

Proposition d'une grille d'analyse d'interactions tutorales dans un dispositif multimodal en ligne.

Vincent, Caroline

Laboratoire ICAR, ENS de Lyon
caroline.vincent@gmail.com

1 INTRODUCTION

L'introduction des technologies dans l'enseignement permet de créer des conditions qui ne seraient pas possibles autrement, tel que le fait de mettre en relation, lors d'échanges synchrones à distance, des apprenants du français situés dans un pays étranger avec des tuteurs locuteurs natifs du français.

L'apprenant se trouve alors ponctuellement comme en « immersion » en France, face à un tuteur dont la culture et la langue sont françaises dans son usage quotidien. Cette impression pour l'apprenant d'authenticité et de réalité (Guichon, Kern, Develotte, 2008) peuvent être des facteurs de motivation et de progrès.

Pour le tuteur inscrit dans une communication multimodale et synchrone que ce soit en audioconférence (Lamy & Hampel, 2007) ou en visioconférence (Wang, 2004), cela s'avère un exercice difficile puisqu'il doit coordonner un ensemble complexe d'opérations à des fins d'apprentissage. Plusieurs canaux de communication sont à la disposition du tuteur (textuel, graphique, gestuel, audio, par exemple) et il va les combiner dans son discours (Develotte, Guichon, Vincent, 2010). Discours d'autant plus important qu'il sera à la fois le support de la communication, mais aussi le support pédagogique. La communication multimodale est ainsi au cœur du processus d'enseignement et d'apprentissage et sur cet aspect que nous centrons notre étude.

Notre corpus est constitué d'échanges à distance synchrones entre des tuteurs français et des apprenants américains, dans le cadre de cours de français. Les tuteurs ont à leur disposition différentes ressources (le dispositif de communication lui-même, mais aussi un blog ou des sites internet par exemple), qui sont elles-mêmes composées de canaux (vidéo, audio, clavardage). Nous voulons comprendre comment les tuteurs vont individuellement gérer les ressources multimodales à leur disposition en menant une interaction pédagogique en ligne. C'est-à-dire comment vont-ils organiser l'espace de leur écran d'ordinateur, quelle modalité vont-ils privilégier, pour quelles raisons et quelles conséquences ces choix auront-ils sur l'interaction et la relation avec les apprenants ? Nous voulons comprendre ce qui les pousse dans le développement d'une « compétence sémio-pédagogique », et les conséquences que cela aura au niveau de l'interaction et de la relation avec leur apprenants.

Ce type de corpus, « fortement multimodal » (Develotte, Kern, Lamy, 2010, Norris, 2004) a déjà donné lieu à des études (Traverso, de Fornel, Kerbrat-Orecchioni,...). Cependant, il nous paraît que la multimodalité est souvent traitée sur un aspect ponctuel (par exemple, la gestualité (Cosnier & Develotte, 2010), l'utilisation des smileys et du clavardage (Marruccia, Anis) et non dans sa globalité, dans le cadre d'une interaction tutorale.

Nous nous proposons donc dans cet article de présenter la grille d'analyse que nous avons mise au point dans un double objectif : d'une part analyser un corpus d'interactions pédagogiques en ligne en prenant en charge la globalité de leur multimodalité, et d'autre part participer à la réflexion sur la méthodologie des corpus multimodaux et la formation des tuteurs à la pédagogie synchrone en ligne.

Nous présenterons dans un premier temps le contexte global de notre recherche (corpus d'étude, recueil des données et cadre théorique). Puis, nous précisons nos questions de recherche et la grille d'analyse que nous proposons, afin d'y apporter des éléments de réponses. Cette grille sera illustrée tout du long par deux exemples de pratiques tutorales, afin d'en faciliter la compréhension.

2 CONTEXTE DE RECHERCHE

Le *Français en (l'ère) ligne* est une recherche-action qui permet à des étudiants en Master de Français Langue Etrangère (FLE) de créer et de tutorer des tâches multimédias pour des apprenants étrangers distants par le biais d'échanges synchrones et/ou asynchrones autour de ces tâches.

Dans le cadre de notre article, nous nous proposons d'étudier les échanges recueillis lors de l'édition 2007-2008 du *Français en (l'ère) ligne*¹, entre des étudiants en Master 2 de FLE à l'Université Lumière Lyon 2 et les apprenants américains (niveau intermédiaire) de l'Université de Californie Berkeley. Les tuteurs sont au nombre de onze et les apprenants dix-sept, de sorte que les apprenants, par groupe de deux, interagissent soit avec un tuteur seul soit avec un groupe de deux tuteurs. Du côté américain, l'échange hebdomadaire s'inscrit dans le cadre d'un programme d'enseignement plus large du français. De même, pour les tuteurs, les échanges avec Berkeley représentent quarante-cinq minutes d'une formation hebdomadaire de trois heures. La formation des tuteurs consiste en effet également à préparer des activités, d'abord seul, puis en collaboration avec les autres tuteurs, avant de tutorer en interaction et enfin à acquérir une réflexivité sur leur façon de tutorer.

Les interactions ont lieu par l'intermédiaire de l'outil synchrone de communication Skype². Le choix de ce logiciel a été guidé selon trois critères : la diversité des modalités mises à disposition pour les utilisateurs, l'ergonomie d'utilisation du logiciel et la compatibilité avec l'environnement technique (ordinateurs MAC ou PC, état des ordinateurs mis à disposition par la Faculté, charge sur la bande passante).

Les tuteurs et les apprenants disposent ainsi de trois canaux de communication: le clavier, l'audio et la vidéo, pouvant ainsi écrire, voir, entendre et parler à leur(s) interlocuteur(s). Voici une capture de l'écran d'un des tuteurs qui permet de visualiser l'espace de travail :

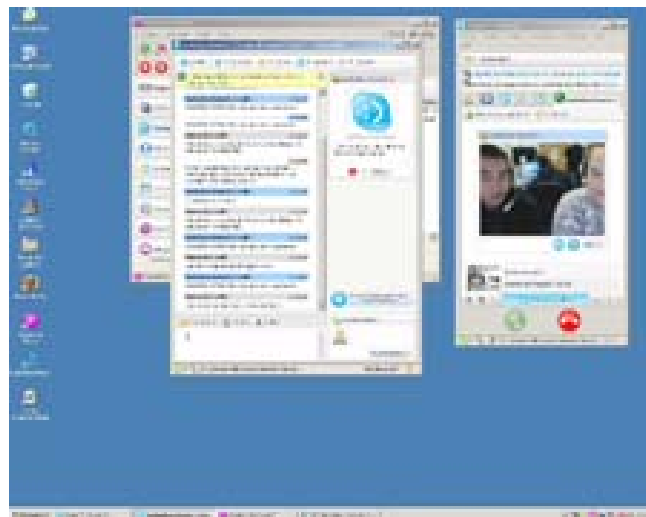


Image 1 : capture d'écran d'un tuteur

3 CADRE THEORIQUE

Plusieurs auteurs issus de différentes disciplines nourrissent nos réflexions, cependant notre perspective est celle des sciences du langage, et plus précisément de l'analyse du discours en interaction.

Nous allons ici présenter succinctement les notions et concepts qui nous ont servi dans cet article pour l'analyse de nos données.

Définissons tout d'abord la terminologie concernant la multimodalité.

Nous appelons « outils » les ressources matérielles manipulées dans un but de production ou de réception sémiotiques : le casque, le micro, la webcam, le papier, le stylo, l'écran, les haut-parleurs. Les « supports » sont les plans de séance, les documents papier ou numériques, qui contiennent des informations nécessaires au déroulement de la séance. Afin de différencier les termes « canal » et « mode », nous nous appuyons sur la définition de la communication multimodale de Kerbrat-Orecchioni, qui décrit celle-ci « comme à la fois multicanale et plurisémiotique » (p.178). Dans le cadre de notre étude, les interlocuteurs utilisent différents canaux à travers ils émettent et réceptionnent différentes ressources sémiotiques. Nous considérons donc les canaux comme les différents formats que peut emprunter la communication, en l'occurrence le clavardage, la vidéo et l'audio, et les modes comme les moyens de communication qui vont être sollicités pour l'émission et la réception des éléments sémiotiques. Prenons trois exemples afin de clarifier ces définitions : lorsqu'un interlocuteur écrit sur le canal clavardage, il utilise le mode textuel. Pour celui qui réceptionne le message, il aura recours au mode visuel. Pour produire un geste à travers le canal vidéo, on utilisera le mode mimo-gestuel. En réception, c'est le mode visuel qui sera de nouveau sollicité. Enfin, pour le mode de production oral (à travers le canal audio), c'est le mode auditif qui sera requis.

Il nous paraît important de distinguer les modes de production des modes de réception puisque ceux-ci ne sollicitent pas les mêmes ressources et n'agissent pas aux mêmes niveaux, comme nous l'avons vu dans les exemples ci-dessus. Le mode visuel en réception inclut par exemple les modes textuel ou posturo-mimo-gestuel en production.

Récapitulons ces choix terminologiques dans le tableau ci-dessous :

Outils	Canaux	Modes de production	Modes de réception
Clavier	Clavardage	Textuel	Visuel
Casque Micro Haut parleurs	Audio	Oral	Auditif
Webcam	Vidéo	Posturo-mimo gestuel Visio	Visuel

Tableau 1 : Terminologie de la multimodalité, dans le cadre de notre étude.

Les tuteurs que nous étudions sont tous équipés d'une webcam, et sont filmés par celle-ci à destination des apprenants avec lesquels ils interagissent. De même, ils ont la possibilité de visionner leurs interlocuteurs à l'écran. Nous nous appuyons sur Develotte, Guichon, Vincent (2010) pour analyser la façon dont ils vont utiliser le potentiel offert par cette fenêtre vidéo. En effet, les tuteurs ont le choix de regarder ou non leurs interlocuteurs à l'écran, et de se rendre eux-mêmes visibles ou non. De plus, il ne suffit pas de savoir si la fenêtre vidéo est présente ou masquée sur l'écran des tuteurs, nous voulons savoir si ceux-ci la regardent et exploitent cet espace (en étant expressifs, par exemple). Develotte, Guichon et Vincent ont définis cinq « degrés d'investissement de la fenêtre vidéo », que nous reprendrons dans nos analyses :

- Degré 0* : Le tuteur est hors du champ de la caméra, ou le canal vidéo est inutilisable.
- Degré 1* : Le tuteur ne porte pas son regard sur l'écran de l'ordinateur.

Degré 2 : Le regard du tuteur est porté sur l'écran d'ordinateur où la fenêtre vidéo est affichée.

Degré 3 : Le tuteur regarde la fenêtre vidéo sur l'écran d'ordinateur et de plus, son visage est expressif et/ou il fait des gestes.

Degré 4 : Le tuteur regarde la webcam directement, donnant l'impression à son interlocuteur qu'il est regardé « dans les yeux ».

Pour analyser la gestualité des interlocuteurs en interaction, nous nous appuyons sur Cosnier (2008), qui distingue les « gestes pour accompagner le discours » des « gestes pour piloter l'interaction ».

Les premiers renvoient à « l'activité mimo-gestuelle qui est liée à la constitution de l'énoncé auquel elle s'intègre : en premier lieu par la gestualité déictique ou désignante [...] (geste de pointage) à laquelle s'ajoute la gestualité illustrative qui mime l'action, ou figure dans l'espace certaines caractéristiques du référent [...] (par exemple « escalier en colimaçon »). D'autres gestes co-verbaux sont directement liés à la pensée abstraite, gestes dit pour cela « métaphoriques » (« beaucoup »). Pour terminer il faut mentionner les « quasi-linguistiques » qui sont des équivalents de paroles et sont parfaitement codifiés selon les cultures.

La deuxième catégorie précise les gestes qui servent à piloter l'interaction. Pour assurer mutuellement l'échange, il existe un dispositif d'interaction auquel s'ajoute un dispositif de partage et de maintenance de la parole. Ces dispositifs sont très largement mimo-gestuels et utilisent en particulier les hochements de tête et la mobilité des regards.

Le dernier point souligné par Cosnier dans l'étude du non verbal renvoie à l'empathie : « En situation d'interaction, les locuteurs vont donc, selon les règles du cadrage affectif, gérer leurs propres sentiments, gérer l'expression de ces sentiments réels ou affichés, et s'efforcer de percevoir les mouvements analogues en cours chez leur partenaire. » (p.123).

Si ce système de coordination est important dans les situations pédagogiques en présentiel, il l'est encore davantage dans un contexte d'interaction multimodale à distance où la communication nécessite d'autant plus d'être soutenue et l'effet de présence maintenu.

Enfin, lorsque nous analyserons le contenu pédagogique des productions orales des tuteurs, nous nous appuyons sur les définitions de Guichon et Drissi (2008) : « Par régulations pédagogiques, nous désignons l'ensemble d'actions et d'énoncés déployé par un enseignant de langue pour créer et maintenir des conditions optimales autour d'une tâche d'apprentissage et ménager des opportunités pour amener les apprenants à développer des compétences langagières. » (p.6). L'étude des régulations pédagogiques nous permettra ainsi de comprendre la gestion de l'interaction orale par chacun des tuteurs étudiés.

4 METHODOLOGIE

4.1 Recueil de données

Pour satisfaire une démarche empirique, nous avons tout d'abord récolté le corpus le plus large possible, sachant que nous pourrions le constituer sur des bases plus sélectives lorsque nous aurions élaboré nos questions de recherche.

Les interactions synchrones sur Skype entre les tuteurs et les apprenants ont été enregistrées grâce à des captures dynamiques d'écran réalisées par le logiciel *Screen Video Recorder*³. Il nous a paru intéressant de pouvoir visualiser non seulement les interactions entre les différents interlocuteurs mais aussi toute activité des tuteurs à l'écran (ouverture d'une page web, aménagement des fenêtres dans l'espace de l'écran par exemple). En effet, cela nous permet de connaître la façon dont les tuteurs organisent leur « bureau virtuel » : s'ils ont une activité parallèle sur un site web ou un logiciel, s'ils mettent au premier plan les fenêtres Skype et si oui lesquelles.

Des débriefings entre les tuteurs et les formateurs ont suivi les interactions et ont été filmés à l'aide d'une caméra numérique.

Les tuteurs ont répondu à des questionnaires pré-interactions afin de connaître leur identité et leurs expériences préalables d'enseignement en présentiel ou à distance.

De même, des entretiens semi-directifs ont été réalisés à la fin de l'expérience⁴.

4.2 Corpus présenté dans cet article.

Pour faciliter la présentation de notre grille d'analyse, nous allons étudier un échantillon de résultat à l'échelle de deux tuteurs (T1 et T2) sur une activité. Cette activité est la première de la séance six⁵. La séance six est l'avant-dernière séance de la formation.

Voici la façon dont est prévu son déroulement :

Séance 6 : 11 mars 2008, thème: « les voyages », 5 activités:

- Négociation (10min)
- Échange d'informations (10min)
- Échange d'opinion (10min)
- Échange d'informations (10min)
- Bilan (5min).

Le déroulement des activités nécessite de se rendre sur un site internet et à observer puis décrire des photographies. Les apprenants doivent imaginer le lieu photographié pour ensuite répondre à différentes questions des tuteurs sur ce lieu imaginaire.

Concrètement, les tuteurs auront alors à l'écran : les fenêtres visio et clavardage et le site web où sont hébergées les photographies.

5 QUESTIONS DE RECHERCHE

Plusieurs modes sont mis à la disposition des tuteurs dans leurs espaces de travail : lesquels vont être utilisés par chaque tuteur, dans quelle proportion et avec quel degré d'investissement ?

Quelles conséquences ces différents choix d'utilisation ont sur la nature de l'interaction ?

Nous émettons l'hypothèse que des corrélations pourraient être établies à deux niveaux : Premièrement, le profil initial des tuteurs (compétences individuelles, expériences professionnelles en présentiel ou à distance, habitude de l'environnement informatique) et le contexte des interactions (perturbations extérieures, problèmes techniques, type de tâche, besoin exprimé des apprenants) ont une influence sur la façon dont le tuteur utilise les outils à disposition.

Deuxièmement, les choix d'utilisations de ces différents modes influencent la nature de l'interaction et la relation entre tuteurs et apprenants.

Nous allons dans la section suivante, présenter la grille d'analyse que nous avons établie afin d'apporter des réponses à nos questions de recherche. Nous reprendrons nos hypothèses de recherche afin de structurer notre présentation. Pour chaque hypothèse les pratiques de deux tuteurs seront analysées, afin d'illustrer notre présentation.

Enfin, les représentations des tuteurs sur leur propre façon de tutorer exprimées lors des entretiens ou des débriefings nous permettront de les comparer à nos analyses.

6 GRILLE D'ANALYSE ILLUSTRÉE PAR DEUX EXEMPLES DE PRATIQUES TUTORALES.

6.1 Hypothèse 1 : Les tuteurs vont utiliser différemment les modes mis à leur disposition.

Tout d'abord définissons les différents modes que nous prenons en compte dans notre étude :

Mode textuel	Utilisation du clavardage synchrone pour écrire un message.
Mode Visio ⁶	Image des interlocuteurs visible à l'écran, et les tuteurs participent à des échanges oraux.
Mode Audio seul	L'image des interlocuteurs invisible à l'écran, les échanges sont uniquement oraux.

Tableau 2 : Définitions des trois modes disponibles pour les tuteurs.

Les modes que nous prenons en compte concernent uniquement l'usage en production, et non en réception, puisque nous focalisons notre étude sur les tuteurs.

Pour le mode textuel, nous comptabilisons le nombre d'utilisation du chat du tuteur. Il est toutefois intéressant de savoir que dans tout notre corpus, les apprenants n'ont jamais utilisé le chat en production, mais uniquement en réception.

Pour le mode visio, nous comptabilisons la durée pendant laquelle le tuteur affiche sur son écran la fenêtre vidéo des apprenants, tout en ayant une conversation orale avec eux. La visio est à différencier de la vidéo, qui consiste uniquement à se voir et à voir son interlocuteur, alors que la visio consiste à voir et à parler ou écouter simultanément.

A l'inverse, le mode audio consiste à masquer la fenêtre vidéo des apprenants et communiquer uniquement par l'oral.

Voici les résultats pour les deux tuteurs qui constituent notre échantillon illustratif :

	T1	T2
Temps total de l'activité	9 '40	6 '00
Visio	5 '14 (54%)	6 '00 (100%)
Audio seul	4 '26 (46%)	0
Fréquence d'utilisation du clavardage ⁷	1	0

Tableau 3 : Temps et pourcentage d'utilisation de chaque mode.

Nous observons que le tuteur 2 conserve en permanence la fenêtre vidéo de ses interlocuteurs à l'écran et mène simultanément une conversation orale avec eux. Le tuteur T1, à l'inverse, visualise la fenêtre vidéo de ses interlocuteurs la majorité du temps, mais il a aussi recours à l'audio seul (46% du temps) et au clavardage (une fois).

Nous constatons dans un premier temps que ces deux tuteurs gèrent différemment les ressources multimodales mises à leur disposition.

Le tableau 2 nous informe du temps pendant lequel la fenêtre vidéo des apprenants apparaît sur l'écran des tuteurs. Cependant, cela ne nous permet pas de savoir si les tuteurs regardent la fenêtre, ni si eux-mêmes se rendent visibles pour leurs apprenants. La fenêtre vidéo peut en effet être exploitée, en production et en réception, avec différents degrés d'investissement, tels que nous les avons définis dans le cadre théorique.

6.1.1 Investissement du mode visio.

Voyons comment nos deux tuteurs investissent la fenêtre vidéo pour cette activité :



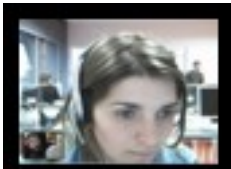

	T1	T2
Temps total de l'activité.	9 '40	6 '00
Degré 0 	3 '36 (37%)	0
Degré 1 	2 '37 (25%)	0 '25 (7%)
Degré 2 	3 '22 (34,5%)	0
Degré 3 	0 '15 (2,5%)	5 '35 (93%)
Degré 4	0	0



Tableau 4 : degrés d'investissement de la fenêtre vidéo.

Les degrés 0 et 1 sont considérés comme une « non-utilisation » de la webcam. En effet, le tuteur ne se rend pas visible pour ses interlocuteurs ou ne les regarde pas. Au regard du tableau 2, le tuteur T1 semblait utiliser de manière mixte la visio et l'audio seul en proportions quasiment égales, mais nous constatons dans le tableau 4 qu'il n'investit la visio seulement 37% du temps. C'est-à-dire que la plupart du temps de l'interaction, soit il n'a pas affiché la fenêtre de ses interlocuteurs à l'écran, soit il ne la regarde pas, s'orientant d'avantage vers l'audio. À l'inverse, le tuteur T2 regarde la fenêtre des apprenants et se rend visible et expressif, 93% du temps. Rappelons que les tuteurs ont un bureau virtuel à gérer mais aussi un bureau « matériel » sur lequel ils placent une feuille sur laquelle est indiqué le plan de la séance et une feuille sur laquelle ils peuvent prendre des notes. Ainsi, le temps durant lequel les tuteurs n'utilisent pas la visio est un temps pendant lequel ils sont concentrés sur d'autres tâches.

Les degrés 2 à 4 graduent l'utilisation de la webcam. Le degré 2 concerne le tuteur T1 pour la grande majorité du temps. T1 regarde donc ses apprenants sur son écran d'ordinateur, mais ne se montre expressif que pendant 15 secondes (degré 3). Au contraire, T2 utilise uniquement la webcam au degré 3 c'est-à-dire qu'il est expressif tout le long de l'interaction, faisant ainsi sentir à ses interlocuteurs qu'ils sont regardés.

Pour compléter notre étude détaillée de l'investissement du mode visio, regardons plus précisément les gestes et expressions des tuteurs. Nous comptabilisons le nombre de hochements de tête, expressions empathiques et gestes illustratifs de chaque tuteur, durant l'activité :

	T1	T2
Temps total de la tâche	9 '40	6 '00
Hochements de tête	9	8
Expressions empathiques	1	3
Gestes illustratifs	2	1

Tableau 5 : Gestes et expressions des tuteurs.

Les deux tuteurs ont globalement une gestualité et une expressivité similaires. En effet, tous deux hochent la tête et illustrent leurs propos par des gestes. Les gestes et expressions des tuteurs revêtent différentes fonctions dans l'interaction. Ils coordonnent les échanges (volonté de prendre la parole, encouragement à la prise de parole, distribution de la parole), ou encore proposent un retour sur leur contenu (validation, incompréhension, demande de répétition, et cetera). Les rires, sourires et hochements de tête participent aussi à l'établissement de la relation interpersonnelle. Il est intéressant de constater que, alors que les tuteurs ne semblaient pas accorder la même importance à l'investissement de la visio (tableaux 3 et 4), tous les deux produisent pour autant des gestes et expressions à destination de leurs interlocuteurs. Nous pouvons donc émettre l'hypothèse que nos tuteurs semblent tous deux conscients d'être regardés par leurs apprenants et de l'impact positif que leur gestualité peut apporter à l'interaction.

Ces comportements sont éclairés par les propos des tuteurs concernant l'utilisation de la webcam, lors des débriefings et entretiens post-interactions. Nous apprenons ainsi que T1 est conscient de ne regarder que très peu ses apprenants, mais que eux le regardent en permanence. Il nous dit ainsi jouer sur son expressivité pour leur donner l'impression que lui aussi les regarde et les rassurer. Il explique de même que pour lui la vidéo est très importante mais qu'il a parfois d'autres choses à gérer, ce qui ne lui permet pas de regarder ses apprenants en permanence.

Nous avons constaté au cours des précédents résultats d'analyses que le tuteur T2 privilégie la visio sur les autres modes. Ce tuteur nous explique qu'il n'est pas à l'aise avec l'environnement informatique et se sent incapable de gérer plusieurs choses à la fois: la conversation orale, écrire sur le clavardage, écrire sur une feuille, lire sur l'écran, lire sur une feuille, ouvrir une fenêtre extérieure. Sa stratégie a donc été d'utiliser le seul mode qui lui a paru essentiel : la visio, pour se voir et s'entendre.

Les entretiens post-interactions nous apprennent que d'après ces deux tuteurs, un discours multimodal, c'est-à-dire combinant le plus de modalités, est le plus efficace en interaction. Ainsi, les stratégies qu'ils adoptent dans le choix de l'investissement des modes sont en partie liées au coût cognitif qu'elles demandent.

Ces premières analyses nous ont permis de comprendre comment les tuteurs investissent le mode visio, et d'entrevoir pour quelles raisons ils choisissent un mode plutôt qu'un autre. Qu'en est-il du mode audio ?

6.1.2 Investissement du mode audio.

Tout d'abord, observons le partage de la parole⁸ des deux tuteurs avec leurs apprenants, pour l'activité étudiée :

	T1	T2
Tuteurs	3'41 (38%)	2'00 (33,5%)
Apprenants	4'02 (42%)	1'53 (31,5%)
Silences	1'57 (20%)	2'07 (35%)

Tableau 6 : Temps de partage de la parole

Les tuteurs T1 comme T2 parlent sensiblement autant que les deux apprenants qui leur font chacun face, pour cette activité. Ce qui les différencie significativement est la part des silences. En effet, pour le tuteur T2, les silences occupent un temps (35%) plus important que le temps de parole des interlocuteurs (33,5% et 31,5%). A l'inverse, le temps des silences (20%) dans l'interaction de T1 représente la moitié des temps de parole du tuteur (38%) et des apprenants (42%). Cette première constatation nous amène à nous interroger sur le contenu des productions orales des tuteurs :

Productions orales des tuteurs	T1	T2
Corrections	9	2
Consignes	18	10
Soutien Socio-affectif	52	4

Total	79	16
-------	----	----

Tableau 7 : contenu pédagogique de la production orale des tuteurs.

La différence la plus frappante qui ressort du tableau 7 est le plus grand nombre de productions orales de T1 (79) par rapport à T2 (16), notamment la proportion du contenu socio-affectif (52 pour T1 et 4 pour T2)⁹. Nous constatons que l'écrasante majorité des productions de T2 sont des consignes (10 sur 16), le tuteur émet en effet peu de corrections ou de soutien socio-affectif. T1 produit en grand nombre chacune des trois catégories que nous avons référencées.

Mettons ces résultats d'analyses du mode audio en perspective de nos précédents résultats: A l'échelle de notre échantillon, T1 est **orienté vers l'audio et la visio** (tableau 3). Il fait peu usage de la webcam en réception puisqu'il ne regarde pas ses apprenants (tableau 4) mais produit pour autant des gestes (tableau 5). Il supporte la communication par la parole, laissant peu de place aux silences (tableau 6), quitte à parler autant que ses deux apprenants réunis. Il se montre peu expressif à la caméra mais soutient pour autant ses apprenants à l'oral, comme nous le montre la part du contenu socio-affectif dans le tableau 7.

T2 est **orienté vers la visio** (tableau 3 et 4) en production et en réception : il regarde la fenêtre vidéo et se rend visible. Les consignes représentent la plus grande partie de ses tours de paroles (tableau 7), puisqu'il donne les feedbacks positifs ou négatifs à ses apprenants plus volontiers par des expressions ou des gestes que par des mots (tableaux 5 et 7).

Les analyses de l'investissement de chaque mode, étendues aux cinq activités de la séance 6 pour un plus grand nombre de tuteur, nous permettront d'établir les « profils » des tuteurs, comme nous venons de le démontrer avec les tuteurs T1 et T2 en déterminant leur « orientation modale ». Cette orientation peut être vers la visio (T2), vers l'audio seul ou mixte (T1). Cela nous permet ainsi d'éprouver notre hypothèse selon laquelle les tuteurs utilisent différemment les modes à leur disposition. Essayons maintenant de comprendre quelles raisons les amènent à ces utilisations différentes.

6.2 Hypothèse 2 : Ces utilisations différentes vont dépendre de certains facteurs.

6.2.1 Le contexte de l'interaction.

Plusieurs informations contextuelles sont importantes à prendre en considération si l'on veut pleinement comprendre les stratégies des tuteurs.

L'activité qui constitue notre échantillon est la première de la sixième séance. Ce jour-là, les apprenants américains sont passés à l'heure d'été et cela n'a pas été anticipé à Lyon. Ainsi, l'interaction a été au dernier moment avancée d'une heure. La première conséquence engendrée fut un stress pour certains tuteurs qui auraient eu besoin de relire le plan de la séance ou de se préparer mentalement à l'interaction. Nous en retrouvons des traces dans les entretiens. La deuxième conséquence fut que les ordinateurs utilisés ne furent pas allumés en avance et connurent un ralentissement le temps de cette première activité.

Il faut ensuite savoir que l'activité consistait à visionner des photographies sur un site web extérieur : les apprenants avaient pour consigne de les décrire et de deviner où elles avaient été prises.

Ainsi, certains tuteurs n'ayant pas pu relire leur plan de séance, n'ont pas précisément en tête les photographies. Les ordinateurs sont ralentis : ouvrir une page web et charger les photographies va donc prendre beaucoup de temps. Ainsi, le choix qui s'offre à eux va être le suivant : masquer les fenêtres vidéo et clavardage en se rendant sur une page extérieure, aménager l'espace pour avoir à l'écran les fenêtres Skype et la page web, ou ne pas ouvrir de page extérieure pour se focaliser uniquement sur l'outil de communication les reliant à leurs interlocuteurs.

Le contexte et le type de l'activité nous permettent ainsi d'affiner le profil de T1 et T2. Comme l'attestent les tableaux 3 et 4, T1 a soutenu l'interaction à l'oral, pour visualiser à l'écran les photographies plutôt

que la fenêtre vidéo de ses apprenants. T2 a maintenu le contact visuel et n'a ouvert aucune page extérieure. Il est donc indispensable de comparer les profils d'utilisation de ces deux tuteurs à ceux d'autres activités où la charge cognitive est moins importante.

6.2.2 Les informations apportées par les questionnaires pré-interaction.

Voici les informations que nous apportent les questionnaires pré-interactions sur les deux tuteurs qui illustrent notre grille d'analyse :

	T1	T2
Age	31	45
Langue maternelle	Française	Française
Familiarité avec l'informatique	Bonne	Très peu
Familiarité avec une plateforme multimodale.	Oui (msn)	Non
Expérience professionnelle de tutorat	Oui, avec différents publics.	Oui, en lycée professionnel depuis de nombreuses années.
Expérience de tutorat à distance	Non	Non
Attentes exprimées	Avoir un bon contact avec les apprenants.	Apprendre à gérer une interaction pédagogique à distance.
Craintes exprimées	Gestion de la relation, étudiants non réceptifs, devoir gérer un problème technique.	Que la communication « ne passe pas », gérer des outils techniques.

Tableau 8 : informations initiales sur les tuteurs.

T1 semble à l'aise avec les outils informatiques et d'avantage préoccupé par la relation avec ses apprenants. T2 est un tuteur chevronné, mais connaît peu d'expériences avec l'informatique et c'est vers cela que se focalisent ses craintes et ses attentes.

Ces informations éclairent une fois de plus nos analyses : il est à prévoir que T2 aura plus de difficultés à gérer les outils multimodaux et nous comprenons donc pourquoi il s'oriente uniquement vers le mode visio. Quant à T1, il est capable de réaliser plusieurs tâches simultanément : ouvrir une page web tout en soutenant l'interaction uniquement à l'oral. Il sera intéressant de suivre l'évolution des utilisations des

tuteurs au cours du temps, en comparant une séance située au début de la formation avec cette sixième séance.

La grille d'analyse nous permet donc de comprendre comment les tuteurs utilisent les modes et pour quelles raisons ils le font différemment. Cependant, nous sommes en droit de nous demander quelles vont être conséquences et à quel(s) niveau(x) ?

6.3 Hypothèse 3 : Ces différentes utilisations vont avoir des conséquences sur la nature de l'interaction et la relation entre tuteurs et apprenants.

Dans les tableaux 5 et 7, nous avons montré que les traces socio affectives occupent une part dans les interactions qui varient selon les tuteurs et les activités. Nous retrouvons les traces socio-affectives dans les productions orales et visuelles (gestes et expressions) et dans le discours des tuteurs lors des entretiens. Nous ne rentrerons pas dans le détail des résultats d'analyses de ces premières entrées, pour nous focaliser sur les « moments non-pédagogiques ». En effet, les tuteurs ont un plan de séance détaillé à respecter et au cours, ou entre les activités, se glissent des tours de parole qui n'ont pas trait à la pédagogie (c'est-à-dire ne font pas partie des activités prévues). Nous faisons l'hypothèse que selon le profil que nous dresserons des tuteurs, ces moments non-pédagogiques auront une part plus ou moins importante dans l'interaction, et à des moments fixes (par exemple lors de l'ouverture, la clôture ou lors des transitions entre les activités).

Voici comment est théoriquement prévue la séance 6 :

Activité 1	Activité 2	Activité 3	Activité 4	Bilan
10 :00	10 :00	10 :00	9 :00	4 :00

Tableau 9 : plan de séance de la séance 6.

Voici maintenant comment les tuteurs T1 et T2 l'ont réalisée :

T1, séance 6 : 35min 24.

Salutations, Ouverture.	Activité 1	Activité 2	Activité 3	Activité 4	Fermeture : conversation Non- Pédagogique.
0 :44	9 :40	6 :01	11 :28	6 :03	1 :40

Tableau 10 : plan de la séance 6 de T1.

T2, séance 6 : 29 min 35.

Saluta tions	Act.1	Act.2	Act.3	Aparté ¹⁰	Act.4	Act. Impro visée	Aparté ¹¹	Act.5	Saluta tions
0 :02	6 :00	4 :20	4 :41	0 :24	7 :20	3 :28	0 :10	2 :22	0 :15

Tableau 11 : plan de la séance 6 de T2.

La comparaison des tableaux 9 et 10 nous apprend que T1 ne réalise pas les activités 4 et 5 mais consacre du temps à l'ouverture et la clôture de l'interaction. La séance de T2 est moins longue et pourtant il réalise toutes les activités prévues (en improvisant même une cinquième). Les séquences d'ouverture et de clôture sont très courtes et les moments d'aparté concernent la gestion de l'interaction. Ces premières analyses nous permettent de penser que T1 est plus orienté vers la relation socio-affective que T2, qui lui privilégie une interaction plus formelle. Cela étaye les résultats des analyses des informations fournies par les questionnaires pré-interactions (tableau 8).

7 CONCLUSION

La recherche que nous avons présentée est en cours de réalisation sur un plus grand nombre de tuteurs et de séances, dans le cadre d'un doctorat. En effet, nous avons comparé les comportements des tuteurs en faisant varier différents facteurs. Nous avons comparé tout d'abord les pratiques de quatre tuteurs sur l'ensemble des activités de la deuxième séance. Ceci nous a permis d'établir des similarités entre les comportements des tuteurs et avec quelles variables cela est en relation (type d'activité, profil du tuteur, contexte de la séance, et cetera). Nous avons appliqué ensuite cette même grille d'analyse sur la sixième séance de la formation, afin d'observer l'évolution des comportements des tuteurs au cours du temps.

Nous avons voulu tout d'abord participer à la réflexion concernant la méthodologie et le protocole de recherche pour les corpus multimodaux, mais aussi en fonction de nos résultats d'analyses, nourrir la recherche sur la formation des tuteurs.

Notre but n'est pas d'établir une hiérarchie entre les pratiques des tuteurs, ni d'inférer sur les effets concernant l'apprentissage des apprenants. En effet, cette ambition ne nous paraît ni pertinente ni réalisable, puisque des variables extérieures à notre étude entrent en compte (par exemple le niveau initial des apprenants ou l'apprentissage du français qu'ils reçoivent en dehors des interactions), et ne permettent pas de comparer l'efficacité des différents styles d'enseignement des tuteurs.

Notre ambition est de pointer la relation qui existe entre les attentes, les capacités, les compétences d'un tuteur et l'interaction qui est réalisée.

REFERENCES BIBLIOGRAPHIQUES.

- Anis, J. (2001). Approche sémio-linguistique des représentations de l'ego dans la Communication Médinée par Ordinateur, *Langages*, 144, pp.20-38.
- Cosnier, J. (2008). *Les gestes du dialogue, La communication état des savoirs*. Auxerre : Sciences humaines, pp.119-128.
- Cosnier, J. & Develotte, C. (2011). Le face à face en ligne, approche éthologique. In Develotte, C., Kern, R. & Lamy, M.-N. (Eds.) *Décrire la conversation en ligne, Le face à face distanciel*. Lyon, ENS Editions, pp.28-50.
- De Fornel, M. (1994). Le cadre interactionnel de l'échange visiophonique. *Réseaux*, 64, pp.107-132.
- Dejean-Thircuir, C., Mangenot, F. (2006). Pairs ou tutrices ? Pluralité des positionnements d'étudiantes de maîtrise FLE lors d'interactions en ligne avec des apprenants australiens. *Le français dans le monde, Recherches et applications, Les échanges en ligne dans l'apprentissage et la formation*, Paris CLE International, pp.75-87.
- Develotte, C., Guichon, N. and Kern, R. (2008). "Allo Berkeley? Ici Lyon... Vous nous voyez bien?" Etude d'un dispositif de formation en ligne synchrone franco-américain à travers les discours de ses usagers. *Alsic*, 11 (2), pp.129-156.
- Develotte, C., Guichon, N., Vincent C. (2010). The use of webcams for teaching a foreign language in a desktop videoconferencing environment, *ReCall*, 22 (3), pp.293-312.

- Develotte, C., Kern, R. & Lamy, M.-N.** (Eds.) (2010) *Décrire la conversation en ligne, Le face à face distanciel*. Lyon, ENS Editions, pp.28-50.
- Guichon, N., Drissi, S.** (2008). Tutorat de langue par visioconférence : comment former aux régulations pédagogiques. *Les Cahiers de l'ACEDLE*, 5 (1), pp.185-217.
- Hampel, R. and Lamy, M.-N.** (2007). *Online Communication in Language Learning and teaching*, Basingstoke Palgrave Macmillan.
- Kerbrat-Orecchioni, C.** (2011) Conversations en présentiel et conversations en ligne : bilan comparatif. In Develotte, C., Kern, R. & Lamy, M.-N. (Eds.) *Décrire la conversation en ligne, le face à face distanciel*. Lyon, ENS Editions, pp.173-195.
- Marcoccia, M.** (2000). « Les Smileys : une représentation iconique des émotions dans la communication médiatisée par ordinateur », in Plantin, C., Doury, M., Traverso, V. (eds), *Les émotions dans les interactions communicatives*, Lyon : ARCI - Presses Universitaires de Lyon, pp.249-263.
- Maître de Pembroke, E. & Legros, D.** (2001). Multimédia, multimodalité et construction des connaissances. In J. Crinon et F. Gautellier (Éds), *Apprendre avec le multimédia et Internet*, Paris: Retz, pp. 193-202.
- Traverso, V.** (2011). Chevauchements de parole, résolution et réparations dans la conversation en ligne in Develotte, C., Kern, R. & Lamy, M.-N. (Eds.) *Décrire la conversation en ligne, Le face à face distanciel*. Lyon, ENS Editions, p117-143.
- Wang, Y.** (2004). Supporting synchronous distance language learning with desktop videoconferencing. *Language Learning & Technology*, 8, pp.90-121.

Tableaux et graphiques.

- Figure 1: Capture d'écran d'un tuteur.
- Tableau 1: Terminologie de la multimodalité, dans le cadre de notre étude.
- Tableau 2: Définitions des trois modes disponibles pour les tuteurs.
- Tableau 3: Temps et pourcentage d'utilisation de chaque mode.
- Tableau 4: Degrés d'investissement de la fenêtre vidéo.
- Tableau 5: Gestes et expressions des tuteurs.
- Tableau 6: Temps de partage de la parole.
- Tableau 7: Contenu pédagogique de la production orale des tuteurs.
- Tableau 8: Informations initiales sur les tuteurs.
- Tableau 9: Plan de séance de la séance 6
- Tableau 10: Plan de la séance du tuteur T1.
- Tableau 11: Plan de séance du tuteur T2.

¹ Le Français en Première ligne : <http://w3.u-grenoble3.fr/le-1-ligne/index.html>

² Skype : <http://www.skype.com/intl/fr/>

³ Screen Video Recorder : <http://www.wordaddin.com/screenvcr/>

⁴ Le corpus recueilli représente environ 25h de captures dynamiques d'écran, 4h45 de vidéos des interactions filmées de l'extérieur, 4h de débriefings, 5h d'entretiens post-interactions.

⁵ Nous avons choisi cette séance car les pratiques sont plus stabilisées que les lors des premières séances, et l'effet de nouveauté atténué.

⁶ Le mode « vidéo seule » n'est pas pris en compte puisqu'on n'en trouve aucune occurrence dans le corpus d'étude.

⁷ Le mode textuel n'étant utilisé qu'une dans notre échantillon, nous ne l'analyserons pas dans cette étude.

⁸ À l'échelle de l'échantillon illustratif que nous présentons, nous nous interdisons bien entendu de tirer des conclusions ou de généraliser sur des données aussi peu exhaustives. À plus grande échelle, il nous sera possible

d'établir si les stratégies d'utilisation adoptées par les tuteurs ont une influence directe sur le partage de la parole et les temps de silences.

⁹ Ces résultats doivent être nuancés du fait que la tâche de T1 dure 3min 40 de plus que celle de T2.

¹⁰ Aparté : T2 demande aux apprenants le temps qu'il reste.

¹¹ Idem