horloges. Pour Copéas, nous avons au moins 3 références temporelles : l'heure du serveur (décalée d'une heure en hiver avec les clients en France) (étant lui-même composé de plusieurs unités centrales, nous avons dû vérifier la synchronisation entre le serveur-maître et les serveurs-esclaves), le temps dans la vidéo et l'heure de la machine client réalisant l'enregistrement de la vidéo (dans cette expérimentation, il y

en avait deux). Au moment de l'intégration des actions dans la base de données, nous avons choisi de traduire chaque estampille dans un référentiel temporel unique de façon à pouvoir reconstituer les séquences d'actions.

6. Conclusion

Ce papier présente nos travaux sur la retranscription multimodale. Nous avons présenté le projet Copéas et les besoins de retranscription qui ont émergé lors de la phase d'analyse des interactions. Nous avons montré que le recueil des données pour les plateformes synchrones est plus coûteux que pour les plateformes asynchrones. Il est essentiellement à la charge du chercheur, mais cela rend le résultat indépendant de la plateforme. La localisation et la perception dans ces environnements synchrones nécessitent d'affiner la notion de groupe. Elles requièrent donc des modèles de représentation spécifiques pour rendre compte du fait que les acteurs sont ensemble ou non et préciser ce qu'ils partagent. L'hétérogénéité des fichiers, la multimodalité et la simultanéité des interactions dans Lyceum nous ont conduit à choisir la vidéo comme moyen d'enregistrement central. Mais la nature de nos analyses a rendu nécessaire la transcription de ces vidéos pour construire une base de données des interactions. L'outil Tasync présenté ici a été développé et testé. Il permet d'aider la transcription des actions issues d'une vidéo pour en injecter les propriétés dans la base de données. Le type des actions étant paramétrable, nous pensons que Tasync peut être exploité dans d'autres contextes.

Ces travaux se poursuivent dans notre projet de recherche MUltimodal Learning Corpus Exchange (Mulce). Nous nous intéressons également à la structuration et à l'échange des corpus (Noras, 2006) et à la découverte de *pattern* dans les interactions multimodales en situations d'apprentissage.

Remerciement

Nous tenons à remercier pour leur participation au programme de formation et / ou au projet de recherche exposés ici, Tim Lewis, Robin Goodfellow et Marie-Noëlle Lamy de l'Open University, GB. Nous remercions également Tarek Djouad qui a effectué son stage dans notre équipe et qui a travaillé sur ces problématiques.

7. Bibliographie

Avouris, N., Komis, V., Fiotakis, G., Margaritis, M., et Voyiatzaki, E., « Logging of fingertip actions is not enough for analysis of learning

activities », Workshop Usage analysis in learning systems, AIED 2005, Amsterdam, July 2005.

Chanier, T., et Vetter, A., Multimodalité et expression en langue étrangère dans une plate-forme audio-synchrone » *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)*, (à paraître), 2006. http://archive-edutice.ccsd.cnrs.fr/edutice-00001436

Djouad, T., Tasync : un outil de transcription multimodale et de visualistaion des actions multimodales, Rapport de Master recherche informatique, Université de Franche-Comté, 2005.

Lamy, M.-N., « Conversations multimodales : l'enseignement-apprentissage de l'oral à l'heure des écrans partagés ». *Le Français Dans Le Monde, numéro thématique Les échanges en ligne dans l'apprentissage et la formation*, F. Mangenot et C. Dejean-Thircuir (coord.), Juillet 2006, (à paraître).

Noras, M., « Un besoin de spécifications des corpus de formation en ligne » Rencontres Jeunes Chercheurs - Environnements Informatiques pour l'Apprentissage Humain, Evry, 11-12 mai 2006.

Rohlfing, F., Loehr, D., Duncan., S., Brown, A., Franklin, A., Kimbara, I., Milde, J-T., Parrill, F., Rose, T., Schmidt, T., Sloetjes, H., Thies, A., Wellinghoff, S., Comparison of multimodal annotation tools - workshop report, Rapport du Wokshop Second Congress of the International Society for Gesture Studies, June 15-18 2005, Université Lyon 2, 2006. http://vislab.cs.vt.edu/~gesture/multimodal/workshop/Report.pdf

Smith, B., et Gorsuch, G.J., « Synchronous computer mediated communication captured by usability lab methodologies: New interpretations » *System*, 32(4) 2004, pp.553-575.

Vetter, A., « Les spécificités du tutorat à distance à l'Open University : enseigner les langues avec Lyceum » *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)*, 07, 2004, pp. 107-129.

Webographie

Lyceum : http://kmi.open.ac.uk/projects/lyceum/

Mulce : http://mulce.univ-fcomte.fr

*Xi*Tools : outils logiciels d'aide à l'analyse de corpus, http://weblex.ens-lsh.fr/projects/xitools/index.htm