Interactions multimodales synchrones issues de formations en ligne : problématiques, méthodologie et analyses

Marie-Laure Betbeder, Maud Ciekanski, Françoise Greffier, Christophe Reffay, Thierry Chanier

▶ **To cite this version:**

Marie-Laure Betbeder, Maud Ciekanski, Françoise Greffier, Christophe Reffay, Thierry Chanier. Interactions multimodales synchrones issues de formations en ligne : problématiques, méthodologie et analyses. STICEF (Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation), 2008, EPAL (échanger pour apprendre en ligne), 15, pp.x-x. edutice-00160314v2

HAL Id: edutice-00160314
https://edutice.hal.science/edutice-00160314v2

Submitted on 23 Feb 2008

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

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

Betbeder, M.-L., Ciekanski, M., Greffier, F., Reffay, C. & Chanier, T. Interactions multimodales synchrones issues de formations en ligne : problématiques, méthodologie et analyses. In Basque, J. & Reffay, C. (dirs), *numéro spécial EPAL (échanger pour apprendre en ligne), Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation (STICEF),* vol. 15, à paraitre, http://sticef.org

Interactions multimodales synchrones issues de formations en ligne : problématiques, méthodologie et analyses

▶ Marie-Laure BETBEDER (LIFC, Besançon), Maud CIEKANSKI (LASELDI, Besançon), Françoise GREFFIER (LIFC, Besançon), Christophe REFFAY (LIFC, Besançon), Thierry CHANIER (LASELDI, Besançon)

■ RÉSUMÉ • Le développement d'environnements de formation en ligne permettant aux utilisateurs de communiquer et de travailler ensemble de façon synchrone a suscité l'engouement de plus en plus de formateurs et d'apprenants depuis la dernière décennie. Ces plateformes de formation génèrent des interactions complexes entre les participants dont les analyses renouvellent le questionnement sur les traces et leurs traitements. Les travaux présentés dans cet article portent sur la transcription et l'analyse de données d'interactions multimodales synchrones issues d'une formation en ligne. Une description de la méthodologie développe le protocole de recueil des données, les conventions de transcription, l'organisation et la représentation des données. Des exemples d'analyses effectuées sur ces données à une échelle macroscopique et microscopique permettent de présenter des résultats et de discuter des atouts et des limites des outils employés.

■ MOTS CLÉS • interactions multimodales synchrones, multimodalité, traces.

■ ABSTRACT • Since the last decade, online synchronous audio-graphic learning environments, which enable participants to communicate and collaborate in order to achieve a common task, have become increasingly popular, particularly in the language learning domain. These online learning environments generate complex interactions between participants. Their analysis renews research on traces and their processing. The work presented in this paper concerns the methodology used to transcribe and analyse synchronous multimodal interactions coming from a particular online training. We describe the protocol to collect data, the original coding scheme to transcribe the video data into a set of participant's actions, part of them being speech acts which occurred into the different modalities of the system (audio, textchat, text editing, and websites), the organization, and the representation of such data. Two examples of analysis, at a macro scale and at a micro scale, give a glimpse of our results, from which the benefits and limits of such approaches and tools are debated.

1. Introduction

Les environnements audio graphiques synchrones sont aujourd'hui très performants et d'utilisation aisée, ce qui les rend de plus en plus populaires en situation de formation. Cependant, leur nouveauté et leur spécificité posent un nouveau défi en termes d'analyse :

- d'une part, en termes de méthodologie ; comment enregistrer les données multimodales issues de tels environnements, quelles conventions de transcription adopter ?
- d'autre part, en termes de représentation des données ; comment le chercheur peut-il retrouver des informations, effectuer des traitements statistiques, et de manière générale, comment peut-il accéder à une relecture des situations jouées étant donné la complexité et l'hétérogénéité de celles-ci pour parvenir à les analyser ?

On sait actuellement peu de choses sur la façon dont apprenants et tuteurs travaillent dans ces environnements. 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 orchestrent différents modes et modalités (audio, clavardage, vote, production, etc.) dans un même fil du discours. À l'inverse des situations en présentiel, la localisation (où ils sont) et la perception (ce qu'ils voient et entendent) des différents participants dans ces environnements synchrones sont critiques et difficiles à saisir à la fois pour le participant et pour le chercheur. Les interactions synchrones et audio couplées à la multimodalité sont donc un objet récent complexe à étudier.

Nous souhaitons à travers cet article présenter notre travail de l'étape d'expérimentation à celle des analyses macroscopiques et microscopiques de la multimodalité en décrivant notre méthodologie (propre à la multimodalité et à la synchronie) et en montrant la nécessité de disposer de représentations de ces données complexes afin d'en rendre possible l'analyse.

Pour cela, nous présentons en section 2 la formation en ligne point de départ de ce travail. Elle est issue du projet de recherche pluridisciplinaire Copéas mené par des chercheurs de deux champs disciplinaires (sciences du langage et informatique). L'expérimentation écologique s'est déroulée dans l'environnement audio graphique synchrone Lyceum, développé et utilisé à l'Open University (GB).

Nous décrivons ensuite notre démarche méthodologique en explicitant le protocole de recueil des données (essentiellement l'enregistrement des sessions par vidéo-écrans). Le besoin de transcription d'un certain nombre de données nous a amenés à spécifier les actions, en travaillant sur les formats des traces (champ largement étudié dans la communauté, par exemple : (Avouris *et al.,* 2005), (Courtin et Talbot, 2006),

(Heraud *et al.*, 2005) (Ollagnier-Beldame et Mille, 2007) et, également, en proposant une représentation des interactions multimodales (Harrer *et al.*, 2007).

A partir de cette transcription, nous pouvons donc nous livrer à des analyses longitudinales (analyser des séquences interactives tout au long du corpus) et synchroniques (analyser les stratégies des acteurs ou le processus de réalisation d'une tâche). À partir d'exemples d'analyses d'interactions multimodales, nous discutons de la nécessité de croiser les représentations et outils pour mieux appréhender des phénomènes complexes tels que, par exemple, les mécanismes d'apprentissage et de collaboration permis et suscités par ces dispositifs.

Les travaux présentés dans cet article s'inscrivent dans le projet Mulce[1], projet de structuration des corpus de formation en ligne en vue de leurs échanges (Reffay *et al.*, 2007) (Noras *et al.*, 2007). Ce projet a pour objectif de définir un format commun de définition et de structuration des différents éléments d'un corpus. Ces éléments incluent la description du scénario pédagogique, du protocole de recherche (composé des questions de recherche et du protocole de recueil), les interactions et traces de la formation, les licences publiques et privées et des analyses. L'objet de cet article est donc de présenter une méthodologie adaptée à la multimodalité, à partir de notre formation en ligne et de nos questions de recherche : protocole de recueil et conventions de transcription, ainsi que représentations de ces données. Nous donnons des exemples d'analyses macroscopiques et microscopiques s'appuyant sur les représentations des données avant de discuter des apports et limites des outils et analyses développées.

2. Contexte

2.1. Copéas

Le projet de recherche bi-disciplinaire Copéas (*Communication Pédagogique en environnement orienté Audio Synchrone*) a permis de réaliser en 2005 une expérimentation écologique qui s'est déroulée sur 16 séances (8 par groupe –selon les niveaux en anglais des étudiants -faux débutants ou intermédiaires-avancés) dans un environnement audio graphique synchrone. L'objectif principal de chaque séance était de développer les compétences d'expression orale dans un contexte professionnel en anglais langue étrangère chez les 14 apprenants du master professionnel FOAD (Université de Franche-Comté). Les apprenants (7 hommes et 7 femmes) avaient entre 25 et 50 ans, avec des niveaux d'études, d'expériences professionnelles, des biographies langagières et des habitudes d'utilisation des outils de communication en ligne très

[1] Projet Mulce (*MULtimodal Learning & teaching Corpus Exchange*) : http://mulce.univ-fcomte.fr

diverses. Le scénario de la formation, conçu par les tuteurs anglophones de l'Open University, proposait des activités collaboratives sur la négociation de critères d'évaluation de sites Internet pédagogiques. À la marge de ces activités synchrones, les participants avaient accès à un environnement asynchrone destiné à l'échange des consignes et travaux entre chacune des séances synchrones.

2.1.1. Lyceum : plateforme audio graphique synchrone

La plateforme audio graphique synchrone utilisée dans cette expérimentation est Lyceum[2]. En tant qu'environnement d'apprentissage, Lyceum permet à un 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 (*chat*) et voir/modifier simultanément des productions textuelles ou graphiques.

L'interface de Lyceum (cf. Fig. 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, de façon à garantir à chacun le même statut dans le groupe et favoriser ainsi un travail collégial. Les activités proposées par le tuteur préconisaient l'utilisation d'un type de module de production collaborative (tableau blanc, traitement de texte, carte conceptuelle), sans empêcher le recours à d'autres modules si nécessaire. Les étudiants avaient la possibilité de communiquer par audio et clavardage comme ils le souhaitaient, à partir du moment où leur participation permettait l'avancement du travail du groupe. Les apprenants ont majoritairement communiqué entre eux en anglais, et dans un anglais normé tant à l'oral qu'à l'écrit.

2.1.2. Les interactions en environnement audio-graphique synchrone

[2] Lyceum : http://kmi.open.ac.uk/projects/lyceum/

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 dans le composant de communication (cadre 2).

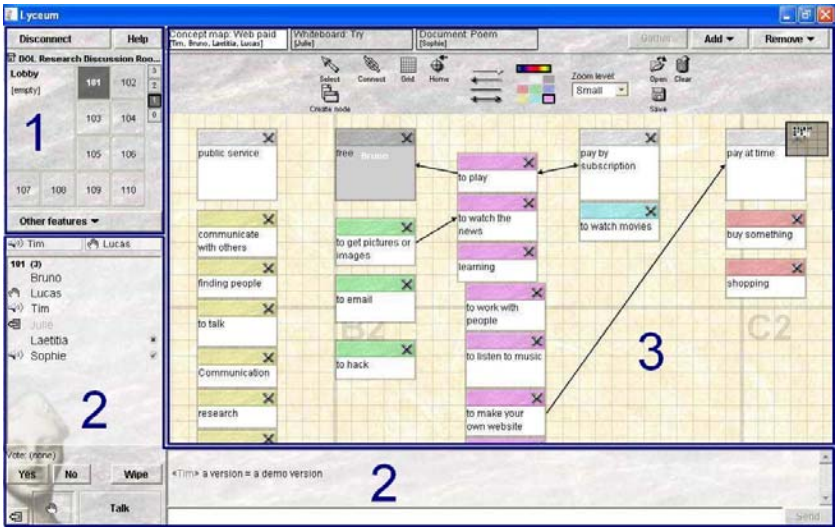


Figure 1 : Interface de Lyceum

Chacun peut, à tout instant, parler en activant 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 : explications de Tim sur "version").

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. Chaque module est visualisable (indépendamment) par chacun grâce aux onglets de la frise supérieure du cadre 3 de la figure 1. 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 auparavant dans le module, et bien sûr, créer, éditer, ou supprimer les objets propres à chaque type de module (ex : dans le Traitement de texte, les objets "paragraphes" ; dans la Carte conceptuelle, les concepts et relations ; dans le Tableau blanc, les traits, formes, textes, etc.).

Le potentiel et la souplesse d'utilisation de ce type d'environnements (Vetter, 2004), constituent un défi pour l'analyse, de par sa complexité, en termes de modes d'interaction et de production, à un double niveau, celui de l'individu et du groupe.

2.2. La multimodalité dans Lyceum

L'environnement audio graphique synchrone proposé par Lyceum permet aux acteurs de recourir à une variété de modes pour communiquer : les modes textuel, parole, graphique, iconique, auxquels s'ajoute un mode spatial qui correspond à la localisation et au déplacement des participants dans les différentes salles et modules de la plateforme (cf. section 2.1.2.). À chaque mode communicationnel correspond une modalité, c'est-à-dire une forme concrète particulière de communication, comme l'illustre le tableau 1.

Modes	**Modalités**
Textuel	Clavardage, traitement de texte, carte conceptuelle, tableau blanc
Parole	Audio
Graphique	Carte conceptuelle, tableau blanc
Iconique	Vote, entrée/sortie, absence momentanée, main levée, prise de parole
Spatial	Localisation (salle + module)

Tableau 1 : Correspondances entre modes et modalités dans Lyceum

La multimodalité se décline sous différentes formes. À un mode peut correspondre une variété de modalités, comme par exemple dans le cas du mode *textuel* qui apparaît dans la modalité *clavardage*, la modalité *traitement de texte*, la modalité *carte conceptuelle* ou encore *tableau blanc*. À certains modes peut ne correspondre qu'une seule modalité, comme dans le cas du mode *parole* qui ne correspond qu'à la modalité *audio*. Enfin, certains modules de production collaborative (carte conceptuelle, tableau blanc) peuvent intégrer plusieurs modes, comme dans le cas de la *carte conceptuelle* avec les modes *textuel* et *graphique* (Chanier et Vetter, 2006). On parlera ici de *communication multimodale* quand elle fait intervenir plusieurs modes, ou quand plusieurs modalités sont associées à un seul mode.

Lyceum propose donc aux acteurs d'interagir à partir d'un ensemble de systèmes sémio-linguistiques particulièrement riche et dont la potentialité, dans le cadre d'un apprentissage de langues par exemple, a donné lieu à diverses études. Ces études ont permis de mettre en lumière, d'une part, l'intégration de la multimodalité dans des scénarii pédagogiques (e.g : (Hampel, 2006)), selon la compréhension qu'ont les participants des fonctions des différents modes et modalités, et, d'autre part, l'organisation de la multimodalité à des fins communicatives (Chanier et Vetter, 2006), (Lamy, 2006). Deux grands traits ressortent de ces travaux : (1) l'organisation hiérarchique de la multimodalité dans les environnements audio graphique synchrones- le mode Parole étant prépondérant par rapport aux autres modes, en termes d'actes réalisés ; (2) la complémentarité forte des modalités audio et clavardage permettant toutes deux de soutenir la communication orale (Vetter et Chanier, 2006).

2.3. Analyser les actions multimodales synchrones

Notre approche scientifique s'inscrit à la suite des travaux de Baldry et Thibault (Baldry et Thibault, 2006) et vise à définir une méthodologie adaptée à l'analyse de la multimodalité. Depuis les années 90, de nombreux travaux (e.g. : (Levine et Scollon, 2004)), notamment en analyse du discours, définissent la multimodalité comme un processus dynamique de construction de sens, indissociable de la notion d'interaction. La multimodalité évolue donc au fur et à mesure de la communication et ne peut être appréhendée comme un texte[3] composite fini. Cela pose question quant à la méthodologie nécessaire à son analyse, dans la mesure où l'étude de chaque unité de la multimodalité *per se* (par ex. l'étude de tout le clavardage, puis tout le traitement de texte, puis tout l'audio) ne permettrait pas de rendre compte du principe d'intégration par lequel les acteurs structurent leur communication.

En outre, la notion *d'interaction multimodale synchrone* questionne les façons dont les acteurs vont travailler et communiquer. La juxtaposition des différents composants (cf. section 2.1.2) implique-t-elle une dispersion de l'attention des apprenants ou invite-t-elle à un comportement polyfocal, intégrant les éléments de communication développés dans des modes différents, et ce dans un même fil de conversation ? Comme le souligne Jones (Jones, 2004), la notion de *polyfocalisation* est cruciale pour comprendre le déroulement de la communication multimodale et déterminer la nature de l'engagement de l'acteur dans les différentes tâches effectuées. Cette notion est d'autant plus importante dans l'environnement étudié que les acteurs peuvent réaliser des actions différentes de façon simultanée, soit dans un même contexte (une même salle, une même tâche), soit dans des contextes différents (des salles, des mo-

3 Nous employons ici "texte" dans un sens très large (Halliday, 1989).

dules, des micro-tâches différentes), pour accomplir une même macro-tâche. Parce qu'il y a synchronie, un même acteur peut également utiliser différentes modalités de façon quasi simultanée pour communiquer (parler et lire en même temps le traitement de texte qui apparaît dans le clavardage, parler et déplacer des objets dans la carte conceptuelle, etc.).

La diversité des actions multimodales questionne la façon de représenter ces phénomènes. Il s'agit donc de définir un code de transcription pour des données hétérogènes (mode verbal : oral et scriptural et mode non-verbal) et de proposer un modèle structuré qui donne un cadre aux futures analyses. En effet, dans Lyceum, cette notion est indispensable car 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. Il est donc important de permettre aux chercheurs de restituer le *contexte*[4] dans lequel ont lieu les interactions pour pouvoir les analyser.

3. Méthodologie

Le dispositif de formation couplé au protocole de recueil des données, et les conventions de transcription influent sur les analyses qui pourront être menées à l'issue de la formation. Le fait "d'avoir la main" sur la plateforme permet de tracer les informations souhaitées (cf. par exemple, (Avouris *et al.*, 2007)). Dans le cas contraire (comme pour le projet Copéas), le protocole de recueil des données permet de pallier les possibles lacunes de la plateforme.

Les conventions de transcription influent sur le grain d'analyse car, selon le niveau de granularité adopté pour les transcriptions, certaines analyses seront possibles, d'autres non. Tous ces choix découlent évidemment des questions de recherche liées au projet. Dans ce projet, celles-ci s'orientent selon deux axes principaux. Le premier concerne l'association de l'audio aux autres modalités, et les spécificités de l'interaction multimodale à distance. Le second a trait au positionnement et à l'instrumentation des activités du tuteur. Ces deux axes nous ont amenés à effectuer les choix suivants.

3.1. Le 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. Le protocole de recueil inclut les enregistrements audio et vidéo,

4 Le contexte étant compris ici comme une entité dynamique définie dans l'interaction par les acteurs (Goffman, 1974), (Goodwin et Duranti, 1992).

les enregistrements des productions individuelles et collectives ainsi que 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 sessions et toutes les salles utilisées ont donc été sauvegardées. Ce choix de protocole exclut par exemple des analyses sur la gestuelle des acteurs (comme par exemple (Smith et Gorsuch, 2004)). En effet, nous avons décidé de ne pas enregistrer les acteurs dans leur environnement quotidien car cela ne correspondait pas à nos questions de recherche. Nous avons également récupéré des logs (connexions à la plateforme) des serveurs de l'Open University et les interactions de clavardage.

À l'issue de l'expérimentation nous avons élaboré et administré un post-questionnaire à l'ensemble des acteurs (étudiants et tuteurs), conduits des entretiens semi-dirigés et des auto-confrontations (*Critical Event Recall*), à partir d'extraits vidéo sélectionnés, auprès d'un plus petit nombre d'apprenants.

Pour ne citer que quelques chiffres et donner un aperçu de la taille du corpus traité, l'ensemble des données compte 37 vidéogrammes d'une durée cumulée de 27 heures, 512 fichiers (productions, audiogrammes des entretiens, questionnaires, etc.) et occupe 35 Go.

3.2. Les conventions de transcription

Les conventions de transcription adoptées induisent le niveau du grain d'analyse. Ce qui est transcrit pourra en effet plus facilement faire l'objet d'une analyse. Par ailleurs, si l'on veut rendre compte des actions dans des situations collectives, il faut être capable de caractériser pour chaque action, outre sa signification : qui la réalise, qui peut la voir, et donc où et quand elle a lieu.

Pour cela, au delà de la différenciation des groupes, l'unité première de « découpage » des enregistrements vidéo est la notion de *session* (8 sessions par groupe). Ensuite chaque session est composée d'*espace-temps* qui caractérisent le lieu et la date/heure d'une action. 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. Dans Lyceum, cette notion est indispensable car 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. Nous avons, en parallèle, divisé chaque session en *séquences pédagogiques*. Ce découpage identifie de manière plus précise mais subjective les grandes

activités d'une session par exemple : salutations, présentation des consignes, travail en sous-groupes, restitution du travail, *débriefing*, bilan. Nous proposons donc deux points de vue, l'un pédagogique et l'autre plus objectif, pour une transcription de la co-présence d'acteurs dans un lieu et durant une période donnée.

Enfin, chaque acte est caractérisé systématiquement par une modalité (audio, vote, clavardage, module de production, etc.), une valeur (ce qui a été dit, écrit ou fait), une date de début et éventuellement une date de fin, l'acteur ayant réalisé cet acte, l'espace-temps dans lequel il a eu lieu et la séquence à laquelle il est rattaché. Les tours de parole audio et les actions de production ont des attributs supplémentaires permettant d'affiner la description de leur réalisation dans ces modes spécifiques.

Concernant la parole, les caractéristiques de l'oral telles qu'elles apparaissent dans Lyceum sont fortement conditionnées par la représentation de l'oral des concepteurs de la plate-forme. Ainsi, des boutons et icônes en rapport permettant aux acteurs de communiquer à plusieurs à distance, pour éviter sans l'interdire le chevauchement qui rend la communication difficile. Les acteurs peuvent ainsi, comme nous l'avons illustré en section 2.1.2, demander la parole, indiquer qu'ils parlent, acquiescer ou refuser par les boutons de vote assurant ainsi la bonne qualité de la communication (Lamy, 2006), (Hampel, 2006). Ces différents éléments donnent lieu à un codage dans la transcription. Au besoin quelques éléments de phonétique ont été transcrits, notamment lorsque la prononciation a joué un rôle dans l'interprétation du message, cette transcription phonétique n'étant donc pas systématique. Nous avons également choisi de coder les silences pour pouvoir les quantifier ou savoir qui reprend la parole après une longue pause. Ainsi dès qu'un silence excède 3 secondes, il est alors considéré comme un tour de parole à part entière attribué à l'"acteur" silence, afin de pouvoir lui donner une place à part entière dans la distribution des tours de parole. Les recouvrements sont finalement peu présents à l'audio et sont majoritairement multimodaux (par exemple un acteur intervient dans le clavardage pendant qu'un autre parle). La date/heure des actions permet cependant de rendre compte du chevauchement entre les actions.

Concernant des actions effectuées dans un module de production collaborative nous n'avions pas de cadre scientifique pour les transcrire. La difficulté principale pour ce type de transcription est de définir le grain de caractérisation de l'action. Par exemple pour un module de traitement de texte quel est l'objet ou l'unité manipulé : l'ensemble du texte, le paragraphe, la ligne le mot ou le caractère ? Tout dépend de l'analyse qu'on veut faire par la suite et de l'utilisation de la transcription. Dans notre cas, l'utilisation de la vidéo permet d'analyser très finement les actions effectuées, les transcriptions n'ont donc pas pour objectif de représenter l'action au niveau le plus fin possible. Pour le traitement de texte nous avons utilisé comme objet la notion de

paragraphe de Lyceum correspondant à une ou deux lignes. Pour le module de carte conceptuelle les types d'objets manipulables sont le concept et la relation et pour le tableau blanc : le trait, la zone de texte, l'ellipse et la punaise. La valeur de l'action correspond à une des quatre activités suivantes : créer, éditer, supprimer ou sélectionner. Au type d'activité "éditer" nous associons une étiquette de fond ou de forme.

3.3. La synchronisation des différentes sources de traces

La figure 2 présente la configuration de connexion des différentes machines utilisées lors des sessions pédagogiques. Une horloge est représentée sur chaque machine susceptible de fournir des traces ou enregistrements des actions de la session. Nous pouvons constater que les différentes unités utilisées par le serveur ont chacune une horloge autonome. La machine de chaque chercheur ayant enregistré les clavardages et la vidéo d'écran des espaces-temps a aussi sa propre horloge. Ainsi, lors du recueil des traces, les estampilles temporelles des actions ne se réfèrent pas à un temps universel : elles ne sont pas synchronisées. En utilisant la redondance de certaines traces, nous avons pu déterminer le décalage existant entre les différentes machines.

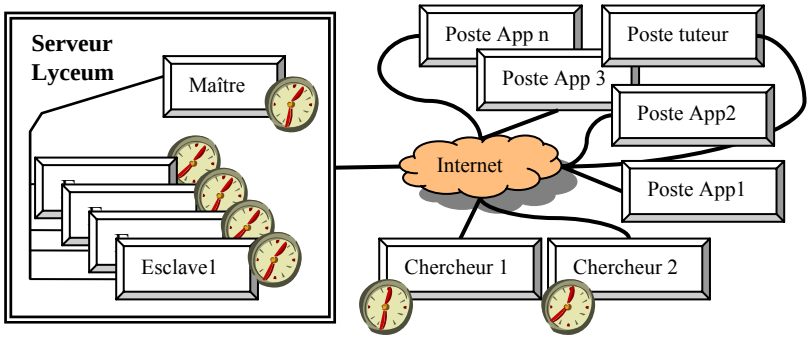


Figure 2 : Connexion des différentes machines Serveurs et Clients

Cette contrainte dans la précision du temps est extrêmement forte dans le cas d'interactions synchrones puisque deux actions consécutives peuvent être séparées de moins d'une seconde, tandis que les horloges étaient désynchronisées de plusieurs minutes.

Lors de la transcription des vidéos d'écran, la date (hh:mm:ss) de chaque action a été donnée dans le référentiel local de la durée de la vidéo donné par le lecteur vidéo ; la date zéro étant le début de l'enregistrement de la vidéo. Il a fallu définir ensuite avec précision, la date et heure (jj/mm/aa hh:mm:ss) du début de l'enregistrement vidéo dans le référentiel temporel de la machine du chercheur. Ainsi, la date de chaque action a pu être replacée dans un temps de référence choisi en appliquant un

décalage rigoureusement calculé. Tous ces calculs de synchronisation ont été faits en supposant que chaque horloge n'a subi aucun décalage au cours de l'expérimentation.

Par ailleurs, nous avons noté que l'exportation de dates et heures de Excel vers Access puis MySQL peut engendrer des petites imprécisions. Pour pallier les risques d'introduction d'imprécisions dans les valeurs temporelles nous opérons une vérification sur un échantillon, à chaque étape de collecte, transcription ou transformation, afin de contrôler qu'en particulier l'ordre des évènements n'est pas affecté.

4. Représentation des interactions

Avant de pouvoir analyser les interactions et tenter de comprendre les phénomènes qui se sont déroulés, il est indispensable de travailler sur "l'organisation, la modélisation et la conceptualisation des traces d'activité, de leur représentation et de leur traitement" (Settouti *et al.*, 2006). Les transcriptions ont été enregistrées dans un premier temps à l'aide d'un tableur par soucis d'efficacité. Cette organisation des données limite cependant leur traitement. Nous avons donc modélisé l'ensemble des données recueillies ou transcrites (Betbeder *et al.,* 2006). Une organisation sous forme de base de données offre au chercheur l'avantage de bénéficier de toutes les données organisées dans un unique format facilitant ainsi la recherche d'informations et lui permettant d'effectuer des fouilles et traitements statistiques.

4.1. Base de données

Sans entrer dans les détails du schéma relationnel (cf. figure 3), ce modèle reprend les différentes notions définies dans les conventions de transcription : sessions, séquence, espace-temps. La table "Actions" est centrale, elle contient les principales caractéristiques d'une action définies dans la section 3.3. Chaque action est liée à un espace-temps ainsi qu'à une séquence. Des tables spécifiques ont été ajoutées pour donner d'autres précisions sur des actions : nous avons par exemple 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 aux actions de productions, permet d'enregistrer le type de module de production collaborative utilisé (tableau blanc, carte conceptuelle ou traitement de texte dans Lyceum), le type d'objet utilisé (lignes, rectangles, ellipses, textes, paragraphes, concepts ou relations), et le document (plusieurs documents d'édition collaborative pouvant être ouverts simultanément) ainsi que le type d'action effectuée (créer, éditer, supprimer, sélectionner, etc.). Enfin les actions sont liées au fichier source (vidéo) et à l'enregistrement du document pour les actions de productions. Cette organisation générale permet alors d'extraire aisément les actions de chaque acteur.

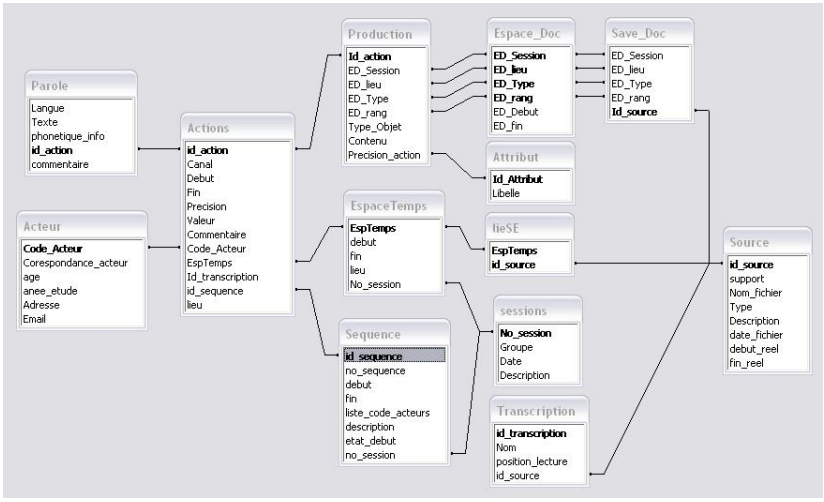


Figure 3 : Schéma relationnel de la base de données

L'interrogation de la base de données permet de calculer par exemple (cf. Tableau 2) la durée moyenne des actes de parole par acteur. On observe que les apprenants des 2 groupes, bien qu'ayant des compétences en langue très différentes, ont des moyennes similaires (min : 7 sec. max : 12 ou 13 sec., moyenne : 9 ou 10 sec.). Les tuteurs des deux groupes (Tim et Rob) ont exactement la même moyenne. Enfin les conventions définies nous permettent de quantifier et de calculer la moyenne des temps de silences entre deux interventions orales (faux débutants : 18 sec. et intermédiaires/avancés : 14 sec.).

Acteur groupe faux débutants	Durée moyenne des actes audio (mm:ss)	Acteur groupe niveau intermédiaire à avancé	Durée moyenne des actes audio (mm:ss)
al	00:12	agnes	00:10
amand	00:11	amel	00:08
chris	00:10	angel	00:11
ed	00:07	aur	00:07
ghis	00:08	bert	00:13
herve	00:07	isa	00:08
remy	00:07	rob	00:14
tim	00:14	sand	00:10
silence	00:18	silence	00:14

Tableau 2 : Durée moyenne des actes audio des acteurs

4.2. Outil de visualisation

Le protocole de recherche ainsi que nos transcriptions nous ont permis d'obtenir une base de données des actions des acteurs riche en informations. Néanmoins cette organisation ne donne pas une visualisation optimale des données utiles pour les analyses. En effet, la représentation offerte par des requêtes de la base ne permet pas la visualisation graphique de la durée des actions, de leurs chevauchements ou enchevêtrements. Les requêtes présentent les résultats "à plat" sans faire ressortir visuellement la durée, les modalités et les acteurs. Nous avons donc développé un outil couplé à la base de données pour améliorer la visualisation des données d'interactions multimodales : il permet, entre autres, de mettre en évidence la durée et la structuration des actions (cf. Figure 4).

Figure 4 : Copie d'écran de l'outil de visualisation, vue par modalité[5].

Ce prototype permet de sélectionner la session et l'espace-temps à visualiser. Deux vues sont proposées, une vue par acteur (chaque acteur est une donnée de l'axe des ordonnées, les modalités sont différenciés par des couleurs différentes) et une vue par modalité (cf. Figure 4), l'axe des abscisses représente le temps. Le graphe généré représente l'intégralité de l'espace-temps que l'on peut "zoomer" et parcourir le long de l'axe temporel.

Dans cette figure, on visualise aisément que l'acteur chris effectue une action de production pendant l'acte de parole de l'acteur tim (le tuteur). Cet outil est au stade de

5 Signification des sigles : As : arrivée/sortie, prod : acte de production, tpa : tour de parole audio, tpc : tour de parole clavardage et v : vote.

prototype, nous avons prévu de développer une version stable courant 2008 et de l'intégrer au sein de la plateforme de partage de corpus du projet Mulce pour visualiser un corpus déposé.

5. Exemples d'analyses d'interactions multimodales

5.1. Analyse macroscopique

L'interrogation de la base de données (cf. section 4.1) nous permet d'effectuer une analyse longitudinale de ce qui s'est déroulé au cours de l'apprentissage. Nous cherchons à dégager des tendances générales en procédant à des calculs statistiques qui combinent les sessions, les acteurs, les actions et les modalités. Le but est également de repérer des écarts par rapport à ces tendances pour mettre en évidence des "comportements" singuliers et tenter de les interpréter.

L'objectif de cette section est de proposer une illustration des résultats obtenus par l'exploitation de la base de données et des analyses. Nous ne prétendons pas faire une analyse exhaustive des interactions multimodales de cette formation.

5.1.1. Une première approche de l'activité

La base de données nous permet d'évaluer l'activité de chaque acteur du groupe des faux débutants en calculant les actes des trois modalités principales : audio, clavardage et production (traitement de texte, carte conceptuelle et tableau blanc). La figure 5 présente ainsi le nombre total d'actes effectués par chacun pour les huit sessions. Cette première mesure un peu grossière sera ensuite couplée avec d'autres observations.

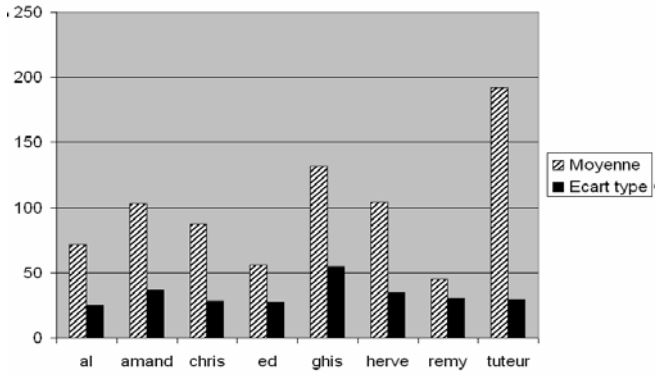


Figure 5 : Nombre moyen d'actes par session pour chaque acteur.
A noter : al a été absent à deux sessions et remy a abandonné après les quatre premières sessions**.**

Ce graphique montre que le tuteur exécute un nombre d'actes très important par rapport aux apprenants en accomplissant à lui seul 27% des actes. Par ailleurs il le fait de façon régulière sur les huit sessions (écart-type faible par rapport à sa moyenne). Parmi les apprenants ghis est le plus actif mais il n'agit pas de façon très régulière sur l'ensemble des sessions (écart-type important). Si l'on détaille l'activité par session et par acteur (cf. Tableau 3) on observe une "sur-activité" pour ghis en session 3 avec 220 actes (i.e. : 26,83% sur le total de la session) contre 181 pour le tuteur. ghis passe en session 4 à 10,04% des actes (49 actes, tuteur 144) : son taux le plus faible sur les huit sessions.

Code Acteur	S3	S4	S4-S3
al	abs	53	
amand	147	65	-82
chris	67	52	-15
ed	58	23	-35
ghis	220	49	-171
herve	127	73	-54
remy	20	29	9
tuteur	181	144	-37

Tableau 3 : Nombre d'actes par acteur en session 3 et 4

Ayant observé cette irrégularité on peut chercher à comprendre ce qui stimule ou freine l'apprenant dans son activité, et à interpréter ces écarts. Pour ce faire des informations complémentaires pour caractériser des différences contextuelles ou pédagogiques entre les sessions 3 et 4 sont nécessaires. On constate d'abord, comme le montre le tableau 3, une baisse générale du nombre d'actes entre les sessions 3 et 4 pour l'ensemble des acteurs. Un examen plus fin des données classées par type de production nous montre que ghis a effectué un nombre d'actes lié à la production très largement supérieur aux autres apprenants en session 3 puisqu'il couvre 50% des actes de production parmi les apprenants. Une analyse plus approfondie de ces actions montre que ghis effectue très majoritairement des éditions de forme (déplacement des objets sur les cartes conceptuelles) contribuant finalement assez peu à la tâche (sur le fond).

5.1.2. Espace-temps et dynamique de groupe

La notion d'espace-temps suscite quelques questions. Comment les acteurs se partagent-ils un espace-temps ? Les outils sont-ils utilisés indifféremment par tous ? Avec la même fréquence ? Comment s'entrelacent les interventions des acteurs ?

Les pourcentages d'actes par acteur (cf. Figure 6) montrent que herve (12,04%) et chris (14,36%) figurent parmi les apprenants actifs. De plus, d'après les écart-types leur activité est régulière sur l'ensemble des sessions. L'observation des courbes de l'activité par session montre des variations complémentaires entre ces deux apprenants avec un cumul des pourcentages des actes accomplis par herve et chris stable (en moyenne par session : 26,4%) et un écart-type faible (3,96).

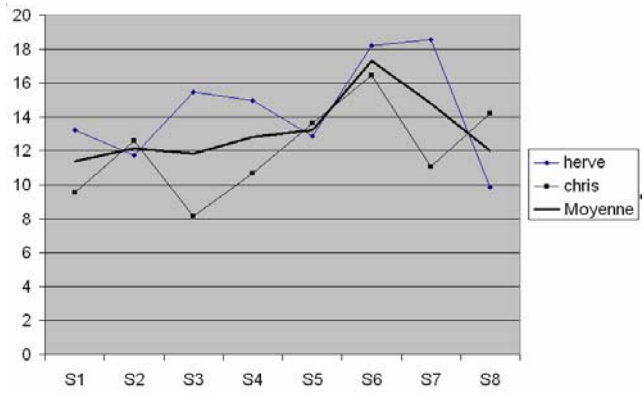


Figure 6 : Pourcentage des actes par session

Notons qu'en session 5 et en session 7 les apprenants ont été placés en sous-groupe, herve et chris restant ensemble. herve et chris couvrent environ un quart des actes à eux deux et d'une façon générale quand l'un diminue son activité, l'autre l'augmente. En terme de dynamique de groupe on peut s'interroger sur le partage d'un espace-temps et les jeux d'influences qui se dessinent entre les apprenants et avec le tuteur. Ce type de statistique en illustre une première approche.

5.1.3. Les actes de dialogue

Pour mener des analyses plus fines, nous avons choisi d'écarter les actes de production pour nous focaliser sur les actes de dialogue conformément au but de la formation qui vise à développer des compétences d'expression orale. Nous avons donc effectué des calculs en retenant les actes audio et les actes de clavardage. Nous cherchons à comprendre également comment s'articulent entre elles ces deux modalités souvent étudiées comme complémentaires (Chanier *et al.*, 2006), (Lamy, 2006).

La question de la participation et la façon de la caractériser est au cœur de cette étude. 878 actes de clavardages et 2570 actes audio ont été effectués. Le tableau 4 représente la taille moyenne (en nombre des caractères) et les variations d'un acte de clavardage ou audio (transcrit). Les écrits sont beaucoup plus courts que les paroles.

Mais pour les deux modalités l'écart-type est très important, cela signifie qu'il y a une dispersion forte dans les tailles des messages. Pour l'ensemble des acteurs du groupe, 52% des messages écrits comprennent de 1 à 15 caractères, 37% de 16 à 40 et enfin 11% plus de 41 caractères. Notons également que 48,5% des messages de plus de 41 caractères ont été rédigés par le tuteur.

	clavardage	audio
Moyenne	20,60	94,26
Min	1	2
Max	170	1473
Ecart-type	19,49	109,20

Tableau 4 : Nombre de caractères par acte de clavardage et audio

Concernant les actes audio, 60% des contenus (transcrits) ont une longueur comprise entre 2 et 80 caractères, 28% entre 80 et 200 caractères. La majorité des messages audio ont donc une longueur plus petite que 200 caractères et sont prononcés en une à deux secondes. Les contenus audio en anglais de plus de 500 caractères sont tous, excepté un, prononcés par le tuteur. Enfin le tuteur a énoncé les 3/4 des contenus audio d'une longueur comprise entre 300 et 539. Ces chiffres montrent que les messages audio ou clavardage les plus longs sont majoritairement accomplis par le tuteur et donne une idée des longueurs moyennement pratiquées dans les espaces-temps.

Le tableau 5 présente des statistiques sur la pratique du clavardage et de l'audio chez les apprenants. Ces chiffres nous montrent qu'il n'y a pas de régularité apparente chez les apprenants entre l'usage du clavardage et de l'audio. Ce n'est pas parce qu'un apprenant parle plus à l'audio qu'il utilise moins le clavardage et inversement.

	audio	audio	clavardage	clavardage
al	279	17,25%	13	1,98%
amand	202	12,49%	246	37,50%
chris	291	18,00%	80	12,20%
ed	170	10,51%	54	8,23%
ghis	368	22,76%	110	16,77%
herve	256	15,83%	132	20,12%
remy	51	3,15%	21	3,20%
total	1617	100,00%	656	100,00%

Tableau 5 : Nombre d'actes et pourcentages d'actes audio et de clavardage par apprenant

Notons un fait singulier chez amand qui a effectué plus d'actes de clavardage que d'audio. Cet apprenant rencontre t-il des difficultés à s'exprimer à l'oral ? Est-ce que le clavardage lui permet de s'effacer et/ou de s'exprimer dans une syntaxe moins rigoureuse qu'à l'oral ? Si l'on regarde la taille des messages, on observe que presque la moitié des messages (121 messages) comprennent un ou deux mots. Exemples : "me too", "is it", "yes". L'autre moitié présente une trentaine de messages qui inclut 3 à 4 mots. L'ensemble ne dépasse pas 69 caractères à l'exception d'un seul qui atteint 170 caractères.

Qu'en est-il de la pratique du tuteur ? Il a effectué 36,6% des actes audio d'une longueur moyenne de 104 caractères (écart-type : 75). Il a également réalisé 24,8% des clavardages d'une longueur moyenne de 29,3 caractères écart-type 21 caractères). Comment ses actes se répartissent-ils au fil des sessions ?

La figure 7 montre comment évolue dans le temps l'usage que fait le tuteur des outils audio et clavardage. Il est intéressant de noter que si le tuteur agit de façon régulière et constante notamment à partir de la session 4, il utilise l'audio et le clavardage de façon variable. En effet, s'il prend souvent la parole en début d'apprentissage, il diminue ensuite régulièrement le nombre de fois où il parle. À l'inverse il utilise de plus en plus le clavardage au fur et à mesure du déroulement des sessions.

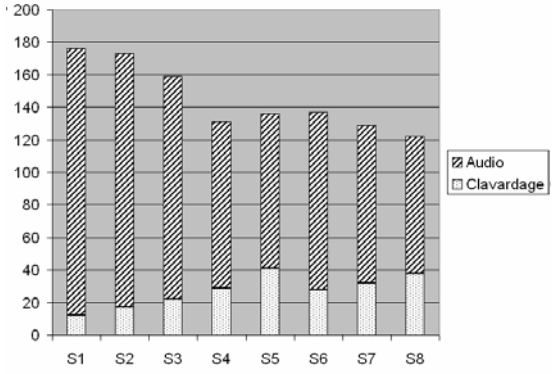


Figure 7 : Nombre d'actes audio et de clavardage du tuteur par session

Comment interpréter ces observations ? On comprend que le tuteur remplit sa mission qui consiste à développer une pratique de l'oral chez les apprenants puisqu'il laisse la parole aux apprenants de plus en plus souvent. Ce qui nous semble plus intéressant est d'observer que le tuteur augmente le nombre de fois où il utilise le clavardage au fur et à mesure qu'il diminue la fréquence de ses prises de parole. On peut interpréter cette observation en pensant que dans un espace-temps donné le tuteur

cherche à garder (consciemment ou pas) un taux de présence constant en communication langagière. S'il laisse plus la parole, il maintient le dialogue par le clavardage.

L'analyse présentée dans cet article illustre le type d'indicateurs relevant d'une vue longitudinale de l'apprentissage que l'on peut extraire des transcriptions, elle montre comment rendre-compte statistiquement des pratiques multimodales et de l'activité des participants. Elle peut aussi permettre d'établir des seuils en deçà desquels l'activité témoignerait d'une participation insuffisante ou d'autres au-delà desquels, elle serait jugée excessive (suractivité d'un participant empêchant les autres de trouver leur place) : autant d'indicateurs qui pourraient alerter le tuteur (qui n'est pas nécessairement présent à chaque instant dans chaque lieu).

5.2. Analyse microscopique

L'analyse microscopique de la multimodalité va permettre de mettre en évidence non plus des tendances générales et longitudinales, mais plutôt des phénomènes singuliers, qui apparaissent à un moment donné, dans un bref espace-temps. Elle permet donc de révéler certaines caractéristiques liées à la synchronie des interactions multimodales. La notion de *contexte* est ici essentielle pour comprendre ces phénomènes : l'utilisation de la multimodalité par les acteurs diffère en effet selon la nature de la tâche à accomplir, selon les modalités de travail préconisées dans le groupe ou les sous-groupes, et selon les outils de production collective utilisés (traitement de texte, carte conceptuelle, tableau blanc).

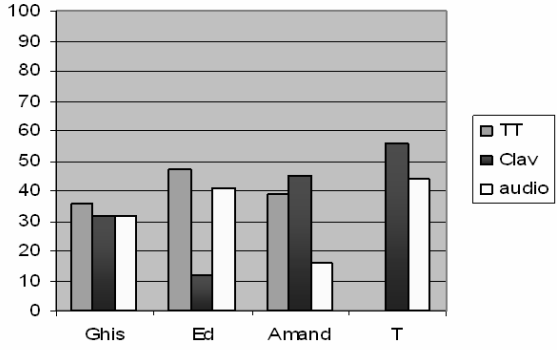
Nous présentons ici diverses analyses mettant en relation trois modalités- l'audio (mode parole), le traitement de texte (TT) et le clavardage (mode textuel). Nous nous intéressons à la façon dont les acteurs (apprenants et tuteur) utilisent ces trois modalités pour produire collectivement un texte, tant au niveau des *pratiques individuelles* (quand et qui utilise quoi ? Dans quelle proportion et pour quel volume ?) qu'au niveau des *pratiques collaboratives* (quelle répartition de l'utilisation de la multimodalité pour quel type de collaboration ?). L'utilisation de la base de données (cf. section 4.1) permet de repérer les espaces-temps dans lesquels les modalités audio, traitement de texte et clavardage sont en co-présence, pour, ensuite, sélectionner parmi ces derniers ceux présentant le plus d'actes de production dans le traitement de texte. L'outil de visualisation (cf. section 4.2), permettant de représenter par acteur l'orchestration de la multimodalité pour un espace-temps, donne à lire les mises en correspondance entre les acteurs et les modes utilisés dans la réalisation collaborative d'une tâche. Nous présentons dans cette partie les premiers résultats d'analyse obtenus à l'aide de ces outils, pour discuter ensuite de leur pertinence et limites pour une analyse microscopique de la multimodalité.

5.2.1. Répondre à un questionnaire en sous-groupe

Notre analyse porte sur l'espace-temps E4T5[6], dans lequel le groupe d'apprenants est divisé en deux sous-groupes de trois apprenants chacun (sous-groupe A et sous-groupe B), travaillant dans deux salles séparées (s.101 et s.102). Le tuteur se déplace de salle en salle pour soutenir le travail en parallèle des deux sous-groupes. Dans cette séquence, chaque sous-groupe travaille à la réalisation d'une même tâche : répondre collaborativement à un questionnaire préparé par le tuteur sur la notion d'interactivité à partir de la consultation d'un site internet fléché par le tuteur. Le questionnaire est mis à disposition des apprenants dans le traitement de texte intégré, les apprenants pouvant ainsi répondre simultanément ou successivement aux différentes questions. Chaque acte d'écriture est signalé aux autres participants par l'inscription en début de ligne des initiales du scripteur. Chaque apprenant peut ainsi voir les réponses apportées au fur et à mesure du processus d'écriture. Cette caractéristique différencie la modalité traitement de texte de celle du clavardage, dans la mesure où l'écrit du clavardage n'est lisible qu'une fois terminé. L'écrit dans le traitement de texte a un degré d'interactivité supérieur à celui du clavardage et conditionne ainsi d'une certaine manière la nature des interactions entre acteurs. L'activité en sous-groupe dure une trentaine de minutes.

5.2.2. La multimodalité dans le sous-groupe A

Pour répondre aux questions posées par écrit sur l'interactivité, le sous-groupe A (tuteur inclus) va réaliser 79 actes qui se répartissent selon les modalités suivantes : 25 actes de parole, 31 actes de clavardage et 23 actes de production dans le traitement de texte. Ces trois modalités ont été utilisées de façon équilibrée pour réaliser la tâche demandée. Toutefois, une analyse des pratiques individuelles montre des différences entre les participants (cf. Figure 8).



6 E4T5 : 4ème espace-temps de la session 5 du groupe des faux-débutants

Figure 8 : Répartition en pourcentage de la multimodalité (Traitement de texte, clavardage, audio) par acteur dans le sous-groupe A.

Si l'apprenant ghis a recours de façon similaire aux trois modalités pour participer à la réalisation de la tâche d'écriture, les apprenants ed et amand montrent, au contraire, une utilisation plus nettement différenciée. ed privilégie l'utilisation des modalités TT et audio au détriment du clavardage, tandis que amand privilégie les modalités TT et clavardage (avec une prépondérance pour le mode textuel), au détriment de la modalité audio. On retrouve chez ed et amand une complémentarité entre les modalités clavardage et audio qui est inversement proportionnelle chez les deux apprenants. Cette focalisation sur les modalités par acteur donne une indication sur la façon dont les acteurs dans un contexte donné perçoivent la fonction de chacun des modes convoqués ; cette indication pouvant être ensuite approfondie par la lecture de la transcription des contenus des échanges.

L'analyse croisée des modalités utilisées et des contenus des interventions (en recourant à la base de données, une fois les données alignées) permet de déterminer les modalités de collaboration du sous-groupe. La similarité des rôles de scripteur et de correcteur chez les trois apprenants et la simultanéité des actions d'écriture dans le traitement de texte révèlent un modèle de collaboration plutôt coopératif où chacun contribue, séparément, à la réalisation d'une tâche commune. L'outil de visualisation permet également de repérer ce phénomène : la visualisation par modalités pour l'espace-temps étudié donne à lire la concentration modale des acteurs au fur à mesure de la réalisation de la tâche (forte concentration dans le TT avec ponctuellement recours à l'audio et au clavardage, ce qui indique une négociation a posteriori entre les apprenants sur leur production à plusieurs mains).

5.2.3. La multimodalité dans le sous-groupe B

Pour répondre aux questions sur l'interactivité, le sous-groupe B va réaliser 85 actes dont 62 à l'audio, 6 dans le clavardage et 17 dans le traitement de texte. Contrairement au sous-groupe A, l'utilisation des trois modalités (TT, clavardage et audio) n'est pas ici équilibrée. Elle démontre plutôt une forte différenciation entre les modalités en faveur de l'audio. L'analyse des pratiques individuelles met au jour une répartition singulière des rôles et des modalités (cf. Figure 9).

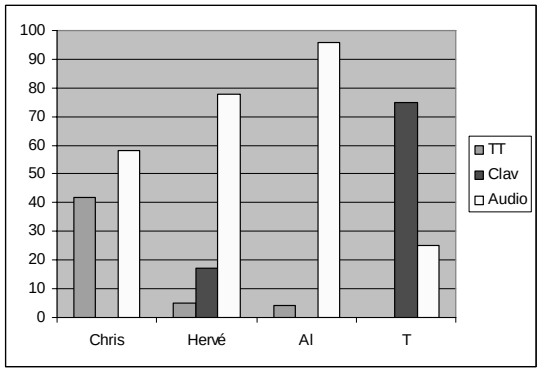
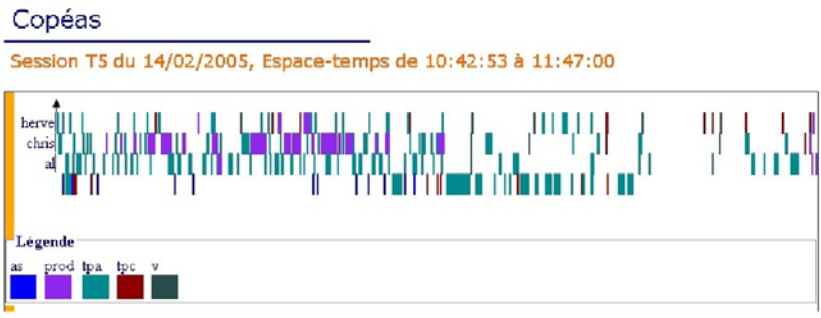


Figure 9 : Répartition en pourcentage de la multimodalité (Traitement de texte, clavardage, audio) par acteur dans le sous-groupe B.

Le calcul en pourcentage de la répartition des modalités selon les acteurs montre une forte utilisation de l'audio chez les apprenants au détriment du clavardage (cette modalité est d'ailleurs absente ou quasi absente chez les apprenants dans l'espace-temps étudié). La correspondance entre les modalités clavardage et audio est ici très différente chez le tuteur qui privilégie le clavardage (comme dans le cas du sous-groupe A) et va nettement moins utiliser l'audio (25 % au lieu des 40 % lors de sa participation au sous-groupe A). L'utilisation que les apprenants font de la modalité TT est très différente de celle observée dans le sous-groupe A puisqu'un seul apprenant (chris) l'utilise. Cela tend à montrer que les apprenants ont eu des rôles différents et qu'ils n'ont pas tous été scripteurs, même s'ils en avaient techniquement la possibilité. Le sous-groupe B a semble-t-il privilégié la bimodalité audio/TT pour accomplir la tâche demandée. La visualisation par acteur pour l'espace-temps étudié corrobore cette analyse et montre une participation beaucoup plus grande par chris dans la modalité TT qui succède et précède des actes à l'audio chez les deux autres apprenants (cf. Figure 10).



Légende : as : arrivée/sortie, prod : production, tpa : tour de parole audio, tpc : tour de parole clavardage, v : vote

Figure 10 : Copie d'écran de l'outil de visualisation, vue par acteur pour l'espace-temps donné.

L'analyse des actions et de leurs contenus permet d'aller plus loin dans la compréhension du modèle de collaboration ici adopté. L'audio est la modalité privilégiée pour négocier les rôles et le contenu de ce qui va ensuite être écrit dans le traitement de texte. Pendant que l'un écrit les réponses discutées à l'audio, les deux autres occupent un rôle de correcteur (parfois à l'écrit, souvent à l'audio). Le modèle est ici plutôt collaboratif dans la mesure où tous contribuent ensemble en se focalisant sur la même action en même temps à la réalisation de la tâche. Au contraire, dans le sous-groupe A, les apprenants ont réalisé en parallèle des actions similaires (répondre aux différentes questions dans le TT) pour ensuite en discuter ensemble ponctuellement à l'audio.

5.2.4. Multimodalité, intramodalité et intermodalité

L'analyse microscopique de l'organisation de la multimodalité selon un contexte donné (une tâche, un outil de production collaborative, un groupe) rend possible une compréhension fine des enchevêtrements des différentes modalités convoquées par les acteurs. Elle permet de repérer des moments où l'organisation du discours est plutôt *intramodal* (i.e. la plupart des interactions qui conditionnent le fil du discours et l'accomplissement de la tâche se réalisent dans un seul mode). Cela signifie également que les acteurs vont davantage porter leur attention sur une seule modalité à ce moment de l'interaction.

L'analyse de ces interactions fortement intramodales sur l'ensemble du corpus permet de repérer deux facteurs explicatifs :

- *les stratégies individuelles.* Les apprenants choisissent pour communiquer et travailler d'utiliser de manière privilégiée une modalité au détriment des autres. C'est le cas d'amand qui va systématiquement utiliser le clavardage au détriment de l'audio, comme nous l'avons vu précédemment ;
- *l'organisation naturelle du discours* dans laquelle à un acte de communication dans un mode donné va répondre un acte de communication dans le même mode. Le sous-groupe A a ainsi privilégié une organisation plutôt intramodal de la multimodalité pour répondre au questionnaire, d'abord en travaillant dans le traitement de texte, puis à l'audio. Ces différents exemples montrent des "densités modales" (Norris, 2004) différentes, selon la compréhension que les acteurs ont de la situation dans laquelle ils interagissent.

Plus ponctuellement, la multimodalité peut s'organiser entre différentes modalités pour accomplir un même fil de discours, c'est ce que nous nommons ici l'*intermodalité*. Elle apparaît dans le corpus principalement lors des travaux en sous-groupe. Il s'agit principalement dans le corpus d'une organisation bimodale (audio et TT), à l'image de la modalité de travail privilégiée par le sous-groupe B dans laquelle un apprenant initiait le fil de discours dans une modalité (audio) puis un autre apprenant continuait ce même fil de discours dans une autre modalité (TT). Il apparaît ainsi que quelle que soit la forme que prenne la multimodalité dans Lyceum celle-ci réponde toujours à des règles de structuration définies plus ou moins explicitement dans le groupe. Chaque sous-groupe montre ainsi la mise en place de normes de communication qui lui sont propres et que les individus négocient au fur et à mesure de leur participation et de leur propre appropriation de l'environnement, détournant parfois les consignes d'utilisation qui leur étaient proposées (cf. sur ce point (Lamy, 2007)), comme nous avons pu le constater en analysant les modalités de travail des différentes configurations des groupes dans le corpus.

L'analyse microscopique des interactions multimodales met en lumière le rôle prépondérant que joue la notion de *contexte* dans la définition et la compréhension des actions multimodales. Comme nous l'avons vu, ce sont les acteurs qui décident et négocient entre eux par leurs interactions la forme et la signification qu'ils donnent à la multimodalité. La notion du point de vue de l'acteur est donc à prendre en compte dans la méthodologie mise en œuvre pour l'étudier. D'une façon plus générale, les analyses microscopiques montrent que la multimodalité gagnent à être étudiée comme un processus dynamique et plaident pour la recherche de cadres d'analyse qui prennent l'action comme fil directeur.

5.3. Bilan

Les différentes représentations présentées dans la section précédente, basée sur les données recueillies et transcrites nous permettent d'effectuer des analyses d'une part statistiques et macroscopiques telles que vu par exemple dans la section 5.1 et plus fines telles que celles présentées en section 5.2. Elles font ressortir la difficulté d'interprétation des traces, champ largement étudié (par exemple, Avouris *et al.*, 2005), (Courtin et Talbot, 2006), (Heraud *et al.*, 2005)).

Ces représentations fournissent une aide à l'analyse, même si bien souvent l'accès au contexte de formation ou au contenu des actes de paroles sont indispensables. Il est également intéressant de coupler ou croiser les différentes représentations ou outils et analyses. Les modes de représentations, base de données et outil de visualisation, permettent d'isoler et de repérer des tendances telles que l'utilisation privilégiée par un acteur d'une modalité de collaboration ou des règles de communication multimo-

dales au sein de sous-groupes qui peuvent donner lieu par la suite à des analyses microscopiques.

Dans le cadre de ce projet nous avons également développé un outil de recherche de patterns (Betbeder *et al.*, 2007) afin d'étudier quelles sont les séquences effectuées de manière récurrente. Cet outil basé sur l'algorithme Winnepi (Mannila *et al.*, 1997) permet notamment d'étudier les aspects multimodaux des interactions. Les questions sous-jacentes à ce travail étaient : Comment s'articule les interactions multimodales ? Peut-on repérer par exemple qui reprend la parole après un silence ? Qui initie une conversation ? Plus généralement, peut-on trouver des règles qui montrent comment les acteurs utilisent les différentes modalités pour leur travail collectif. Nous pensons que la recherche de patterns, appliquée à un grand volume d'interactions, peut faire ressortir des suites d'actions récurrentes qui peuvent aider le chercheur dans son analyse. Il peut également être utilisé pour confirmer ou infirmer la récurrence *d'éléments* étudiés dans une séquence précise.

Par exemple, dans la section 5.2., l'analyse microscopique révèle des pratiques d'utilisation en sous groupe des trois modalités. L'outil de recherche de patterns nous permet de noter que ces entrelacements de modalités ne sont apparus (de manière récurrente) qu'en séquence en sous-groupe. En effet, 75.6 % des patterns sont unimodaux (un acte audio de l'acteur x suivi d'un acte audio de l'acteur y implique un acte audio de l'acteur z), i.e. les acteurs privilégient la même modalité pour construire le discours. Le nombre de patterns bi-modaux (audio et production) est significativement plus important dans les 2 séquences en sous-groupe (et principalement dans le sous-groupe B). Les règles d'association obtenues dans (Betbeder *et al.*, 2007) montrent des interactions entre trois acteurs qui alternent entre audio et production. Par exemple : un acte audio de l'acteur Al suivi d'un acte de production de l'acteur Chris implique un acte audio de l'acteur Herve.

Le recours aux différents outils développés pour étudier de façon croisée la complexité des interactions multimodales d'apprentissage tend à plaider pour une méthodologie qui intégrerait cette variété d'outils, les limites de chaque analyse proposée ici (macroscopique et microscopique) trouvant souvent réponse par le changement d'échelle.

6. Conclusions et perspectives

La multimodalité est aujourd'hui un objet de recherche qui de par sa complexité permet de renouveler le questionnement sur l'analyse des traces et leur traitement. Ainsi, de nombreux travaux issus de différentes disciplines cherchent à définir un cadre méthodologique pertinent qui permette l'analyse et l'interprétation de données multimodales (Avouris *et al.*, 2005), (Baldry et Thibault, 2006), (Lamy, 2007).

Les tendances observées dans les analyses et leur interprétation sont pour nous indissociables du contexte pédagogique dans lequel les usages ont été effectués. Les usages caractérisés par "l'existence de groupes d'usagers ayant conscience d'appartenir à une communauté d'intérêts" (Bruillard et Baron, 2006) sont ainsi différents selon les groupes et les rôles des acteurs. On constate en effet que l'appropriation et l'utilisation de la multimodalité prennent des formes multiples selon les utilisateurs (individuel ou groupe), et ce même dans un contexte identique.

Les analyses et outils présentés dans cet article présentent des limites si l'on veut comprendre des phénomènes liés à la dynamique de groupe. L'analyse de processus de groupe demande par exemple de "mesurer" les chevauchements des actes, de calibrer si des échanges représentent des interactions plus ou moins fortes, etc. Ces questions nécessitent de disposer d'outils qui traitent des données temporelles ainsi que des outils pour la reconnaissance de forme : reconnaissance de motifs d'actions, reconnaissance de contenus d'échanges parlant du "même" sujet. Des outils de traitement automatique du langage nous permettraient d'évaluer la distance sémantique entre deux échanges (oraux et/ou écrits) pour mesurer si les acteurs parlent plus ou moins de la même chose, nous aideraient à différencier les degrés d'interactivités (du simple échange de politesse à la co-construction des connaissances), et d'affiner ainsi la compréhension de la structuration de la multimodalité.

Nos travaux sur les données d'interaction recoupent ceux de (Champin *et al.*, 2003) qui cherchent à définir une théorie de la trace en proposant une définition des systèmes à base de traces comme étant des "systèmes informatiques permettant et facilitant l'exploitation des traces". Le contexte, le protocole de recherche et les interactions présentées ici (issues du projet Copéas) constituent l'un des ensembles de données disponibles dans le projet Mulce. La plateforme développée au terme de ce projet vise à proposer un ensemble de corpus échangeables et intégrera des outils d'analyse et de visualisation. Le schéma opératoire : collecter, transformer, analyser et visualiser est de ce fait intégré au projet. De cette façon les traces natives transformées et contextualisées deviennent réellement pertinentes pour l'analyse.

Remerciements

Mulce est un projet soutenu par l'ANR (ANR-06-CORP-006). Il rassemble des équipes des laboratoires LASELDI et LIFC (Université de Franche-Comté), CREET (The Open University) et LIP6 (Université Paris 6), coordonnées respectivement par Thierry Chanier, Christophe Reffay, Marie-Noelle Lamy et Jean-Gabriel Ganascia.

Nous remercions N. Azizi et A. Dami pour le développement de l'outil de visualisation.

BIBLIOGRAPHIE

Tous les liens ci-dessous ont été vérifiés le 20 janvier 2008.

(Avouris *et al.*, 2005)

AVOURIS, N., KOMIS, V. FIOTAKIS, G., MARGARITIS, M., VOYIATZAKI, E. (2005). Logging of fingertip actions is not enough for analysis of learning activities, *Actes du Workshop Usage Analysis in learning systems, AIED 2005*, Amsterdam**,** Pays-Bas, juillet 2005.

(Avouris *et al.*, 2007)

AVOURIS N., FIOTAKIS G., KAHRIMANIS G., MARGARITIS M., KOMIS V. (2007). Beyond logging of fingertip actions: analysis of collaborative learning using multiple sources of data, *Journal of Interactive Learning Research JILR*, Special Issue: Usage Analysis in Learning Systems: Existing Approaches and Scientific Issues, Vol. 18 (2), p.231-250.

(Baldry et Thibault, 2006)

BALDRY A., THIBAULT P. (2006). *Multimodal Transcription and Text Analysis, a multi-media toolkit and coursebook with associated on-line course.* London, Equinox.

(Betbeder *et al.*, 2006)

BETBEDER M.-L., REFFAY C., CHANIER T. (2006). Environnement audio graphique synchrone : recueil et transcription pour l'analyse des interactions multimodales, *Actes de JOCAIR 2006*, Amiens, 6-7 juillet 2006, p. 406-420. http://edutice.archives-ouvertes.fr/edutice-00085646

(Betbeder *et al.*, 2006)

BETBEDER M.-L., TISSOT R, REFFAY C. (2007). Recherche de patterns dans un corpus d'actions multimodales, *Actes de la conférence EIAH*, Lausanne, Suisse, 27-29 juin 2007, p.533-544. http://edutice.archives-ouvertes.fr/edutice-00158881

(Bruillard et Baron, 2006)

BRUILLARD E, BARON G -.L. (2006). Usages en milieu scolaire : caractérisation, observation et évaluation. In Grandbastien M. et Labat J.-M. (dirs), *Environnements informatiques pour l'apprentissage humain*, col. Hermès Lavoisier, p. 270-284.

(Champin et al., 2003)

CHAMPIN P.-A., PRIÉ Y., MILLE A., (2003). MUSETTE: Modelling USEs and Tasks for Tracing Experience. *ICCBR'03: Workshop ''From structured cases to unstructured problem solving episodes''* ICCBR'03: NTNU, p. 279–286.

(Chanier *et al.*, 2006)

CHANIER, T., VETTER, A., BETBEDER, M.-L., REFFAY, C. (2006). Retrouver le chemin de la parole en environnement audio-graphique synchrone. Dans Dejean-Thircuir, C. & Mangenot, F (Dir.), "Les échanges en ligne dans l'apprentissage et la formation", *Le Français dans le Monde, Recherche et applications* (40), p.139-150. http://edutice.archives-ouvertes.fr/edutice-00084388

(Chanier et Vetter, 2006)

CHANIER T., VETTER A. (2006). Multimodalité et expression en langue étrangère dans une plate-forme audio-synchrone, *Alsic*, vol. 9, p. 61-101. http://alsic.u-strasbg.fr/v09/chanier/alsic_v09_08-rec3.htm

(Courtin et Talbot, 2005)

COURTIN, C., TALBOT, S. (2005). An Architecture to Record Traces in Instrumented Collaborative Learning Environments, *Actes IADIS de CELDA 2005*, Porto, Portugal, 14-16 décembre 2005, p.301-308.

(Goffman, 1974)

GOFFMAN, E. (1974). *Frame analysis: an essay on the organization of experience*. New York: Harper & Row.

(Goodwin et Duranti, 1992)

GOODWIN, C., DURANTI, A. (1992). Rethinking Context: An introduction. In Goodwin, C. & Duranti, A (dirs). *Rethinking Context: Language as an interactive phenomenon*, Cambridge, CUP, p. 191-227.

(Halliday, 1989)

HALLIDAY, M.A.K. (1989). Part A. In Halliday M.A.K. and Hasan R. (dirs). *Language, Context and Text: Aspects of language in a social-semiotic perspective.* Oxford. Oxford University Press, p. 55-79.

(Hampel, 2006)

HAMPEL R. (2006). Rethinking task design for the digital age: A framework for language teaching and learning in a synchronous online environment, *Recall*, Vol. 18, p. 105-121.

(Harrer *et al.*, 2007)

HARRER A., ZEINI S., KAHRIMANIS G., AVOURIS N.,. MARCOS J. A, MARTINEZ-MONES A., MEIER A., RUMMEL N., SPADA H. (2007). Towards a Flexible Model for Computer-based Analysis and Visualization of Collaborative Learning *Activities, Proc. CSCL 2007*, 16-21 July 2007, New Jersey, USA.

(Heraud *et al.*, 2005)

HERAUD, J.-M., MARTY, H.-C., FRANCE, L., CARRON, T. (2005). Une aide à l'interprétation de traces : application à l'amélioration de scénarios pédagogiques, *Actes de EIAH 2005*, Montpellier, mai 2005, p. 237-248.

(Jones, 2004)

JONES, R. (2004). The problem of Context in Computer-Mediated Communication. In Levine P. & Scollon, R. (dirs). *Discourse & Technology, multimodal discourse analysis*, Washington, Georgetown University Press, p. 20-33.

(Lamy, 2006)

LAMY M.N. (2006). Conversations multimodales : l'enseignement-apprentissage à l'heure des écrans partagés. *Revue Le Français Dans le Monde, Recherche et Applications*, n°40, p. 129-138.

(Lamy, 2007)

LAMY, M-N. (2007). Multimodality in Online Language Learning Environments: Looking for a methodology. In Baldry, Anthony Montagna, Elena (dirs.). *Interdisciplinary Perspectives on Multimodality: Theory and practice*. Proceedings of the Third International Conference on Multimodality. Campobasso: Palladino, p. 237-254.

(Levine et Scollon, 2004)

LEVINE, P. SCOLLON, R. (dirs) (2004). *Discourse & Technology, multimodal discourse analysis*, Washington , Georgetown University Press.

(Mannila *et al.*, 1997)

MANNILA, M., TOIVONEN, H., VERKAMO, A. (1997). *Discovery of Frequent Episodes in Event Sequences*, Springer Science+Business Media B.V., Formerly Kluwer Academic Publishers B.V., p 259 – 289.

(Noras *et al.*, 2007)

NORAS M., REFFAY C., BETBEDER M.-L. (2007). Structuration de corpus de formation en ligne en vue de leur échange, *Actes de la conférence EIAH*, Lausanne, Suisse, 27-29 juin 2007. http://edutice.archives-ouvertes.fr/edutice-00154372

(Norris, 2004)

NORRIS S. (2004). Multimodal Discourse Analysis: A Conceptual Framework. In Levine P. & Scollon, R. (dirs). *Discourse & Technology, multimodal discourse analysis*, Washington, Georgetown University Press, p. 101-115.

(Ollagnier-Beldame et Mille, 2007)

OLLAGNIER-BELDAME M., MILLE A. (2007). Faciliter l'appropriation des EIAH par les apprenants *via* les traces informatiques d'interactions ?, *Revue Sciences et Technologies de l´Information et de la Communication pour l´Éducation et la Formation.*

(Reffay *et al.*, 2007)

REFFAY C., CHANIER T., NORAS M., BETBEDER M.-L., (2007). Contribution à la structuration de corpus de formations en ligne pour un meilleur partage en recherche. *Communication au Colloque EPAL : Echanger Pour Apprendre en Ligne*, Grenoble, juin.

(Settouti *et al.*, 2006)

SETTOUTI, L.-S., PRIE, Y. MILLE, A. MARTY, J.-C. (2006). Systèmes à base de traces pour l'apprentissage humain, *Actes de TICE 2006*, Toulouse, octobre.

(Smith et Gorsuch, 2004)

SMITH, B., GORSUCH, G.J. (2004). Synchronous computer mediated communication captured by usability lab methodologies: New interpretations, *System*, 32(4), p.553-575.

(Vetter, 2004)

VETTER A. (2004). Les spécificités du tutorat à distance à l'Open University, *ALSIC*, Vol 7, p.107-129. http://alsic.u-strasbg.fr/v07/vetter/alsic_v07_06-pra2.htm

(Vetter et Chanier, 2006)

VETTER A., CHANIER T. (2006). Supporting oral production for professional purpose, in synchronous communication with heterogeneous learners, *ReCALL*, Vol 18, p. 5-23. http://edutice.archives-ouvertes.fr/edutice-00080316