

Multimodalité et expression en langue étrangère dans une plate-forme audio-synchrone

Thierry CHANIER, Anna VETTER
Université de Franche Comté, France

Résumé : *Le développement actuel d'environnements audio-synchrones pose un nouveau défi, non seulement aux concepteurs et aux tuteurs de formations en langues à distance, mais encore aux chercheurs en analyse du discours. Il devient en effet indispensable de comprendre comment s'organise la communication à finalité pédagogique dans ce type d'environnement. L'article aborde ce champ encore peu exploré sous deux angles : l'un méthodologique et l'autre analytique. Nous fournissons, d'une part, un cadre méthodologique pour l'analyse conversationnelle multimodale à partir des notions de média, mode et modalité et, d'autre part, nous définissons les composantes d'une plate-forme synchrone. La partie analytique synthétise d'abord les premiers résultats quantitatifs de l'expérimentation Copéas sur l'usage combiné des modalités audio et clavardage et son impact sur la participation des apprenants. L'analyse indique ensuite, en s'appuyant sur des séquences transcrites, comment s'organise la communication pédagogique lorsque les différents modes (verbaux et non verbaux) s'associent pour structurer et soutenir des conversations en L2. Elle montre que le libre choix des modes et des modalités soutient la production verbale des apprenants. Dans ces conversations, le mode parole occupe tantôt une place prépondérante, tantôt est totalement absent des échanges au profit des modes texte, graphique et iconique. Enfin, nous élargissons le champ de description du discours multimodal en considérant le mode spatial qui s'impose comme le contexte dans lequel chaque transaction doit être située. Ces différents exemples constituent l'occasion de mettre en évidence deux questions connexes : celle de la conception et de l'animation pédagogique dans ces environnements de communication multimodaux à travers le phénomène de polyfocalisation ; et celle de la transcription du discours multimodal.*

- 1. Introduction
- 2. Discours multimodal, enseignement et apprentissage
- 3. Espaces et moyens de communication multimodaux dans une plate-forme audio-synchrone
- 4. Mesure de la participation orale et équilibrage des modalités audio et clavardage
- 5. La multimodalité dans tous ses états
- 6. Conclusions
- Références
- Annexes

1. Introduction



La technologie est en train, encore une fois, d'offrir aux pédagogues de nouveaux environnements dans lesquels ils vont pouvoir scénariser et animer des formations en langue. Les systèmes d'information et de communication combinent en un seul progiciel des outils garantissant à un groupe d'utilisateurs en ligne une communication audio et textuelle de très bonne qualité, avec des outils de production collective et, enfin, le choix dynamique du regroupement dans un ou plusieurs lieux ou salles virtuels. Souvent baptisés "systèmes de visio / audio conférence", "classe virtuelle", ces systèmes, que nous appellerons "plates-formes audio-synchrones", se distinguent donc nettement du simple clavardage (*textchat*), des logiciels de visio / audio conférence limités à deux personnes ou encore de la vidéoconférence avec ses salles, non virtuelles, spécialement aménagées.

De tels environnements ne sont pas à la portée des individus ni des enseignants isolés. Ce sont les institutions de formation qui, à la suite d'un choix raisonné, peuvent d'un seul coup offrir à tout leur personnel la possibilité de faire travailler simultanément, autrement dit de façon synchrone, les apprenants en groupe, que ceux-ci soient localisés à l'intérieur de l'institution ou, ce qui est bien plus intéressant, dans des lieux extérieurs dont leur domicile, chacun disposant d'un ordinateur standard et d'une liaison Internet normale (premier niveau de vitesse de l'ADSL ou équivalent suivant les pays).

De ce fait, s'ouvrent de nouvelles situations pour les formateurs en langues. L'expression orale considérée comme primordiale dans ces environnements multimodaux vient élargir le champ des possibles en Formation À Distance (FAD). L'enjeu porte sur la constitution de groupes de la taille de ceux rencontrés en formation intensive en présentiel, de groupes où la présence de natifs est systématisée malgré l'éloignement géographique des interlocuteurs, de groupes où la parole et plus généralement la production en L2 s'équilibrent mieux entre chacun des participants.

Mais pour que ce potentiel devienne réalité, il faut, à l'heure où se déploient de tels dispositifs à des échelles nationales, que les formateurs s'en saisissent pour mettre en scène la communication en L2 en évitant la reprise caricaturale de certains aspects du présentiel. Il

devient donc urgent de comprendre les particularités du discours multimodal que l'on peut scénariser et animer dans ce type d'environnements.

La recherche en ce domaine et, plus précisément, dans celui se consacrant à la communication médiée par les systèmes informatiques (*Computer Mediated Communication*, CMC), a particulièrement étudié la conversation textuelle par clavardage. Ces travaux permettent, avec le recul du temps, de se faire une idée précise des possibilités offertes par ce moyen de communication directe et synchrone [Calico05]. Mais, même si un faisceau convergeant de résultats montre l'impact que peuvent avoir certaines formations qui en font largement usage sur les compétences orales, ce n'est que depuis peu que des environnements technologiques offrent la possibilité de mettre les apprenants en situation simultanée de compréhension et production orales dans de vrais dialogues, ou plutôt polylogues. Les premières recherches rassemblant audio et clavardage mettent ces deux modalités de communication en concurrence [Sykes05] ou se limitent à des échanges individuels apprenant-enseignant [Blake05]. Nous nous intéressons ici à l'usage combiné de ces modalités dans des échanges de groupe et, plus encore, à l'étude de leur association à d'autres modes et modalités dans des espaces en ligne où se construisent des discours d'un genre nouveau. La compréhension de l'agencement de cette multimodalité est sans doute la clef pour le passage à un enseignement complet (dans les quatre compétences) des langues en ligne.

Notre étude commencera dans la section suivante par un survol des publications récentes sur le discours multimodal, thème qui connaît un regain d'intérêt dans plusieurs domaines et, notamment, en contexte pédagogique. Nous y fixerons notre terminologie sur la multimodalité et préciserons les enjeux didactiques de ces recherches tels qu'énoncés par les collègues. Dans la section suivante, nous définirons la notion de plate-forme synchrone, en listant ses composants, en établissant des critères pour le choix de tels progiciels. Composants et fonctionnalités seront illustrés à partir de l'exemple d'une plate-forme utilisée dans notre projet de recherche *Copéas* (COmmunication PÉdagogique en environnement orienté Audio Synchrone). La section 4 décrit justement la formation à distance en anglais qui a servi de terrain d'observation aux chercheurs. Elle aborde la question de la prise de parole en L2 dans des groupes fortement hétérogènes, en résumant des résultats en cours de publication qui se concentraient sur la mesure quantitative des prises de paroles à l'audio et sur l'équilibre des participations dans les modalités audio et clavardage.

La section suivante développe le corps principal de l'article en s'appuyant sur les notions, le vocabulaire, les résultats introduits précédemment, complétés par un rappel de termes empruntés à l'analyse conversationnelle monomodale. Nous examinons alors de façon détaillée cinq extraits de discours tirés de nos transcriptions, qui font fortement varier de l'un à l'autre le jeu des modes et modalités utilisés par les participants et, surtout, leurs imbrications profondes. Suivant la classification des travaux de recherche sur le discours en ligne, établie par Herring [Herring04], le présent article s'inscrit dans le paradigme de l'analyse conversationnelle et porte sur le domaine des interactions. Notre objectif sera double : d'un point de vue linguistique, intégrer à l'analyse conversationnelle les éléments de description de

ce nouveau type de discours et, d'un point de vue didactique, tirer les premiers enseignements de la façon d'orchestrer la production en L2.

2. Discours multimodal, enseignement et apprentissage

Des chercheurs de disciplines fort variées se passionnent pour le discours multimodal ([LevineScollon04] ; [KressVanLeeuwen01]) avec des perspectives que nous classerons, par souci de simplification, en deux catégories. Une partie d'entre eux postule le fait que le discours est intrinsèquement de nature multimodale. La disponibilité des technologies numériques élargit considérablement, après l'ère des premiers magnétophones et caméras, les possibilités de capture et d'analyse des différents aspects de la communication verbale et non verbale dans les discours quotidiens à des fins d'études ethnologiques, linguistiques ou en communication. Le reste des chercheurs s'interroge sur la nature des échanges se produisant à l'intérieur des environnements technologiques multimodaux, notamment au travers des différents outils de communication développés sur Internet.

Dans cette deuxième catégorie, certains auteurs sont même principalement motivés par l'étude des discours survenant en contexte pédagogique que cela soit en présentiel, en cours de science par exemple [KressJewitt01], ou en ligne, comme c'est le cas ici. Cette recherche, assez récente, croise la voie ouverte dix à quinze ans plus tôt par les spécialistes des Interactions Humain-Machine (IHM) pour lesquelles on parlait déjà de communication multimodale [BellikTeil92]. La variété des spécialités disciplinaires des chercheurs et de leurs objectifs d'étude rend difficile la définition consensuelle de ce qu'est la communication multimodale. Si les notions de "média", "mode" et "modalité" sont systématiquement considérées comme centrales, elles le sont sous des acceptions différentes.

2.1. Média, mode et modalité

Le **média** est une ressource matérielle utilisée dans la production / diffusion de produits ou événements sémiotiques. Cette ressource correspond aussi bien au support matériel qu'à l'outil / dispositif technologique qui acquiert et diffuse l'information ([BellikTeil92] : 1 ; [KressVanLeeuwen01] : 22). Ce terme recouvre donc les objets tels que encre, papier, caméras, ordinateurs ou gants, lunettes, stylos utilisés en réalité virtuelle. Dans notre cas, les médias d'entrée seront le clavier, la souris et le microphone ; ceux de sortie, l'écran et le casque audio.

Les **modes** sont des ressources sémiotiques qui concourent à la réalisation simultanée du discours et des interactions. Le mode textuel, le mode parole (*speech*), le mode geste, le mode image, etc. renvoient chacun à des systèmes de signes dotés d'une grammaire (terme à considérer dans un sens très large) qui en décrit l'organisation et l'usage. En ce qui concerne le terme **modalité**, les acceptions varient considérablement. En IHM, le mode est d'abord perçu dans son aspect communication et la modalité se présente alors comme une forme concrète

particulière de communication. Dans certains travaux plus orientés vers la sémiotique, où l'on s'intéresse à la fois à cet aspect communication et à la construction de significations, la modalité est plus une relation qui, partant d'un ensemble de signes / d'un énoncé dans un mode donné, exprime la validité de la représentation qu'il réalise ([VanLeeuwen04] : 15). On peut retrouver alors la notion courante de modalité en linguistique, bien que l'on ne se limite pas à la présence de marqueurs linguistiques comme les verbes modaux. Suivant les acceptions choisies on classera alors la musique comme une modalité en IHM [BellikTeil92] ou un mode ([KressVanLeeuwen01] ; [LevineScollon04]).

Dans la plate-forme synchrone que nous présenterons dans la partie suivante, nous considérerons les modes texte, parole, graphique, icônes et même un mode "spatial", que nous définirons en section 5.4, et qui correspond grossièrement à la gestuelle et au déplacement des locuteurs en salle de classe. Reprenant l'acception utilisée en IHM, nous parlerons des modalités attachées à chacun des modes. Ainsi le mode texte apparaît dans la "modalité clavardage", la "modalité traitement de texte", voire la "modalité carte conceptuelle" ou tableau blanc. À un mode peut donc correspondre plusieurs modalités ou une seule (cas de la parole avec la modalité audio). Un outil peut intégrer plusieurs modes (cas de la carte conceptuelle avec le texte et le graphisme). On parlera de communication multimodale, comme en IHM, quand elle fait intervenir plusieurs modes, ou quand plusieurs modalités sont associées à un seul mode.

2.2. La multimodalité dans les situations didactiques

Pourquoi s'intéresser à la multimodalité dans les situations didactiques ? Pour Kress et Jewitt [KressJewitt01], le langage n'est qu'un des modes, minoritaire même à leurs yeux, apparaissant en classe de science. Ce qu'ils appellent le "mode action" est une composante fondamentale de la classe. La gestuelle, la désignation d'objet, les mouvements du corps dans la classe, y compris le positionnement de l'individu, enseignant ou élève, par rapport aux autres participants, font l'objet d'une étude précise, au même titre que les éléments langagiers, les schémas, etc. Le recours à ce mode action et à ses différentes modalités permet aussi d'expliquer des concepts abscons, ou de montrer qu'ils ont été assimilés (cf. une gestuelle explicative du mouvement des molécules dans un gaz). De même, l'étude des schémas associés aux textes des élèves permet de vérifier ce qu'ont appris ceux-ci et comment ils ont compris les concepts du cours. La communication multimodale est donc au cœur des processus d'enseignement et d'apprentissage et devrait faire l'objet d'une recherche didactique, ce qui sous-entend la mise au point de méthodes de transcriptions de cette multimodalité.

En apprentissage / enseignement des langues, on pourrait *a priori* penser que l'on n'est concerné que par le verbal, autrement dit par les modalités où interviennent parole et texte. Mais, déjà en présentiel, adopter un tel point de vue serait faire fi de la gestuelle qui complète, remplace ou duplique la communication verbale [CalbrisPorcher89] et qui, pour nombre de didacticiens, fait partie de la problématique d'apprentissage d'une L2 et lui est spécifique. *A fortiori* dans les types d'environnements multimodaux considérés ici, les actes verbaux sont en correspondance permanente avec les actes non verbaux. Par exemple, à une demande

d'explication d'une notion difficile en L2, un apprenant peut répondre en accomplissant une action sur un graphique d'une carte conceptuelle. Comment évaluer sa compétence en compréhension orale sans tenir compte de cela ? Comment suivre ce discours et le transcrire sans représenter les interactions entre les différents modes et modalités ? D'autre part, cette possibilité de choix de modes et de modalités de communication offerte à l'apprenant ne joue-t-elle pas comme un facteur important dans la prise de risque à s'exprimer en L2 ?

À un autre niveau, il faut nous interroger sur les finalités pédagogiques des formations que nous concevons en L2. Dans le cas de *Copéas*, il s'agissait d'une formation en langue de spécialité où se mêlaient objectifs d'apprentissage de l'anglais et maîtrise des nouveaux modes d'interaction en ligne. Ce rapprochement ne vaut pas que pour l'enseignement de spécialité. Par exemple, on délivre aujourd'hui des diplômes de littérature anglaise à des étudiants qui n'ont lu qu'une fraction des livres étudiés par leurs aînés. En contrepartie, ces nouveaux étudiants ont été confrontés aux autres médias véhiculant cette littérature. Nos formations doivent s'adapter aux nouveaux types de discours auxquels sont exposées les générations nées avec Internet, discours qui occuperont une place de plus en plus grande dans leur future vie sociale. Thorne et Payne [ThornePayne05] ont bien illustré le défi auquel vont être confrontés les enseignants, issus d'une autre génération, et l'intérêt que peuvent revêtir pour l'apprentissage en L2 de nouveaux médias tels les *blogs*, les *wikis* et la *baladodiffusion* (*podcasting*), où interagissent téléphone portable, baladeurs audio et réseaux informatiques.

Pour revenir à des technologies qui placent au premier plan la communication synchrone, Jones [Jones04] rapporte que, lors de ses cours de rédaction en anglais se déroulant en salle machines à l'université à Hong-Kong, l'essentiel des conversations menées par ses étudiants s'accomplit avec des personnes non présentes dans la salle et sur des sujets hors contexte du cours ! Pourtant, explique-t-il, les étudiants accomplissent leur travail et, tout en composant collaborativement leurs écrits, ils effectuent des recherches en ligne sur des matériaux de référence pour le cours, téléchargent des notes du cours, écoutent l'enseignant donner des explications sur le style des écrits et, en même temps, conversent avec des amis ou des étrangers par clavardage. Cette **polyfocalisation**, comme la nomme Jones ([Jones04] : 27), reprenant en cela Scollon, est également présente dans les échanges de notre formation et, même si l'attention des apprenants semble bien se limiter à l'objet du cours (ce qui peut comprendre la tâche et le hors-tâche), elle complexifie notablement la tâche de suivi par l'enseignant.

Ainsi donc, plutôt que de demander aux responsables réseaux des établissements de formation de fermer les ports de communication des systèmes de clavardage utilisés par les étudiants, peut-être vaut-il mieux commencer par étudier la nature des échanges multimodaux dans la perspective d'un usage régulier et contrôlé en formation en langues !

3. Espaces et moyens de communication multimodaux dans une plate-forme audio-synchrone

Dans cette partie, nous définissons la notion de "plate-forme synchrone", et nous donnons une liste des principales fonctionnalités et de leur utilisation en situation de communication et interaction multimodale. Les usages prévus ne correspondant pas toujours à ceux observés, nous exposons le point de vue des apprenants. La plate-forme *Lyceum* développé par l'*Open University*, pour les besoins de sa formation à distance nous sert d'illustration, mais notre propos se veut plus général.

3.1. Composants et usages de ces plates-formes

Dans le milieu de la FAD, on a coutume de dénommer "plate-forme de téléapprentissage" (*Learning Management System*, LMS), ces environnements numériques déployés par les institutions de formation pour permettre aux apprenants et formateurs de travailler en ligne. Même si la plupart de ces plates-formes dispose entre autres d'un outil de clavardage, le mode d'utilisation y est essentiellement asynchrone, chacun travaillant à son rythme. De même nous parlerons de **plate-forme (audio)synchrone** pour désigner un collecticiel dédié aux interactions synchrones en groupes sur Internet rassemblant dans une même interface au moins trois composants activables simultanément du côté client :

- le composant spatial pour se déplacer dans les salles et sous-salles de travail ;
- un second composant rassemblant divers outils de communication synchrone dont les modules audio et clavardage ;
- un troisième composant rassemblant différents outils de production collaboratifs.

La composante de communication audio étant systématiquement présente dans ce type de plate-forme, nous parlerons simplement par la suite de "plate-forme synchrone". À ces trois composants s'ajoute, dans une autre partie du système, un module d'administration pour la gestion des comptes, des statuts et la réservation des espaces de travail. Ce module est généralement à usage exclusif des formateurs. La définition que nous venons de donner permet de distinguer ces environnements des logiciels de communication synchrone que les internautes peuvent utiliser sans passer par des serveurs. (Les logiciels de communication synchrones ordinaires permettent de clavarder à plusieurs mais limitent l'utilisation de l'audio à un nombre très restreint d'interlocuteurs, deux pour *MSN*, cinq pour *Skype*, par exemple, ou se réduisent à une simple conférence téléphonique sur Internet sans modalité supplémentaire.)

Dans la figure 1, visualisant l'interface de l'utilisateur dans *Lyceum* [Lyceum06], chaque composant est situé dans un cadre numéroté. Ces trois cadres sont intégrés et ne peuvent se chevaucher. Dans le cadre 1, l'utilisateur peut naviguer d'une salle à l'autre et savoir qui est dans quelle salle. Le cadre 2 rassemble les outils de communication synchrone. L'espace du cadre 3 est dédié à la production collaborative où l'utilisateur, grâce à la frise supérieure, peut

sélectionner le document à visualiser et savoir qui visualise quoi. Afin d'éviter des quiproquos, un utilisateur peut rassembler tous les participants de la salle devant le même document (cf. figure 1, bouton **Gather** en haut à droite de la fenêtre).

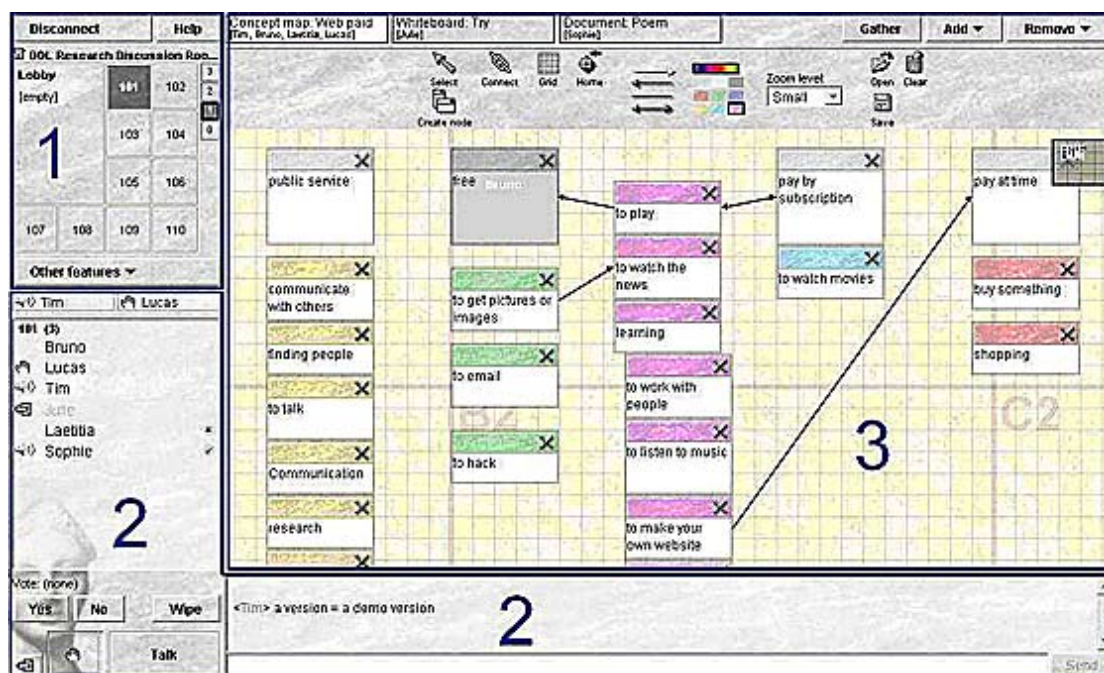


Figure 1 - Les trois composantes d'une plate-forme synchrone, à partir de l'exemple *Lyceum*. Le lecteur trouvera une présentation plus détaillée de *Lyceum* dans [Vetter04].

Le tableau 1 donne l'appréciation par les apprenants du degré d'utilité de certaines fonctionnalités par composants. Lors de l'administration des questionnaires à la fin de la formation, les 14 apprenants ont eu à répondre à la question suivante : "si vous aviez à choisir une plate-forme synchrone, quelle importance / utilité accorderiez-vous à chacune des fonctionnalités / outils pour guider votre choix ?". À cette date, ils avaient non seulement l'expérience de la formation en langue mais aussi celle d'une autre formation collaborative synchrone dans *Centra* [Centra06], plate-forme déployée par le ministère de l'Éducation Nationale. Parmi les outils de communication synchrone, le clavardage est noté comme tout à fait indispensable, comme le système d'icônes, dit de "vote", pour indiquer approbation et dénégation (figure 1 : **Laetitia** et **Sophie**). Sont considérés comme fort utiles également, la possibilité de lever la main pour demander la parole (figure 1 : **Lucas**), le repérage de la personne qui parle (figure 1 : **Tim**) et le bouton pour parler plutôt que la possibilité de laisser le micro ouvert en permanence. Parmi les outils de production collaborative, le traitement de texte intégré a été plébiscité, la carte conceptuelle un peu moins. Bien qu'utile et support de nombreuses activités, la carte conceptuelle a été jugée moins évidente à manier. Les apprenants ont estimé important de savoir à tout moment qui était en train d'effectuer une action dans un outil (figure 1, cadre 3 : **Bruno** modifie la boîte **free**).

En majorité, les apprenants ont jugé utile (mais pas indispensable) le fait qu'il y ait plusieurs salles annexes disponibles pour un même groupe. Du côté des formateurs, cette fonctionnalité est considérée comme tout à fait indispensable pour scénariser un cours comme une suite de

plénières et de travaux en sous-groupes, justement dans les salles annexes. Notre expérimentation montre le grand niveau d'engagement des apprenants dans des dialogues et des tâches qui fréquemment ne sont pas directement sous la vue du tuteur.

Spatial (1)		Outils de communication synchrone (2)	
Salles de réunion annexes	7,5	Audio	
Outils de production collaborative (3)		Clavardage	9,4
Traitement texte intégré	8,6	Icônes pour voter oui / non	9,7
Tableau blanc	7,7	Icônes pour demander la parole	8,9
Voir qui agit dans un outil	7,7	Bouton pour parler	8,9
Carte conceptuelle	6,3	Savoir qui parle	8,0
		Distinguer les droits de parole	3,7

Tableau 1 - Fonctionnalités de l'interface utilisateurs, classées par composant et dont les degrés d'utilité ont été notés par les apprenants sur une échelle de 10.

On notera dans le tableau 1 l'avis des apprenants visant à écarter la distinction des droits de parole entre participants suivant les statuts. La plate-forme *Lyceum* offre à tous, formateurs et apprenants, les mêmes droits, ce qui autorise les approches pédagogiques basées sur la collaboration.

À ce propos, ce que nous avons appelé "plate-forme synchrone" est aussi dénommé par d'autres "systèmes de visio / audio conférence". L'utilisation du terme "*conference*" et de ses dérivés est courant en langue anglaise. Quant il s'agit de qualifier des systèmes comme *Lyceum*, certains auteurs, désirant mettre l'accent sur la présence de modalités non textuelles aux côtés de l'audio, les dénomment alors "*audio-graphic conferencing systems*" [HampelBaber03]. Au-delà de ces variations terminologiques (la nôtre mettant l'accent sur la situation de téléformation), il importe de repérer les finalités premières de ces environnements. La plupart distinguent nettement les droits des utilisateurs selon que l'on est "animateur" ou "participant". Même si leurs concepteurs essayent de les présenter aussi sous l'angle éducatif, ces progiciels n'ont pas été élaborés avec cet objectif. Ils induisent généralement une vision transmissive du savoir. Le scénario pédagogique typiquement mis en avant par leurs promoteurs, ressemble à s'y méprendre à celui du cours magistral en amphithéâtre, la participation des apprenants se limitant à poser épisodiquement des questions à l'orateur, le "maître de conférences". La présence d'outils de production collective dans une partie de ces systèmes ne change rien à ce biais. L'utilisateur averti remarquera que, dans certains systèmes, un outil comme le tableau blanc ne permet pas de déplacer les objets graphiques après leur création ..., ce qui réduit à néant toute possibilité de négociation entre les participants !

Le tableau 2 expose le point de vue des apprenants *a posteriori* sur l'adéquation des outils à des objectifs discursifs ciblés, souvent rencontrés dans l'apprentissage collaboratif en ligne. La

distinction entre outils de communication synchrone et outils de production y ressort clairement. On gardera en mémoire ces valeurs lorsque ces mêmes outils apparaîtront dans les séquences d'interaction présentées dans la suite de l'article.

	Audio	Clavar- dage	Vote	Main levée	Signal absence momentanée	Traitement de texte	Carte concep- tuelle	Tableau blanc
Gérer le dialogue	+++	+	++	+++	+			
Décider	+++	++	+++	+				
S'exprimer sur le plan socio-affectif	+++	++	+					
Faire du remue-ménages	+	+	NA	NA		+	++	++
Produire ensemble en laissant une trace visible	NA		NA	NA	NA	+++	+++	++
Montrer quelque chose de préparé	NA	NA	NA	NA	NA	+++	+++	+++

Tableau 2 - Appréciation par les apprenants de l'adéquation d'outils de communication ou de production à des objectifs discursifs.

3.2. L'intégration de tels environnements dans les institutions de formation

Suivant de peu la vague des plates-formes asynchrones utilisées en FAD, les différentes offres concurrentes de systèmes synchrones arrivent sur le devant de la scène. Les formateurs sont les premiers concernés par les choix que vont opérer leurs institutions. Nous avons, dans la section précédente, indiqué les fortes contraintes induites par les fonctionnalités de ces systèmes sur la scénarisation pédagogique des activités courantes. Mais il est un autre type de critères qui intéresse les formateurs au quotidien.

Les plates-formes synchrones qui comprennent bien un côté serveur (institution) et client (utilisateur) s'apprécient d'abord sur les critères suivants.

- Nombre maximum d'utilisateurs simultanément en session dans toutes les salles. Le fait de devoir gérer des flux multimédia nécessite des infrastructures coûteuses.

- Facilité et souplesse de la réservation des salles de travail par les utilisateurs autorisés.
- Facilité de l'installation du logiciel client et de son utilisation immédiate quelle que soit la connexion Internet (dans une institution ou par un fournisseur d'accès) et la configuration matérielle du poste client.
- Qualité de l'audio pour l'utilisateur de base doté d'un micro-casque sans autre composant additionnel que ceux livrés en standard sur les ordinateurs.

Le caractère impératif de ces quatre critères et, en particulier les coûts induits pour l'infrastructure serveur, nécessitera un engagement politique fort de l'institution pour l'intégration de ces technologies, voire la mutualisation de moyens avec d'autres institutions. Notons que ce type de plate-forme correspond bien (comme les plates-formes asynchrones, mais de façon encore plus immédiate) au développement actuel de la société de l'information : l'institution de formation doit d'abord penser aux conditions de travail de ses formateurs et apprenants lorsqu'ils ne sont pas dans ses murs, qu'ils soient en situation de FAD ou, simplement, en dehors des heures de cours en présentiel.

4. Mesure de la participation orale et équilibrage des modalités audio et clavardage

Les discours multimodaux que nous étudierons dans la section suivante sont extraits de la formation en langue autour de laquelle s'est construit le projet de recherche *Copéas*. Nous en rappelons ici les principales caractéristiques de façon à ébaucher le contexte dans lequel s'insèrent ces discours. Nous citons quelques conclusions de nos travaux récents ([VetterChanier05] ; [ChanierEtal06]) concernant les modalités audio et clavardage.

4.1. La formation en langue Copéas ou comment retrouver le chemin de la parole en L2

Le public concerné rassemblait les stagiaires d'un master consacré à l'ingénierie des dispositifs de formations ouvertes et à distance. Comme c'est souvent le cas dans les formations professionnelles de 3^e cycle, le groupe d'apprenants était fortement hétérogène, la moitié de ses membres relevant de la formation continue, d'où une disparité d'âge (25 à plus de 50 ans), de niveau d'études, d'expérience professionnelle et, bien sûr, de biographie langagière. Sur ce dernier point, nous retrouvons à une extrémité certains bilingues russes, polonais, espagnol, un titulaire d'une maîtrise d'anglais et, à l'autre extrémité, des stagiaires ayant cessé tout apprentissage des langues depuis des années. Précisément, sur 14 apprenants, neuf ne pratiquaient plus aucune langue étrangère, dont quatre depuis 30, 25, 22 ou 15 ans, trois entre 7 et 5 ans. Les résultats d'une autoévaluation effectuée sur le référentiel européen dispersaient le groupe entre les niveaux A1 et C1 en anglais.

Nous avons conçu avec les collègues de l'*Open University*, une formation à distance en anglais sur objectifs spécifiques, tirant parti des connaissances professionnelles du groupe

(connaissances des notions en FAD, expérience de travail en groupe à distance) pour les mobiliser sur des apprentissages mêlant objectifs professionnels (découverte de nouveaux environnements technologiques) et langagiers. Sur ce dernier point, il s'agissait donc de permettre aux apprenants de (re)trouver le chemin de la parole à travers la production de discours oraux ou écrits.

La formation s'est étalée sur 10 semaines, début 2005. Les apprenants ont été répartis en deux groupes équilibrés par niveaux, le groupe des faux débutants et celui de niveau intermédiaire à avancé. Deux formateurs ont participé à la conception, centrée sur l'analyse et l'évaluation de sites Internet pouvant être utilisés en FAD. Chaque groupe a retrouvé son tuteur pendant huit séances *Lyceum* d'une heure à une heure trente chacune ; chacun se connectant depuis son domicile personnel en utilisant son ordinateur et sa liaison Internet propres. Ces séances synchrones centrées sur des activités précises, donnaient souvent lieu à des préparations personnelles hors ligne. Chaque séance alternait rassemblement du groupe entier dans une même salle *Lyceum* avec travaux en demi-groupe dans des salles distinctes. La formation totale avait été conçue comme devant correspondre à une vingtaine d'heures.

Le dispositif expérimental a été programmé pour enregistrer l'intégralité des interactions survenues dans l'environnement audio-synchrone. Notre méthode de recueil et de transcription des données nous a permis de mesurer la durée des prises de parole audio de chaque participant en calculant le temps de parole en continu^[1] et le nombre de tours de parole. Nous avons alors intégré à ces tours de parole audio, ceux provenant du clavardage de façon à étudier le discours produit dans ces deux modalités. Rappelons brièvement les principales conclusions de nos études en les resituant dans la problématique des recherches internationales.

4.2. Le phénomène d'égalisation entre l'audio et le clavardage

Les recherches sur les interactions synchrones en langue étrangère ont démontré le bénéfice des conversations synchrones à l'écrit sur la participation orale des apprenants. La plupart de ces recherches se sont inscrites dans le cadre d'une comparaison entre les interactions en classe et les interactions en ligne et Ortega rappelle qu'elles ont abouti à deux conclusions importantes [Ortega97] :

- les apprenants participent davantage en ligne qu'en présentiel [Kern95] ;
- la répartition de la parole est plus homogène (**égalité**) dans les conversations synchrones à l'écrit entre les apprenants que dans le groupe-classe en présentiel ([Warschauer96] ; [Fitze06]).

D'autres études ont avancé que les structures linguistiques produites lors de ces conversations étaient plus complexes et davantage formelles que celles produites à l'oral en présentiel ([Kitade00] ; [Sotillo00]). Enfin, Payne et Whitney ont montré que le clavardage contribue à l'amélioration des compétences orales [PayneWhitney02]. Cependant, Thorne et Payne

[ThornePayne05] rappellent que ces résultats ont été tirés d'expériences qui se sont déroulées dans une seule modalité de communication, le clavardage.

L'intérêt de l'expérimentation présentée par notre dispositif de recherche *Copéas* est de questionner une situation de formation en langue présentant une forte disparité de niveaux linguistiques dans le cadre d'un environnement multimodal qui combine deux modalités (audio et clavardage) sur le plan quantitatif (la participation orale des apprenants) et qualitatif (la complexité du discours professionnel). Sur le plan qualitatif, on se reportera aux conclusions de [VetterChanier05] qui montrent que des conversations notionnelles de haut niveau peuvent être tenues même par les étudiants de faible niveau linguistique.

Les aspects quantitatifs de notre recherche nous permettent d'affirmer que l'hétérogénéité des niveaux linguistiques et des parcours des apprenants n'est pas un frein à leur engagement dans la participation orale lors d'une formation à distance dans un environnement audio-synchrone. Concernant le groupe de niveau intermédiaire-avancé, cette affirmation n'est pas nouvelle et vient seulement confirmer ce que d'autres études ont déjà noté à propos de groupes hétérogènes de niveau B en présentiel. En revanche, les résultats sont plus surprenants concernant le groupe le plus faible sur le plan linguistique. On a, en effet, constaté que chez ces derniers, certains compensent le fait de s'exprimer relativement peu dans la modalité audio par un usage intensif du clavardage. On retrouve bien ici le phénomène d'égalisation, mais en considérant l'utilisation combinée des deux modalités.

Plus précisément, l'examen des temps totaux de participation audio comptabilise un tiers de temps supplémentaire pour les apprenants du groupe intermédiaire à avancé par rapport à ceux du groupe de faux débutants. Mais l'examen de l'évolution des temps de parole par session montre que la durée de production audio des faux débutants est multipliée par trois entre la première et la dernière séance, pour atteindre en fin de formation un taux comparable à celui de l'autre groupe en début de formation. En ce qui concerne le clavardage, on observe que les faux débutants s'expriment presque deux fois plus dans cette modalité, et que le nombre moyen de mots par intervention est supérieur à celui du groupe plus avancé.

Les courbes de la participation dans le clavardage révèlent un comportement très différent des groupes. Le groupe de faux débutants semble trouver son rythme à partir de la troisième séance, le taux de participation dans le clavardage accompagnant, dans les hausses comme dans les baisses, le taux de participation audio. Il semble donc que le clavardage constitue, pour ceux-ci, un soutien significatif pour la conversation et peut-être même une alternative à la prise de parole en modalité audio. Dans le groupe le plus avancé, le faible indice de participation dans le clavardage montre que cet outil n'est *a priori* pas en lien direct avec l'audio.

Enfin, lors de l'examen des comportements individuels des apprenants, nous avons constaté que le choix de l'une ou l'autre des modalités audio ou clavardage pour s'exprimer n'est pas systématiquement lié au niveau linguistique. En effet, on observe une propension à utiliser

plus le clavardage que la moyenne chez certains apprenants du groupe des faux débutants qui s'expriment sans trop de difficulté à l'oral dans la modalité audio. Au contraire certains rechignent à utiliser cette modalité. On pourrait presque parler ainsi de "clavardophiles" et "clavardophobes".

5. La multimodalité dans tous ses états

Dans cette section, nous partons à la découverte de certains traits du discours multimodal dans les plates-formes synchrones au travers d'extraits de séances *Copéas*. Dans chacune des situations, nous cherchons à savoir ce qui a motivé la sélection par les participants d'un sous-ensemble de modes ou de modalités, au-delà de celles initialement prévues par le scénario pédagogique. Le lecteur gardera également en mémoire l'appréciation des apprenants sur l'importance qu'ils accordent aux outils et modes de communication associés (cf. tableau 1 et tableau 2).

5.1. Premiers éléments du cadre d'analyse des conversations et conventions de transcription

Cette étude requiert la construction d'un cadre d'analyse des conversations. Pour ce faire, il est certes possible d'emprunter des concepts et systèmes de représentations relatifs à l'analyse du discours monomodal. Du point de vue terminologique, nous suivrons la tradition (voir par exemple [Stenström94]) en distinguant les structures du discours suivantes, listées de la plus générale à la plus spécifique.

- Chaque **session** *Lyceum* correspond à un cours. La session se compose d'une suite de **séquences** dont le découpage correspond soit à des changements de salles et de participants, soit à des changements de thèmes prévus par le scénario pédagogique.
- Une séquence est composée d'une ou plusieurs **transactions**. Ce découpage indiquant les grandes étapes de ce que l'on a coutume d'appeler une "conversation" entre un nombre fixé de participants dans une salle donnée.
- Une transaction se décompose en un ensemble d'**échanges**, l'échange étant la plus petite unité mettant en scène l'interaction. Dans sa forme la plus simple, un échange comporte deux tours de parole consécutifs impliquant deux locuteurs différents.
- Un **tour de parole** correspond à un flux verbal exprimé en continu par un seul locuteur dans une modalité unique. Un tour de parole est composé d'un ou plusieurs actes de langage.
- L'**acte** de langage est la plus petite unité du discours. On peut lui faire correspondre une intention de communication unique. De façon plus générale, on parlera "d'actes" pour caractériser les actions élémentaires accomplies dans les modes non verbaux, actions qui viendront se mettre en correspondance avec les tours ou les actes de parole.

Ces notions traditionnelles seront adaptées pour rendre progressivement compte de la spécificité du contexte et du discours multimodal. Nous introduirons simultanément les premiers éléments spécifiques à la représentation des interactions entre modes et modalités. Cette représentation s'appuie elle-même sur les transcriptions des tours de paroles survenus dans chacune des modalités. Pour les distinguer nous avons été obligés d'utiliser un système de codes, dont voici un extrait (voir le détail en annexe 1).

- Une ligne de transcription d'un tour de parole est notée sur quatre cellules de tableau correspondant respectivement au moment d'occurrence du tour par rapport au début de session (en minutes : secondes), à l'identifiant du tour, à l'identifiant de l'interlocuteur et au contenu de la transcription (cf. tableau de la figure 2).
- L'identifiant d'un tour de parole dans la modalité audio est préfixé par **aud** suivi du numéro d'occurrence dans une session (**aud27** est donc le 27ème tour de parole audio). La trame de fond de la cellule est en **mauve**. Pour la modalité clavardage le préfixe est **clav** (trame de fond **orange**). Celui pour le vote est **vot** (trame de fond **jaune**).
- Le contenu des tours de parole utilise les conventions suivantes. Les pauses d'une seconde sont notées par des "+". Les changements de code linguistique apparaissent entre accolades. Les chevauchements entre tours de parole adjacents figurent entre parenthèses et l'intonation montante est notée par une flèche montante "↑" (il n'est en effet pas possible d'introduire la ponctuation, ici un point d'interrogation, qui est une marque de l'écrit). L'icône du vote est transcrite par une affirmation ou une négation en anglais.
- Lorsqu'un **silence** audio de plus de trois secondes survient entre deux tours, ce silence est codé comme un tour de parole audio avec, en lieu et place de l'identifiant de l'interlocuteur le code **sil**, suivi dans la cellule de contenu de sa durée en secondes.
- En ce qui concerne le codage des interlocuteurs : le groupe intermédiaire à avancé est noté **R**, celui des faux débutants **T**, d'après la première lettre du prénom de chacun des tuteurs ; les apprenants du groupe **R** sont notés **AR** suivi d'un numéro ; de même **AT3** est l'identifiant du troisième apprenant du groupe des faux débutants **T** ; les tuteurs de chacun des groupes sont identifiés respectivement par **TutR** et **TutT**.

Le lecteur trouvera peut-être, en première lecture, ce système de codage un peu lourd, mais rappelons qu'il est destiné à transcrire des données sémio-linguistiques fortement hétérogènes et à préserver l'impératif d'anonymat des participants. Sa simplification pour les besoins d'exposé dans un article nous exposerait à des incohérences à l'intérieur d'un système conçu pour la transcription de milliers de tours de parole (cf. [ChanierEtal06]).

5.2. Primauté de l'oral et de la modalité audio

Dans un environnement audio-synchrone, le mode parole prime tous les autres. C'est la raison pour laquelle, avant le début de la formation, chaque participant vérifie en se connectant au serveur de la plate-forme que son système audio fonctionne correctement. Cette réussite conditionne sa participation ultérieure. De plus, chaque séance débute par le rituel consistant à vérifier que chacun entend correctement les autres et est également entendu en retour. La modalité de vote permet de préciser la qualité de sa réception. Cette dernière modalité est fréquemment utilisée, par l'enseignant qui a régulièrement besoin de vérifier la bonne compréhension de chacun. Le vote offre une réponse claire, qui engage l'apprenant. L'absence de réponse se décèle aisément. On notera la différence avec la situation de la classe en présentiel, où l'enseignant cherche des indices de compréhension en croisant rapidement le regard des apprenants, voire formule une demande de confirmation orale en s'adressant à tout le groupe. Les retours correspondants sont alors peu aisés à interpréter ou vérifier.

Mais si important que soit le mode parole dans ce type d'environnement, il ne saurait avoir le monopole. Il a besoin d'être soutenu par d'autres modes et modalités.

5.3. Jeu entre modalités qui viennent soutenir l'oral

L'exemple sélectionné (figure 2) au milieu de nombreux autres, offre une bonne illustration des interactions des modes iconique et textuel avec le mode parole. Il s'agit de la première séquence de la seconde session du groupe des plus avancés (R). Le premier tour de parole de l'extrait survient après 18 minutes et 40 secondes de conversation. Dans un but de simplification, nous avons retiré de l'extrait les silences.

Attardons-nous sur le contenu de ces échanges. AR4 explique en réponse à une sollicitation du tuteur (TutR) qu'il est, le week-end dernier, allé écouter un concert des *Enfoirés*. Il occasionne alors une interrogation de nature très culturelle car les deux non-natifs (TutR et AR1) ne connaissent pas ce rassemblement d'artistes, qui pourtant représente l'événement musical le plus populaire en France. Les réponses provenant du vote confirment la connaissance des natifs (français) et l'ignorance des autres. Pendant ce temps, AR6 utilise le clavardage pour donner une explication supplémentaire. AR4 indique alors qu'il va essayer de traduire le nom du groupe. Un recouvrement, sur les deux tours de parole adjacents, apparaît avec le début de l'intervention du tuteur qui demande à l'apprenant s'il a apprécié ce concert. Les natifs continuent à donner, dans le clavardage, des précisions sur l'événement et le groupe, pendant que AR4 se livre à la traduction promise, à savoir *restaurant of the heart* (qui correspond aux *Restos du Cœur*). L'apprenant non-natif (AR1) exprime son accord en utilisant la même modalité, le clavardage, ce qui clôt l'échange sur les *Enfoirés*. Le tuteur demande ensuite à cette même personne ce qu'elle a fait de son côté (audio à t = 19:33). Son explication orale sur son travail au restaurant pour gagner sa vie suscite une reprise du tuteur de la traduction des *Restos du Cœur* (aud88). Il y a donc reprise en mode parole d'une information tapée dans le clavardage avec un glissement sémantique, exprimant ainsi la compassion du tuteur pour la

situation de AR1. Cette intervention provoque le rire de l'apprenant et une reprise avec *heart*, qui est aussi acceptable dans la situation, ce que concède le tuteur (aud90) en marquant à nouveau sa compassion (sur un mode ironique, bien sûr).

Le schéma au bas de la figure 2 fait ressortir la structure du discours. Chaque tour de parole y est représenté. Un tour de parole peut se réduire à un seul acte. C'est le cas systématiquement pour les votes ou, par exemple, pour la question du tuteur (aud80). D'autres tours incluent plusieurs actes, comme celui de AR4 qui juxtapose réponse et question (aud79). Les accolades délimitent les échanges. Ainsi l'échange ouvert par la question de AR4 posée à l'audio est clos par l'acceptation de AR1 dans le clavardage (clav27). Il est intéressant de noter le fait qu'un échange peut trouver sa conclusion dans une autre modalité que celle utilisée dans le tour de parole d'ouverture. On en retrouve un autre exemple dans l'échange ouvert à l'audio par le tuteur (aud83) et clôt par AR4 à l'aide du vote (vot27). On notera également l'économie représentée par cette dernière modalité (et plus généralement le mode iconique) qui, ne perturbant pas le flux audio, permet aussitôt au tuteur de rouvrir un autre échange en s'adressant à AR1 (aud85) ... qui vient de conclure dans la troisième modalité, comme nous venons d'en parler. Participer au discours implique bien la nécessité pour chacun de suivre ce qui se passe dans les trois modalités ! Les relations entre elles ne manquent pas, comme le symbolisent les flèches du schéma.

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
19:33	aud85	TutR	ok how about AR1↑ what did you do this week ↑ +

19:37	aud86	AR1	euh so this week euh I worked euh very hard I + euh sometimes I worked in a restaurant + euh it's a restaurant of my friends euh + where there are euh so where they euh there are a lot of people + euh I helped them
20:04	aud88	TutR	it's not the euh the restaurant of the heart ↑ AR4 + has talking about then ++
21:00	aud89	AR1	<rires> no ↑ <rires> it's not the restaurant of the heart + no +
21:08	aud90	TutR	(pity)

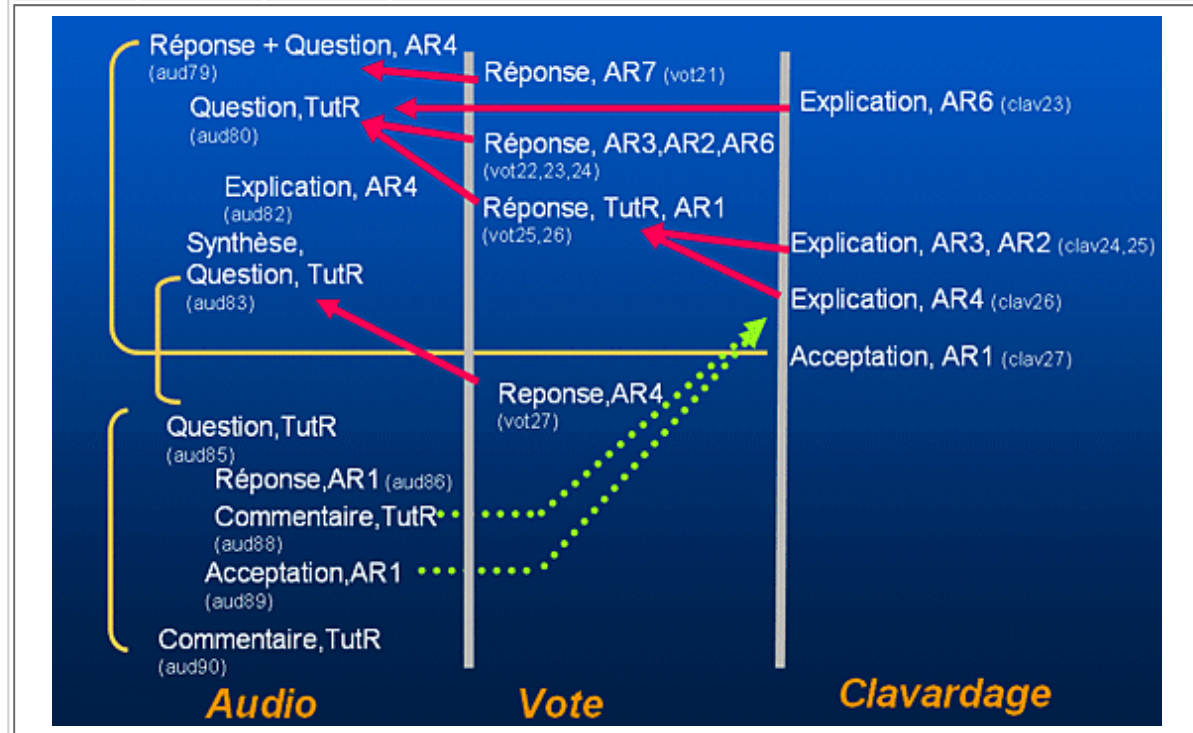


Figure 2 - Extrait de transcription et (image) mise en évidence des liens entre les tours de parole et les actes de langage dans les trois modalités.

Une flèche pleine, apparaissant en rouge, indique que le tour de parole à l'origine de la flèche, réalisé dans une certaine modalité, est dans une relation de "dépendance conditionnelle", aussi appelée "séquentielle" ([Bange92] : 44) avec le tour pointé à son extrémité, lui réalisé dans une autre modalité. C'est le cas, par exemple de AR7 lorsqu'il^[2] répond (vot21) à AR4 (aud79). La notion de "paire adjacente" du discours monomodal doit être réinterprétée dans le contexte multimodal. D'une part parce que la paire précédente se déroulait dans deux modalités distinctes et, d'autre part, parce que la question du tuteur (aud80) suscite des réactions immédiates de plusieurs locuteurs dans plusieurs modalités (vote et clavardage), certaines pouvant être interprétées comme des réponses, d'autres comme des explications. Le tuteur participe d'ailleurs lui-même à une réponse aux côtés des autres dans le vote (vot25). Cet ensemble de votes génère à son tour des explications multiples dans le clavardage.

La deuxième catégorie de flèches du schéma, apparaissant en pointillés (en vert), symbolise une référence faite par un locuteur dans un tour de parole réalisé dans une certaine modalité au contenu d'un autre tour d'une modalité différente. C'est le cas de TutR et AR1 (aud88 et

aud89), qui reprennent à l'audio ce qui est apparu dans le clavardage à l'initiative de AR4 (clav26). On a donc ici un exemple d'une nouvelle forme de polyfocalisation (cf. section 2.2), qui survient dans un même environnement.

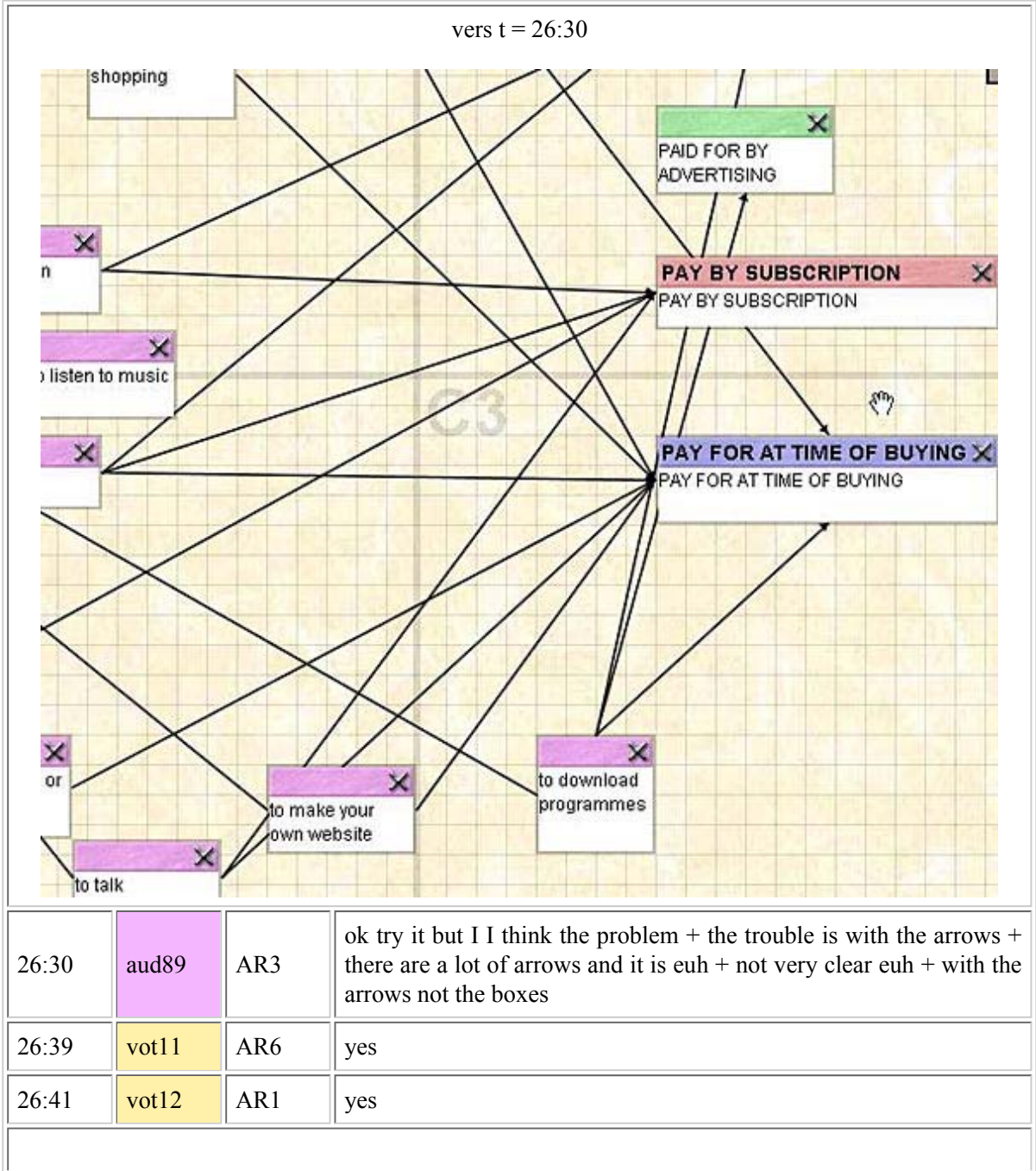
Cet exemple est représentatif des échanges mêlant les trois modalités, audio, clavardage et vote. Les votes indiquent clairement l'état de compréhension des participants et va susciter un échange de clarification sur le clavardage. L'usage de cette modalité permet de ne pas interrompre le fil du discours oral et de fixer par écrit le sens de ce que l'on veut dire, en limitant les ambiguïtés de l'oral. Ce deuxième fil de discours dans le clavardage s'opère au vu de tous, ce qui permet au tuteur d'y faire allusion dans le fil du discours oral. Le terme "fil de discours" est peut-être un peu abusif, car tous les thèmes abordés dans chacune des modalités sont convergents. Dans cet exemple, il n'y a qu'un seul polylogue construit par l'ensemble des participants. Ce contenu thématique unique, décliné dans les trois modalités, requiert bien une polyfocalisation de chacun des participants mais limite les surcharges cognitives, ce qui ressort bien des enchaînements rapides des tours de parole avec références croisées. On mesure ainsi toute la richesse de ces jeux de langue qui peuvent s'opérer grâce à la multimodalité et aussi l'apport des deux modalités vote et clavardage à l'audio.

5.4. Quand modes verbaux et non verbaux se répondent

Abordons maintenant un exemple (figure 3) dans lequel modes verbaux (parole) et non verbaux (graphique) interagissent. Lors de la sixième session du groupe des plus avancés (R), quatre apprenants travaillant dans une salle annexe ont reçu pour consigne de construire une carte conceptuelle déclinant les usages de la Toile, en distinguant les usages gratuits de ceux donnant lieu à des formes différenciées de paiement. C'est une tâche complexe et ce à deux niveaux différents : d'un côté, le classement des usages de la Toile dans une catégorie ou une autre va donner lieu à des débats nourris et à l'expression de désaccords entre participants sur ce qui peut être considéré comme payant ou gratuit ; d'un autre côté, l'augmentation du nombre de nœuds et de liens dans la carte conceptuelle, associés à des couleurs différentes pour visualiser les typologies d'usage, constituent rapidement un réseau dense et peu lisible. Le groupe saisissant la perte de lisibilité (préjudiciable à la présentation programmée de la carte en plénière, cf. la première image de la figure 3) décide de réorganiser l'ensemble. Suivant les habitudes du travail collaboratif, l'un des apprenants se propose de coordonner le travail. C'est ce qu'exprime AR3 en aud89, qui reçoit rapidement l'assentiment des autres au moyen du vote (cf. début de la transcription en figure 3). AR3 se livre d'abord seule à ce travail en suivant les conseils du groupe, qui continue à discuter oralement.

L'extrait se poursuit huit minutes après son début avec une nouvelle proposition de réorganisation qui, impliquant trop d'actions différentes, va provoquer une réaction de tout le groupe. Celui-ci décide alors de se répartir le travail. En aud121, AR6 propose d'agir à la suite de AR3. Puis pendant le silence de 191 secondes (aud122), tout le groupe passe à l'action pour aboutir à la carte dont la deuxième image de la figure 3 donne une vue partielle. Suite à cela, l'un des membres exprime sa satisfaction (aud125), le coordinateur acquiesce, suivi des autres.

Ces approbations closent les transactions ouvertes en aud122 et aud89. On notera au passage que le fait de donner à chaque silence audio un statut de tour de parole est la condition nécessaire pour assurer la continuité de transcription du flux audio, tout en permettant d'établir un rapport avec ce qui est fait simultanément dans les autres modalités. Notre méthode de recueil et de transcription des données nous a permis de calculer les temps de prises de parole audio de chaque participant en mesurant le temps de parole en continu et le nombre de tours de parole. Ce temps en continu est obtenu en isolant les silences entre les tours de parole, silences que l'on peut classer en deux catégories. Dans la première, on trouve les hésitations entre deux tours de parole des participants, phénomènes que nous distinguons des pauses incluses dans un même tour de parole. Le deuxième type de silence s'explique par la concentration des acteurs dans des tâches mobilisant des outils de production collective de la plate-forme. Ainsi, lorsque les apprenants sont occupés à lire ou à modifier une production écrite, on constate des "blancs" à l'audio.



[...]			
34:04	aud120	AR3	yes we have we also have to euh double + double some boxes for example the box with listen to music we keep it and we we do it again in yellow and we add it + for free + I will + I do it +
34:06	vot15	AR1	yes
34:22	aud121	AR6	ok I will do the boxes for + pay for at time for buying
34:29	aud122	sil	191
	AR6 déplace les boites bleues + AR3 s'occupe des boites jaunes + AR1 et AR4 participent aux déplacements.		
[...]			
vers t = 39:00			
39:00	aud125	AR4	I think now it's very nice and very useful + don't you ↑ ++
39:07	aud126	AR3	yes at least we can read it and it is very clear + it is big but it is clear +
39:15	aud127	AR6	yes + it's better like that +

Figure 3 - Extrait de transcription avec, l'image de la carte conceptuelle à deux instants de son élaboration.

Le groupe a donc géré l'organisation et la répartition des tâches dans la "modalité carte conceptuelle" et la modalité audio. Pour être plus précis, les deux modalités (et les deux modes correspondants) construisent un discours multimodal unique dans lequel les transactions sont composées d'échanges qui mêlent tours de parole audio et actions dans la carte, d'une façon un peu comparable à l'occurrence de paires adjacentes dans un discours à modalité verbale unique.

L'exemple, tel que reporté dans la figure 3 pourrait peut-être laisser croire, à tort, qu'existerait une forme de séquentialité et d'indépendance entre modes verbaux d'un côté (représentés par les modalités transcrites dans la figure 3) et le mode graphique de la modalité carte conceptuelle. Cela est dû au fait que nous ne pouvons faire figurer ici que l'image finale de la carte et non chacun des actes ayant permis de la construire. Un examen détaillé de la session ferait ressortir le fait que la plupart des actes accomplis dans la carte font l'objet d'une discussion et d'une approbation, l'acte pouvant soit précéder la discussion (proposition dans un échange), soit la suivre (acceptation), soit intervenir au milieu d'une transaction pour exprimer un désaccord (en défaisant, par exemple, ce qui avait été fait), etc. Une retranscription parfaitement intégrée des différentes modalités impliquerait de détailler les actions évoquées pendant le silence aud122 en actes élémentaires, comme nous l'avons fait pour le traitement de texte dans l'exemple de la section 5.5. Ce travail de recherche est en cours.

On notera que pour mener à bien cette tâche complexe, les participants ont eu recours pour la discussion à la seule modalité audio (avec l'appui économique de la modalité vote). Le clavardage n'a à aucun moment été utilisé. Le tuteur, souvent occupé dans l'autre sous-groupe, n'est pratiquement pas intervenu. Peut-être les apprenants se sont-ils sentis suffisamment confiants dans leurs compétences orales pour économiser la gestion d'une autre modalité verbale (pourtant utile par ailleurs) ? Cette question est à rapprocher d'une des observations faites en IHM sur le fait que les utilisateurs n'hésitent pas à utiliser la multimodalité dans des environnements destinés à l'action multimodale, mais qu'ils évitent la multiplication des modes pour exprimer une commande précise ([Oviatt99] : 76).

5.5. Quand la parole perd la voix ou la multimodalité textuelle

En contrepoint à la remarque précédente qui tentait des comparaisons avec la problématique IHM, nous discutons ici un exemple dans lequel, d'une part, l'audio va être écarté au profit du clavardage et, d'autre part, ce mode textuel est en interaction avec une autre modalité de même mode, le traitement de texte. Audio et clavardage semblent donc à nouveau se concurrencer, avec un choix opposé cette fois. Le discours, certes unimodal, mais multimodal reste entièrement dans le verbal. Nous nous éloignons donc des situations IHM, qui se concentrent sur les actions non verbales à accomplir dans une interface.

5.5.1. Une transaction ouverte à l'audio et close à l'audio, mais sans dialogue audio

Dans l'exemple suivant, extrait d'une séance de travail du groupe des faux débutants (T), la transaction est ouverte et close à l'audio, mais sans dialogue audio (figure 4). Le tuteur ouvre à (t = 37:04, aud138) une transaction en donnant pour consigne au groupe d'énoncer dans l'outil traitement de texte, noté TT dans le tableau (la trame de fond de la cellule est en **bleu**) les 10 critères que doit respecter un bon site Internet pédagogique. Il a écrit le premier critère et propose à chacun de continuer ainsi. Son objectif est double : d'une part, obtenir un texte final qui soit acceptable sur le plan de la langue et de la sémantique professionnelle, et d'autre part, faire construire cet écrit collectivement en incitant les participants à dialoguer entre eux. Seul est énoncé explicitement ce deuxième objectif. En formulant cette demande de dialogue entre les participants, le tuteur choisit l'expression **talk to each other**, ce qui semble indiquer qu'il s'attend à ce que la discussion ait lieu sur le mode parole (modalité audio), comme il vient de le faire lui-même^[3]. La transcription montre cependant qu'à la fin de son tour de parole (t = 38:03), plus de 4mn (257 secondes exactement) s'écouleront avant que le silence audio ne soit rompu ... par lui en aud140 !

Les interactions ont pourtant bien eu lieu, mais sous forme textuelle, à la fois dans le traitement de texte et dans le clavardage. Le tuteur le note d'ailleurs dans le clavardage en faisant remarquer que tout le monde est bien silencieux (t = 41:20, clav63), mais travaille dur (t = 41:53, clav64) ! Par ce dernier tour de parole, le tuteur vient en quelque sorte reconnaître que ses demandes initiales ont bien été prises en considération, mais en utilisant une modalité langagière différente. Il clôt ainsi la transaction qu'il avait lui-même ouverte. Il cherche ensuite à ouvrir une nouvelle transaction en utilisant encore le canal audio (t = 42:40, aud140) en demandant au groupe de reformuler le critère 2, suivant en cela son premier objectif, la construction d'un texte cohérent. Les participants continuent leur travail sans avoir apparemment entendu. Le tuteur reformule sa demande en utilisant le mode textuel dans le clavardage (t = 43:12, clav65). C'est dans cette même modalité qu'il clôt cette seconde transaction, très peu de temps après (t = 43:21, clav66), avec un acte de validation du travail accompli. Une partie du groupe l'avait en fait parfaitement entendu et préparait une nouvelle formulation au moment où le tuteur répétait sa demande à l'écrit. Les apprenants avaient reformulé le critère 2 en passant de **to give a question to the author** à **you can contact the author of the web site or a teacher**.

Il y a donc bien eu un refus tacite d'adoption du mode parole par le groupe, mais toutes les consignes ont été respectées : le travail de nature langagière a été accompli collectivement (point 1) et des conversations ont eu lieu (point 2). Le groupe a "seulement" choisi de rester dans le même mode, textuel, pour accomplir ces deux séries d'interactions. Revenons maintenant sur ces deux points.

Point 1 : écriture coopérative

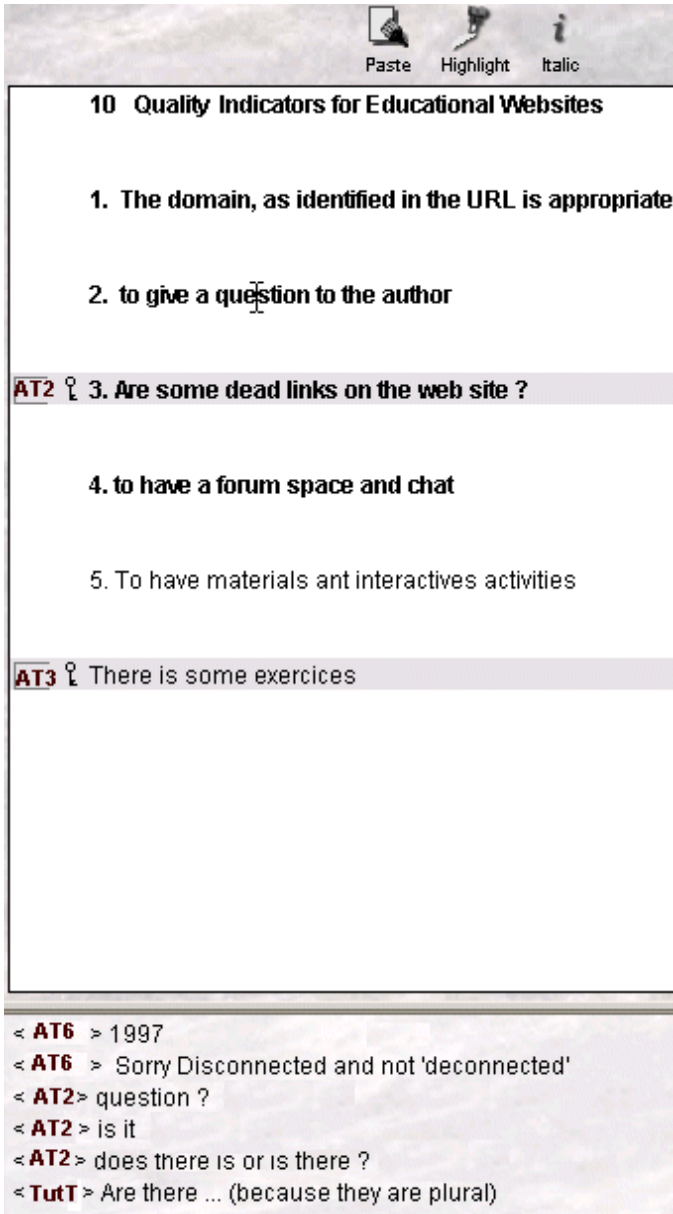
Tout d'abord, le travail (énoncé des 10 critères apparaissant dans le traitement de texte) a été fait de façon entièrement collective, avec la contribution régulièrement croisée de chacun, à tout moment un scripteur énonçant un nouveau critère, ou reprenant la formulation d'un critère précédemment écrit par quelqu'un d'autre. Plusieurs scripteurs ont même travaillé simultanément, ce qui crée alors une situation de polyfocalisation d'un autre type. L'image de la figure 4, illustre ce point. On y voit une partie de l'écran *Lyceum* à (t = 40:25). À cet instant, on observe l'interaction simultanée de AT2 et AT3 dans le traitement de texte. Comme le montre la transcription, non reproduite dans l'extrait de gauche, AT3 s'apprête à modifier un énoncé initialement commencé par AT2, alors que AT2 revient corriger le critère 2 dont elle était le premier auteur.

Point 2 : une conversation fractionnée où clavardage et traitement de texte se répondent

Il est intéressant de revenir sur l'historique de cette correction du critère 3 car elle a donné lieu à une conversation fractionnée dans le temps (et multimodale) entre le tuteur et AT2. Ceci illustrera le point 2.

AT2 à (t = 39:08, clav57) ouvre une transaction en formulant sur le clavardage une demande d'aide étalée sur trois tours de parole. Elle s'interroge sur la formulation de l'interrogative du critère 3 qu'elle avait tapée dans le traitement de texte à (t = 38:22, *Does there is a some dead links on the website ?*). Ce critère se rapporte à la qualité des liens Internet, vivants ou morts, d'un site. Puis elle continue à saisir d'autres critères (t = 39 et seq.). À (t = 40:17, clav60), le tuteur lui répond dans la même modalité, le clavardage, en donnant la formulation correcte et en pointant l'erreur de nombre : *Are there... (because they are plural)*. L'apprenante, qui était bien en attente de réponse tout en poursuivant son travail, saisit la remarque apparue sur le clavardage pour aussitôt abandonner sa tâche d'écriture courante et revenir modifier ce critère 3. Après deux modifications successives, ce critère apparaîtra sous la forme correcte *are there some dead links on the website?* (t = 40:27). Le tuteur poursuit le dialogue dans le clavardage par une invite de passage au mode actif et non interrogatif et en introduisant un adjectif plus précis (t = 40:59, clav62) : *or: All the links are reliable*. Ce dernier échange dans la transaction ouverte par AT2 restera lettre morte.

37:04	TutT	aud138	[...] if you could just note down + ten ++ aspects + of a website that tell you it may be + it should be + a good + educational + site ++ so I've started personally + with one [...] + so do you want to carry on and fill in the next ↑ + fill in the rest ↑ +++ talk to each other to work out what should be in there
38:03	sil	aud139	257
[...]			
38:22	AT2	TT	3. Does there is a some dead links on the website ?

		(ajout)	
39:01	AT3	TT (ajout)	4. to have a space for a forum space
39:08	AT2	clav57	question ?
39:14	AT2	clav58	is it
39:24	AT2	clav59	does there is or is there ?
39:38	AT2	TT (modif)	4. to have a forum space and chat
39:38	AT4	TT (ajout)	5. To have materials ant interactives activities
[...]			
40:17	TutT	clav60	Are there ... (because they are plural)
40:11	AT2	TT (ajout)	There is some exercices
			

40:27	AT2	TT (modif)	3. Does Are there is some dead links on the website ?
[...]			
40:39	TutT	clav61	e.g. there are no dead links
40:59	TutT	clav62	or: All the links are reliable
[...]			
41:20	TutT	clav63	you are all very quiet (me too...)
[...]			
41:53	TutT	clav64	But you are working very hard : -)
[...]			
42:40	TutT	aud140	ok hum + just to say I don't fully understand I don't completely + understand + point + two +++ can we have a think how we can + reformulate that ↑ ya ↑
42:57	sil	aud141	50
[...]			
43:12	TutT	clav65	can we reformulate point 2
[...]			
43:21	TutT	clav66	brilliant!

Figure 4 - Extrait de transcription avec ajout des actes effectués dans le traitement de texte ; l'image est un aperçu du traitement de texte et du clavardage à (t = 40:25).

Cet exemple met en valeur l'intérêt linguistique de l'outil traitement de texte et sa complémentarité avec le clavardage. En autorisant la sauvegarde de textes sous une forme permanente, il invite à la construction d'un discours plus soutenu que celui éphémère du clavardage. Il offre un système de production de textes hautement collaboratif qui brise la linéarité de l'écriture, invite à l'écriture à plusieurs mains, en visualisant qui tient "les plumes". Cette écriture simultanée sur des lignes disposées de façon non consécutives spatialement, par une succession d'ajouts ou de modifications d'énoncés aux auteurs multiples, crée un nouvel espace d'interactions scripturales qui ne suit plus la linéarité temporelle ou spatiale des énoncés produits avec les outils de communication audio ou clavardage. Ces nouvelles "conversations" peuvent ou non se doubler d'un dialogue plus conventionnel dans ces deux systèmes de communication.

5.5.2. Pourquoi privilégier un mode plutôt qu'un autre ?

La non-utilisation du mode parole par les apprenants peut s'expliquer de plusieurs façons. Le groupe était engagé dans une tâche prenante de nature textuelle dans le traitement de texte. Rester dans ce mode unique évitait sans doute une surcharge cognitive. Deux modalités étaient utilisées (traitement de texte et clavardage), le clavardage permettant de gérer un

dialogue ayant certains des traits de l'oralité, tout en demandant moins d'efforts en production que le mode parole. La multimodalité n'est donc pas rejetée mais la tâche (purement langagière, cette fois, au contraire de l'exemple avec la carte conceptuelle) devenant complexe, les apprenants semblent avoir fait l'économie de modalités et de modes qui peuvent être considérés comme concurrentiels dans cette situation sur le plan illocutoire (on peut faire les mêmes choses en les énonçant dans l'un des deux modes).

Si les apprenants se satisfaisaient de ces deux modalités textuelles, le tuteur a semble-t-il été impressionné par le silence audio, qui s'est prolongé au-delà de l'exemple reporté en figure 4. À tel point même que l'énoncé des dix critères étant terminé, il a cru bon de demander à chacun, avant de poursuivre la session, si leur système audio marchait bien. Pourtant tout au long de cette transaction, les apprenants ont montré régulièrement que, non seulement, ils entendaient bien le canal audio mais qu'ils écoutaient également le tuteur en dialoguant dans le clavardage ou en interagissant dans le traitement de texte.

5.5.3. Polyfocalisation et perte d'orientation

L'impression de "parler dans le vide" ou de solitude du tuteur pouvait venir du fait que dans ces espaces multimodaux, les énoncés ne contiennent souvent pas le nom du ou des destinataires [4]. Dans un groupe ayant l'habitude de travailler ensemble à distance, une requête peut d'ailleurs être formulée à tous en espérant que la première personne disponible et compétente pourra y répondre. Un participant doit donc souvent inférer de l'observation des tours de parole de chacun et des interactions dans différentes modalités s'il a bien été entendu, voire écouté. Cette situation de polyfocalisation peut, si l'on y est peu habitué, introduire le doute, voire une perte d'orientation.

C'est peut-être comme une volonté de réintroduire des repères explicites dans la conversation que l'on peut interpréter, au second degré, les deux interventions de validation du tuteur que nous avons déjà citées (clav63 et clav66). Au premier degré, celui-ci accomplit des actes pédagogiques et sociaux rituels en validant le travail accompli et en soutenant positivement les participants (actes reconnus comme indispensables par les apprenants lors des entretiens post-formation). Au second degré (qui représente la valeur perlocutoire de l'acte de langage), il clôt lui-même, explicitement, les transactions qu'il a ouvertes.

5.6. Le pseudo-mode spatial ou se repérer dans les différents espaces à la voix

Il serait abusif de conclure à partir de l'exemple précédent que la modalité traitement de texte s'accommode mal du mode parole, ni même que les apprenants cherchent dans tous les cas à minimiser le nombre de modalités. Nous allons voir dans un nouvel exemple, d'une part, que le mode parole peut devenir prévalent y compris lorsque le traitement de texte est utilisé et, d'autre part, que les apprenants n'hésitent pas à alterner les modalités, en particulier spatiales. Mais il convient peut-être avant tout de s'expliquer sur l'expression "mode spatial".

5.6.1. Distinguons trois échelles spatiales

On se souvient que Kress et Jewitt [KressJewitt01] placent le "mode action" au centre de leurs études sur le présentiel. Pour ce qui est de l'environnement audio-synchrone de la plate-forme *Lyceum* nous pouvons distinguer grossièrement trois "échelles" spatiales différentes :

- les fenêtres de l'écran de l'ordinateur de l'utilisateur (non compris celle de la plate-forme), ainsi que l'environnement matériel de l'utilisateur ;
- les différentes salles de l'environnement de la plate-forme ;
- les différents cadres d'une même salle.

Dans la première dimension de l'échelle, il peut s'agir d'une fenêtre de navigateur dans laquelle l'utilisateur examine un site Internet ou fait appel à un moteur de recherche pour trouver la définition d'un mot ou d'une expression, à moins qu'il ne décide d'utiliser le dictionnaire papier posé à côté de son clavier. À ces exemples, ajoutons le recours à des systèmes de clavardage extérieurs par des apprenants désireux de mener des apartés pendant la séance principale *Lyceum*, autre exemple de polyfocalisation, cette fois proche de la situation évoquée par [Jones04]. Tous ces exemples ont été vécus à de multiples reprises dans la formation *Copéas*, chacun augmentant l'espace de la plate-forme de ses lieux et outils familiers.

La seconde dimension de l'échelle recouvre l'élément fondamental de structuration à l'intérieur de la plate-forme : un utilisateur, à l'ouverture de sa connexion, arrive dans le hall (*lobby*), lieu de prédilection pour les rassemblements de groupe et les rituels de salutation avant de partir travailler par groupe entier dans une salle particulière ou encore de se disperser en sous-groupes dans des salles différentes. Les autres modes dépendent de cette organisation spatiale : les "portes" entre salles sont closes et épaisses, on ne s'entend donc pas d'une salle à l'autre ; en revanche, on peut voir à chaque instant qui est où. Méthodologiquement, l'enseignant devra d'abord scénariser ses séances de formation à ce niveau, en décidant à quel moment et où chacun travaillera en groupe ou en sous-groupe. Cette structuration n'empêche pas de se perdre momentanément, ce qui génère des dialogues où on retrouve bien toute la déictique spatiale (cf. Tableau3).

aud116	AR5	AR1 are you alone ↑ + where are the others ↑ ++
aud117	AR1	what ↑ ++
aud118	AR5	are you + alone ↑
aud119	AR2	they are in the lobby + look ++

Tableau 3 - Conversation audio tenue par des apprenants présents dans une même salle.

Dans la troisième dimension de l'échelle, on trouve la surface principale de l'écran

compartimentée en cadres, certains étant présents en permanence sans recouvrement (sur la figure 1, cadre 2, il s'agit des espaces dédiés aux outils de communication directe), d'autres se recouvrant et étant dédiés aux outils de production collective (figure 1, cadre 3 ; carte conceptuelle, tableau blanc, traitement de texte). Il peut s'ensuivre des quiproquos savoureux lorsque certains parlent d'éléments d'un outil de production qu'ils ont sous les yeux, alors que les autres en regardent un second. L'environnement *Lyceum* offre cependant tous les éléments de repérage nécessaires pour savoir qui utilise quoi à un instant donné et même la commande *Gather*, que nous avons évoquée, pour rassembler tout le monde dans un même outil^[5].

À l'énoncé de ces éléments, on saisit toute leur importance et leur influence sur les interactions, dont beaucoup apparaissent sous forme de traces verbales. On devine l'accès à des ressources linguistiques qui peut avoir lieu ainsi. D'un certain point de vue, on peut en parler comme du "contexte", par contraste avec le "texte" qui s'écrit dans les outils et modalités plus verbales :

Perhaps the biggest barrier to a useful understanding of the context in which computer-mediated communication occurs is the tendency for the attention of analysts of CMC to stop at the screens edge, for people to regard "virtual realities" and "material realities" as separate things. ([Jones04] : 24).

Jones parle ainsi de la première dimension de l'échelle, que nous venons de distinguer. On peut aussi remarquer que chaque dimension spatiale a sa propre signalétique, son propre code de signes et d'actions qui interfère directement dans la communication. À ce niveau de considération, on peut parler d'un "pseudo-mode spatial", qui se déclinera dans différentes modalités de communication. Ainsi dans le tour de parole audio aud119 du tableau 3, ne faut-il pas comprendre l'énoncé *look* comme s'accompagnant d'un geste de pointage (virtuel) vers un signe visible de tous à l'écran, signe qui permet de savoir où sont les autres ?

5.6.2. Mode, pseudo-mode et modalité

L'exemple détaillé dans l'annexe 2 va nous permettre de pousser plus loin notre réflexion. Il devient nécessaire d'identifier les contextes spatiaux dans lesquels se construisent les discours. Plusieurs commentaires s'imposent après examen des transcriptions. Les apprenants du groupe des faux débutants ont de leur propre chef décidé d'accomplir une tâche complexe en changeant à plusieurs reprises de lieu (espace de troisième dimension) afin d'aller voir des ressources destinées à alimenter la discussion dont le thème a été fixé au départ. C'est un fait surprenant. En effet, un tuteur y réfléchirait souvent à deux fois avant de faire changer ces interlocuteurs d'espace en cours de conversation, avec le risque d'en perdre en route. Nous avons d'ailleurs observé, en transcrivant les huit sessions des deux groupes, des négociations tuteurs-apprenants lors desquelles ces derniers désiraient éviter d'avoir à préparer des plénières en jonglant sur plusieurs lieux. On notera que ces déplacements nécessitent une bonne collaboration entre les interlocuteurs, collaboration qui s'observe dans les dialogues.

La modalité audio est la seule modalité verbale à être partout présente dans une même salle. Elle n'accapare pas la vue des apprenants qui peuvent concentrer leurs regards sur les outils de production collective. Ceci explique sans doute pourquoi les apprenants, devant limiter la complexité de la tâche, ont ignoré le clavardage et choisi la production orale, pourtant linguistiquement plus demandeuse. Au contraire des situations observées en IHM, ils n'ont pas cherché à simplifier l'usage de la langue. En ignorant le clavardage, ils ont certes ignoré les remarques lexicales du tuteur, mais ils se sont corrigés à l'oral sur des points de prononciation, comme en témoigne la fin de l'extrait.

Les lieux interviennent constamment dans notre extrait. Les locuteurs font directement allusion à tel ou tel lieu (lorsqu'un interlocuteur demande aux autres d'aller dans un certain espace, verbalise une action propre à un lieu spécifique, etc.). Ils se déplacent d'un endroit à un autre. Enfin, ils peuvent discuter de la pertinence d'une définition dans un espace (le traitement de texte) qui reprend des informations aperçues dans un autre (le tableau blanc). Nous avons parlé de "pseudo-mode" spatial et non de "mode", suivant en cela Kress et Jewitt, du fait que l'environnement n'offre aucune modalité reposant sur le spatial. Pour passer d'un espace à l'autre, il faut par exemple, recourir à la modalité souris en appuyant sur un bouton. Les participants ne peuvent pas accomplir un geste, en utilisant par exemple un gant comme dans certains environnements IHM, où le mode spatial a droit de cité en tant que tel.

Cependant il est nécessaire de chercher à coder dans les transcriptions ce pseudo-mode spatial. En effet, les différents lieux interagissent constamment avec le discours, non seulement comme contexte conditionnant les interprétations, mais aussi parce que les interactions avec les modes verbaux sont fréquentes. Nous avons commencé ici à le faire. Il faudrait aller plus loin et représenter les autres niveaux de lieux : la salle courante où est assurée la continuité de la modalité audio, les autres salles pour les besoins de co-référence, etc. Pour terminer, signalons que notre codage agit bien au niveau du pseudo-mode et du contexte. Dans la partie gauche de la figure 5 de l'annexe 2, on désigne un niveau de lieux avec les personnes qui s'y trouvent et cet espace englobe un ensemble de tours de parole (échanges et transactions), sans qu'il soit possible d'introduire des actions spatiales qui interviennent au niveau des tours de parole.

6. Conclusions

Dans cet article, nous nous sommes intéressés à l'étude du discours multimodal à l'intérieur d'une plate-forme audio-synchrone dans un contexte pédagogique d'apprentissage des langues en groupes.

En citant le regain d'intérêt des chercheurs de sciences du langage pour le discours multimodal et la voie ouverte 15 ans auparavant par ceux du domaine des interactions humain-machine, nous avons fixé les acceptions des termes "média", "mode" et "modalité". Nous avons défini ce que nous entendions par "plate-forme synchrone" en illustrant les trois composants fondamentaux, le composant spatial, les outils de communication et les outils collaboratifs de

production. Le choix d'une telle plate-forme étant un choix stratégique au niveau d'un établissement, nous avons proposé des critères d'appréciation et un classement des 12 fonctionnalités principales regroupées par composant et classées par degrés d'utilité et enfin un tableau de correspondance entre outils de communication et de production et objectifs discursifs.

Le rappel des caractéristiques de la formation en anglais autour de laquelle s'est construit le projet de recherche *Copéas* nous a permis de montrer tout l'intérêt de ces environnements technologiques en apprentissage des langues. Nos études récentes nous ont permis de mesurer précisément les productions langagières de deux groupes d'apprenants fortement hétérogènes. L'étude comparative de l'usage combiné de l'audio et du clavardage indique l'effet d'égalisation de ces deux modalités, égalisation qui se mesure particulièrement dans le bon niveau de participation du groupe des faux débutants. Ces mesures, accomplies sur un nombre limité de sujets, demanderaient bien sûr à être confirmées. Mais elles ne constituent pas le centre de cet article et n'ont été synthétisées ici que pour mieux introduire deux modalités de base de l'environnement d'apprentissage et mieux contextualiser notre propos.

Dans la partie centrale de l'article, nous avons illustré comment, au-delà de ces deux modalités verbales, c'est surtout la possibilité offerte à chaque apprenant de s'exprimer dans différents modes et modalités qui a fortement soutenu la production verbale en L2.

Ce parcours a commencé par l'examen d'un exemple prototypique dans lequel un discours se construit par imbrication de tours de parole effectués dans plusieurs modalités, en l'occurrence audio, clavardage et vote, l'imbrication s'observant dans la construction des échanges et les phénomènes de références. Le second exemple a étendu ce phénomène d'imbrication en mettant en correspondance modes verbaux et non verbaux. Le troisième exemple a opéré un retour sur les modes verbaux et illustré de façon surprenante comment l'audio, malgré sa prééminence peut perdre la voix et laisser la place à un discours construit sur des modalités uniquement textuelles. Le traitement de texte perd à cette occasion son statut de simple présentoir pour devenir un outil interactif aux côtés du clavardage. Dans le dernier exemple nous avons élargi le champ de description du discours multimodal en considérant le mode spatial qui s'impose comme le contexte dans lequel chaque transaction (et ses sous-composants) doit être située. Nous en avons parlé en fait comme d'un pseudo-mode, le distinguant ainsi du mode spatial rencontré en IHM, et en avons décrit les trois dimensions principales. Nous avons observé l'influence des deuxième et troisième dimensions sur le discours.

Lors de la présentation de ces exemples de discours multimodaux, deux préoccupations principales nous ont guidés.

La première, de nature pédagogique, explique l'insertion de remarques sur la scénarisation et l'animation d'activités pédagogiques dans de telles plates-formes. L'occurrence de phénomènes variés de polyfocalisation dans le discours a retenu notre attention. Pour les deux

groupes d'apprenants étudiés ici, quel que soit leur niveau linguistique, ces formes de polyfocalisation ne semblent avoir gêné ni leur compréhension, ni leur production. Ces phénomènes ont, en revanche, posé plus de problèmes de suivi au formateur. Cela confirme ce que la littérature de recherche considère déjà comme un nouveau défi pour les enseignants en langue ([Lewis06] ; [Lamy06]). Dans le même esprit, nous avons cherché dans chaque exemple à savoir ce qui expliquait le choix par les interlocuteurs de tels modes ou modalités, plutôt que d'autres. Cette compréhension est aussi nécessaire à la scénarisation et à l'animation pédagogique. Insistons sur le fait que notre exposé ouvre plus de questionnements, destinés à susciter l'intérêt des didacticiens et à servir d'invite à des recherches ultérieures plus ciblées, qu'elle ne prétend apporter de réponses à partir desquelles établir des prescriptions pédagogiques.

Notre seconde préoccupation recouvrait des objectifs linguistiques sur les systèmes de transcriptions du discours multimodal. Nous avons évoqué les limites des méthodes de l'analyse conversationnelle monomodale, et pointé les caractéristiques, les relations, ainsi que la nature des éléments du discours qui devaient être pris en compte. Le travail exposé ici n'est qu'une ébauche que d'autres recherches doivent approfondir afin de dégager les caractéristiques du discours multimodal propre à ces environnements. Dans cette optique, la transcription des actes non verbaux est la condition nécessaire à la mise en correspondance précise du verbal et du non-verbal. C'est un sujet de recherche auquel nous nous consacrons actuellement.

À plus long terme, il faudrait arriver à identifier les situations et activités les plus propices aux apprentissages en L2, sans doute comprendre celles où les tâches conduisent à une complémentarité des modes, ou au choix récurrent d'un mode particulier par l'apprenant (assignation), à la redondance entre les modes, ou à un choix indifférent de tel ou tel mode (équivalence). Ces questions reprennent précisément le modèle CARE (Complémentarité, Assignation, Redondance, Équivalence) conçu en IHM [Caelen00]. À nous de le repenser dans le contexte de ces plates-formes audio-synchrones et de l'apprentissage des langues.

Remerciements

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, ainsi que Laurence Jeannot, Christophe Reffay et Marie-Laure Betbeder de l'université de Franche-Comté. Le projet *Copéas* constitue une partie du projet de recherche ODIL (Outils Didactique pour les Interactions en Ligne) et est, à ce titre, partiellement financé par l'ACI "Éducation et Formation" du Ministère de la Recherche et de la Technologie.

Références

Les liens externes étaient valides à la date de publication.

Bibliographie

[Bange92]

Bange, P. (1992). *Analyse conversationnelle et théorie de l'action*. Paris : Hatier / Didier.

[BellikTeil92]

Bellik Y. & Teil, D.(1992). "Définitions terminologiques pour la communication multimodale". *Conférence*

http://www.limsi.fr/Individu/bellik/publications/ihm92_1.rtf

[Blake05]

Blake, R. J. (2005). "Bimodal CMC : The Glue of Language Learning at a Distance". In *Special Issue on Computer-mediated Communication and Foreign Language Learning : Context, Research, and Practice, CALICO Journal*, vol. 22, n° 3. pp. 497-512.

[Caelen00]

Caelen, J. (2000). "10 ans de recherches en multimodalité(s)". *Colloque sur les Interfaces Multimodales - 09 / 10 Mai 2000*. IRIT. Toulouse.

http://www-clips.imag.fr/geod/User/jean.caelen/transparents_fichiers/Multimodalite2000.ppt

[CalbrisPorcher89]

Calbris, G. & Porcher, L. (1989). *Geste et communication*. Paris : Crédif / Hatier.

[Calico05]

Calico (2005). *Special Issue on Computer-mediated Communication and Foreign Language Learning : Context, Research, and Practice, CALICO Journal*, vol. 22, n° 3.

[ChanierEtal06]

Chanier, T., Vetter, A., Betbeder, M.-L. & Reffay, C. (à paraître 2006). "Retrouver le chemin de la parole en environnement audio-graphique synchrone". In Mangenot, F. & Dejean-Thircuir, C. (dir.) (à paraître).

[Fitze06]

Fitze, M. (2006). "Discourse and Participation in ESL Face-to-Face and Written Electronic Conferences". *Language learning and Technology*, vol. 10, n° 1. pp. 67-86.

<http://llt.msu.edu/vol10num1/fitze/default.html>

[HampelBaber03]

Hampel, R. & Baber, E. (2003). "Using internet-based audio-graphic and video-conferencing for language teaching and learning". In Felix, U. (dir.). *Language Learning Online : Towards Best Practice*. Lisse : Swets & Zeitlinger. pp. 171-192.

[Herring04]

Herring S. C. (2004). "Computer-Mediated Discourse Analysis : An Approach to Researching Online Behavior". In Barab, S. A., Kling, R. & Gray, J. H. (dir.). *Designing for Virtual Communities in the service of Learning*. New York : Cambridge University Press. pp. 338-376.

[Jones04]

Jones, R. H. (2004). "The problem of context in computer-mediated communication". In Levine & Scollon. pp 20-33.

[Kern95]

Kern, R. G. (1995). "Restructuring classroom interaction with network computers : Effects on quantity and characteristics of language production". *The Modern Language Journal*, vol. 79, pp. 457-476.

[Kitade00]

Kitade, K. (2000). "L2 learners' discourse and SLA theories in CMC : collaborative interaction in Internet chat". *ReCALL (Computer assisted language learning)*, vol. 13, n° 2. pp. 143-166.

[KressJewitt01]

Kress, G. & Jewitt, C. (2001). *Multimodal teaching and learning : the rhetorics of the science classroom*. Londres : Continuum.

[KressVanLeeuwen01]

Kress, G. & Van Leeuwen, T. (2001). *Multimodal discourse : The Modes and Media of Contemporary Communication*. Londres : Arnold.

[Lamy06]

Lamy, M.-N. (à paraître 2006). "Conversations multimodales : l'enseignement-apprentissage de l'oral à l'heure des écrans partagés". ". In Mangenot, F. & Dejean-Thircuir, C. (dir.) (à paraître).

[VanLeeuwen04]

Van Leeuwen, T. (2004). "Ten reasons why linguists should pay attention to visual communication". In (Levine & Scollon, 2004). pp. 7-19.

[LevineScollon04]

Levine, P. & Scollon, R. (2004). *Discourse and technology : multimodal discourse analysis*. Washington, DC : Georgetown University Press.

[Lewis06]

Lewis, T. (à paraître 2006). "Teacher autonomy and the online tutor". *CALICO Journal*.

[MangenotDejean06]

Mangenot, F. & Dejean-Thircuir, C. (dir.) (à paraître). *Les échanges en ligne dans l'apprentissage et la formation, Le Français Dans Le Monde - Recherches et applications*, juillet 2006. Paris : Cle International.

[Ortega97]

Ortega, L. (1997). "Processes and outcomes in networked classroom interaction : defining the research agenda for L2 computer-assisted classroom discussion". *Language Learning and Technology*, vol. 1, n° 1. pp. 82-93.

<http://lilt.msu.edu/vol1num1/ortega/default.html>

[Oviatt99]

Oviatt, S. L. (1999). "Ten myths of multimodal interaction". *Communications of the ACM*, vol. 42, n° 11. pp. 74-81.

http://www.cse.ogi.edu/CHCC/Publications/ten_myths_of_multimodal_interaction_oviat.pdf

[PayneWhitney02]

Payne, J.S. & Whitney, P.J. (2002). "Developing L2 oral proficiency through synchronous CMC : Output, working memory and interlanguage development". *CALICO Journal*, vol. 20, n° 1, pp. 7-32.

http://php.scripts.psu.edu/faculty/j/s/jsp17/articles/calico2002/Payne_&_Whitney_02.pdf

[Sotillo00]

Sotillo, S. M. (2000). "Discourse Functions and Syntactic Complexity in Synchronous and Asynchronous Communication". *Language learning and Technology*, vol. 4, n° 1. pp. 82-119.

<http://lilt.msu.edu/vol4num1/sotillo/default.html>

[Stenström94]

Stenström, A. B. (1994). *An introduction to spoken interaction*. Londres : Longman.

[Sykes05]

Sykes, J. (2005). "Synchronous CMC and Pragmatic Development : Effects of Oral and Written Chat". In *Special Issue on Computer-mediated Communication and Foreign Language Learning : Context, Research, and Practice*, *CALICO Journal*, vol. 22, n° 3. pp 399-432.

[ThornePayne05]

Thorne, S. L. & Payne, J. S. (2005). "Evolutionary Trajectories, Internet-mediated Expression, and Language Education". In *Special Issue on Computer-mediated Communication and Foreign Language Learning : Context, Research, and Practice*, *CALICO Journal*, vol. 22, n° 3. pp 371-398.

http://language.la.psu.edu/%7Ethorne/thorne_payne_calico2005.pdf

[Vetter04]

Vetter, A. (2004) "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)*, vol. 7, n° 1. pp. 107-129.

http://alsic.u-strasbg.fr/v07/vetter/alsic_v07_06-pra2.htm

[VetterChanier05]

Vetter, A. & Chanier, T. (2005). "Supporting oral production for professional purpose in synchronous communication with heterogeneous learners". *Communication EUROCALL (EUROpean Computer Assisted Language Learning) 2005 conference*, 24-27 août, Jagiellonian University : Cracovie, Pologne. En cours de soumission à la revue *ReCALL*.

[Warschauer96]

Warschauer, M. (1996). "Comparing face-to-face and electronic discussion in the second language classroom". *CALICO Journal*, vol. 13, n°2. pp. 7-26.

<http://calico.org/journalarticles/Volume13/vol13-2and3/Warschauer.pdf>

Logiciels

[Centra06]

Centra (2006). *Site présentant la plate-forme audio-synchrone Centra One*.

<http://www.centra.com/french/products.asp>

[Lyceum06]

Lyceum (2006). *Site présentant la plate-forme audio-synchrone Lyceum*.

<http://kmi.open.ac.uk/projects/Lyceum/>

Annexes

Annexe 1. Code de transcription

- a. alternance codique : elle est signalée par les accolades : {oui}.
- b. ARN : identifiant du énième apprenant du groupe R.
- c. asN : énième action d'entrée ou sortie d'une personne dans une salle.
- d. ATN : identifiant du énième apprenant du groupe T.
- e. audN : énième tour de parole audio.
- f. Chevauchement : les chevauchements à l'intérieur d'un tour de parole sont insérés dans l'intervention principale, entre les signes < et >, avec le nom de l'acteur, séparé du texte par une virgule : ... you can write + <APx, yes + but + > but.
- g. clavN : énième tour de parole dans le clavardage.

- h. Inaudible : les passages inaudibles ou incertains sont signalés par des xxx.
- i. Intonation : l'intonation montante est signalée par une flèche montante ↑.
- j. Majuscule : pas de majuscules dans le texte sauf exceptions pour les acronymes.
- k. nav : action survenant dans le navigateur.
- l. Pause : les pauses sont signalées par le signe + (1s) ++ (2 s) +++ (3 s).
- m. Prononciation : les problèmes phonétiques sont signalés par des crochets : [first].
- n. R : identifiant du groupe intermédiaire à avancé.
- o. Silence : si la pause dure plus de 3 secondes, elle apparaît comme un tour de parole audio avec la durée en secondes : sil 5s.
- p. T : identifiant du groupe des faux débutants.
- q. TB : action survenant dans le tableau blanc.
- r. TT : action survenant dans le traitement de texte. Les caractères barrés indiquent le fait que l'utilisateur supprime ce qui a d'abord été tapé.
- s. TutR et TutT : identifiant des tuteurs des groupes T et R.
- t. votN : énième action de vote avec l'icône correspondante.

Annexe 2. Changements d'espaces et modalité clavardage ignorée

L'exemple suivant, emprunté au groupe des faux débutants (T), va nous permettre de pousser plus loin notre réflexion (cf. section 5.6). Il devient nécessaire d'identifier les contextes spatiaux dans lesquels se construisent les discours. Les cellules de fond gris du cadre gauche de la figure 5 indiquent combien de personnes (ou qui) se trouvent avec un contexte spatial donné (contexte de première et deuxième dimension suivant la distinction que nous venons d'introduire). Ainsi à (t = 18:33) les trois apprenants, AT1, AT3 et AT6^[6], seuls dans la salle ont devant les yeux la fenêtre avec le traitement de texte (TT) dans le cadre principal (noté 3 dans la figure 1) et une autre fenêtre ouverte sur leur navigateur sur le site de formation en langue *English Club*. Cette dernière fenêtre, au contraire de l'écran de la plate-forme, n'est pas partagée. Des parties différentes de la Toile peuvent donc être affichées dans le navigateur de chacun. Nous avons noté cette situation spatiale TT et nav (3) dans le cadre gauche de la figure 5. Ils doivent répondre dans le traitement de texte à un questionnaire préparé par le tuteur. Le cadre droit de la figure donne un aperçu de ce questionnaire, qui devra ensuite être montré et commenté à tout le groupe en plénière. Les expressions en italique sont les réponses que les apprenants composent au fur à et à mesure de leurs discussions.

A (t = 19:45), AT6 quitte cet espace pour aller travailler dans le tableau blanc, noté TB dans le

tableau (la trame de fond de la cellule est en vert). Puis revient à (t = 23). La discussion aborde alors le point 2 du questionnaire (*The best of them is*) en aud34. Pour y répondre AT6 invite les autres à "venir" voir ce qu'il a préparé dans le tableau blanc, à savoir une copie d'écran partielle du site où apparaissent les offres de clavardage et de forum offertes par *English Club*. Les trois participants quittent alors le traitement de texte pour regarder le tableau blanc (TT n'est pas fermé, mais la vue sur celui-ci est alors cachée par TB). Un extrait du dialogue montre la satisfaction de AT1 (aud44) qui se propose de donner la définition de l'acronyme ESL (*English as a Second or other Language*), parce qu'il apparaît à de multiples reprises sur l'image extraite du site par AT6.

À (t = 26:49), les trois apprenants et le tuteur, qui vient de les rejoindre, quittent le tableau blanc pour retourner sur le traitement de texte. Ayant lu la définition de ESL, tapée par AT1, les apprenants discutent pour savoir s'il est nécessaire de la faire figurer dans les réponses au questionnaire. Cette transaction se clôt positivement après l'insistance de AT1 (aud61), acceptée par AT3 (aud62). Pendant ce temps, le tuteur saisit l'occasion de taper à plusieurs reprises des informations dans le clavardage en réponse à la question orale posée par AT6 à propos de *mainly*. Ce dernier ne comprend pas cet adverbe, utilisé dans la définition de ESL. Le tuteur aura beau insister par trois fois, personne ne regarde le clavardage. Il quitte la salle à (t = 30:03, as8). La discussion continue ensuite sur le point 3 du questionnaire. On notera une intervention corrective à l'oral de AT3 (aud71) sur une expression mal formulée, toujours à l'audio, par AT1.

18:33	TT et nav. (3)	début du travail du groupe de 3 dans la salle après consignes du tuteur	Interactivity quiz - (It takes two to tango!) ENGLISH CLUB 1. The interactive elements of the English Club site include: <i>Help, Games, Quizzes, Forums, chat room</i> ESL / <i>This mainly used to talk about student learning english while living in a English speaking country</i> 2. The best of them is: <i>Forums, chat room</i> 3. The worst of them is: <i>help only</i>
19:45	TB (AT6)	AT6 part préparer un extrait d'écran du site en cours d'analyse	
23	TT et nav. (3)	puis revient avec les 2 autres	
23:28	AT3	aud34 what do you think it's the best of them ↑	
23:30	sil	aud35 11	
23:41	AT1	aud36 for me the best of them is ++ forum and chat ++ because is euh + is like + euh +++ {non + ouais} +++	
23:59	AT3	aud37 are you ok for this euh answer AT6 ↑	
24:11	AT6	aud39 yes ++ I put euh ++ an information of the whiteboard ++ example of interactivity ++ do you	

			euh do you see euh whiteboard just euh + a moment
[...]	TB (3)		les 3 quittent le traitement de texte pour aller dans le tableau blanc
25:19	AT1	aud44	yes euh + AT6 it's euh + it's good and euh but euh + I think euh + we have to say what is ESL ++ and euh + and and ++ what the definition of ESL ++ I can I can't to do that + on the on the on the document +++
[...] 26:49	TT (4)		les 3 quittent le tableau blanc pour le traitement de texte, AT1 est déconnecté un instant puis se reconnecte, le tuteur les rejoint un moment dans la salle
28:01	AT3	aud52	no I +++ <rires> the ++ the answer of AT1 + is very euh is very long <rires> +
28:14	AT6	aud53	yes ++ it's very long +++
28:22	AT1	aud54	euh + so what is <AT6, what + what ↑> I disconnect <rires> I don't know what but euh {voilà donc} euh ++ ESL euh ++ do you understand that ↑ ++
28:37	AT6	aud55	what's mean euh + mainly ↑ ++ please what's mean mainly ↑
28:45	AT1	aud56	mainly is euh + usually ++
28:48	TutT	clav1	principally
28:50	AT3	aud57	but euh what euh do you write the definition of ESL ↑ it's not the + it's not the question
28:57	TutT	clav2	generally
28:58	AT1	aud58	no because because the + the interactive element + on the whitboard + euh + name ESL games + ESL help + ESL quizz ++

29:15	AT6	aud59	yes +
29:17	AT3	aud60	yes but ++ but it's not important I think +
29:23	AT1	aud61	I euh +++ for me euh ++ it's important ++
29:31	AT3	aud62	ok
29:41	AT3	aud64	do you want to answer at the third ++ euh question ↑
29:44	TutT	clav3	... are both equivalents of "mainly"
29:51	AT1	aud66	euh + yes
30:03	TutT	as8	Le tuteur quitte la salle
30:22	AT1	aud68	I euh + for me it's a games ++ maybe + I don't know <rires>
30:50	AT1	aud70	euh worst euh + of them is euh ++ the more bad +++
30:59	AT3	aud71	yes + it's euh the more bad + but you must say the worst +
31:06	AT1	aud72	yes <rires>

Figure 5 - À gauche, retranscription avec changement de lieu (les silences, à l'exception du premier ont été supprimés par économie de place) ; à droite état du traitement du texte à (t = 31) avec les productions des apprenants en italique.

Notes

[1] Le lecteur se rapportera aux articles précités pour la caractérisation précise de la notion de temps de parole en continu.

[2] Le pronom "il" ne donne aucune indication du sexe du locuteur. Notons que dans la population des apprenants, les nombres d'hommes et de femmes sont équilibrés.

[3] Il peut être intéressant d'ajouter que notre protocole expérimental incluait dans l'étape post-formation, non seulement le remplissage de questionnaires auxquels nous avons fait allusion en section 3, mais également l'interview des apprenants, des deux tuteurs et, enfin, des séances individuelles de reformulations réflexives (*critical event recall*) au cours desquelles un participant verbalisait ses intentions en regardant un extrait de vidéo d'une session auquel il avait participé, extrait que les chercheurs hésitaient à interpréter de façon univoque.

[4] La référence explicite au nom du destinataire existe seulement dans les systèmes de clavardage muni d'une fonctionnalité de conversation privée (ou en aparté).

[5] Ces problèmes de repérage de ce que voit l'autre existent encore à une quatrième échelle plus réduite, à l'intérieur d'un cadre de fenêtre, lorsque l'objet déposé ou produit dépasse les limites du cadre de visualisation. Il ne peut alors être affiché en entier mais peut seulement être déplacé et visualisé partiellement.

[6] Ces trois apprenants sont ceux qui ont abandonné la pratique de la L2 depuis le plus grand nombre d'années (15 à 30 ans) et ont des compétences de niveau A.

À propos des auteurs

Thierry CHANIER s'intéresse à l'apprentissage des langues en environnement informatisé depuis une vingtaine d'années. Il anime à l'université de Franche-Comté un groupe de recherche bi-disciplinaire, dont une partie spécialisée en sciences du langage est intégrée au laboratoire Laseldi, et l'autre en informatique au LIFC. Il est également responsable dans ce même établissement du master professionnel FOAD (Formation Ouverte et À Distance).

Courriel : thierry.chanier@univ-fcomte.fr

Toile : <http://lifc.univ-fcomte.fr/~chanier>

Adresse : UFR SLHS, université de Franche-Comté, 32 rue Mégevand, 25030 Besançon Cedex, France.

Anna VETTER est titulaire d'un doctorat de littérature comparée (1999) portant sur les interférences des codes littéraires et picturaux et d'un DESS (2003) "Ingénierie Pédagogique dans des Dispositifs Ouverts et à Distance"(IPDOD). Elle est actuellement chargée de cours à l'université de Franche-Comté, formatrice de formateurs, conceptrice de formations à distance et conceptrice de contenus pédagogiques en FOAD.

Courriel : anna.vetter@univ-fcomte.fr

Toile : <http://www.supportsfoad.com>

Adresse : 26 grande rue, 25360 Naisey, France.

Date de réception version 1 : novembre 2005 ; version finale : février 2006

Référence de l'article :

Chanier, T. & Vetter, A. (2006). "Multimodalité et expression en langue étrangère dans une plate-forme audio-synchrone". *Apprentissage des langues et systèmes d'information et de communication (ALSIC)*, vol. 9, n° 1. pp. 61-101. http://alsic.u-strasbg.fr/v09/chanier/alsic_v09_08-rec3.htm, mis en ligne le 30/06/2006.



[ALSIC](#) | [Sommaire](#) | [Consignes aux auteurs](#) | [Comité de rédaction](#) | [Inscription](#)

© *Apprentissage des Langues et Systèmes d'Information et de Communication*, juin 2006