



HAL
open science

Ce que l'oculométrie peut apporter dans une approche écologique des échanges en ligne. Une discussion épistémologique et une étude de cas

Marco Cappellini, Yu-Yin Hsu

► To cite this version:

Marco Cappellini, Yu-Yin Hsu. Ce que l'oculométrie peut apporter dans une approche écologique des échanges en ligne. Une discussion épistémologique et une étude de cas. EPAL - Echanger Pour Apprendre en Ligne, Jun 2018, Grenoble, France. hal-02023002

HAL Id: hal-02023002

<https://hal.science/hal-02023002>

Submitted on 18 Feb 2019

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

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

CE QUE L'OCULOMETRIE PEUT APPORTER DANS UNE APPROCHE ECOLOGIQUE DES ECHANGES EN LIGNE. UNE DISCUSSION EPISTEMOLOGIQUE ET UNE ETUDE DE CAS.

Marco Cappellini

Aix Marseille Université

Laboratoire Parole et Langage (UMR 7309 CNRS)

Yu-Yin Hsu

Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

Résumé : Dans cet article nous souhaitons développer une réflexion théorique et épistémologique sur l'approche écologique à la lumière de récentes possibilités ouvertes par le développement d'outils d'oculométrie. Plus spécifiquement, nous nous intéressons à ce que les données d'oculométrie peuvent apporter dans l'étude du développement des compétences techno-sémio-pédagogiques (Guichon, 2012) dans une telle approche. L'article présente une première partie développant cette réflexion épistémologique d'un point de vue théorique et une deuxième partie montrant une application de ces principes à une étude exploratoire dans un dispositif basé sur le modèle du *Français en (première) ligne* (Mangenot, 2013) pour le chinois langue étrangère.

Mots clés : compétences techno-sémio-pédagogiques, approche écologique, multimodalité, vidéoconférence, oculométrie.

1. Introduction

Par oculométrie (en anglais *eye-tracking*) on entend toute technique qui permet de recueillir des données et d'observer les mouvements des yeux pendant différentes activités humaines. Si l'oculométrie remonte à un passé assez lointain (fin du XIXe siècle) avec des méthodes d'observation directe, c'est seulement avec le développement d'outils exploitant la lumière infrarouge pour détecter la focale des pupilles qu'une observation systématique a été possible. Ces outils ont été dans un premier temps invasifs et contraignants, mais ils sont devenus récemment de plus en plus discrets et donc intégrables dans des dispositifs pédagogiques. Cette évolution a donné lieu à un intérêt dans le domaine de la télécollaboration (O'Rourke, 2012) et différent.e.s chercheur.euse.s ont mobilisé un ensemble varié de cadres conceptuels, allant d'approches cognitives basées sur la théorie de l'alignement (par exemple

Michel & Smith, 2018) à des approches plus socioculturelles intégrant les dimensions sociales et de médiations des artefacts (par exemple Shi et al., 2017).

Dans cet article nous développons une réflexion épistémologique sur le rôle que l'oculométrie peut jouer dans l'approche écologique. Plus spécifiquement, nous nous intéressons à ce que les données d'oculométrie peuvent apporter dans l'étude du développement des compétences techno-sémio-pédagogiques (Guichon, 2012) dans une telle approche. L'article présente une première partie développant cette réflexion épistémologique d'un point de vue théorique et une deuxième partie montrant une application de ces principes à une étude exploratoire dans un dispositif basé sur le modèle du *Français en (première) ligne* par visioconférence (Develotte et al., 2008) pour le chinois langue étrangère.

2. L'oculométrie dans une approche écologique

2.1. La multimodalité des dispositifs de vidéoconférence pour l'enseignement/apprentissage des langues étrangères

Les recherches s'intéressant à l'utilisation de la vidéoconférence dans l'enseignement/apprentissage des langues étrangères ont souvent fait de la multimodalité un objet d'étude privilégié. La multimodalité peut être définie comme la présence simultanée de plusieurs canaux de perception et de plusieurs modes sémiotiques qui forment un ensemble (par exemple Jewitt et al., 2016). Il est important de souligner que cette multimodalité se retrouve dans toute forme de communication, écrite ou orale (Thibault & Baldry, 2006). Dans les environnements en ligne et dans la télécollaboration, la complexité de la multimodalité dans la communication se trouve accentuée par la présence de plusieurs « éléments » (que nous conceptualisons comme « affordances » ci-dessous) porteurs de sens (Liddicoat, 2011).

Les cadres épistémologiques mobilisées dans la littérature pour étudier la multimodalité sont on ne peut plus larges, allant d'approches semi-expérimentales manipulant les éléments de cette multimodalité (par exemple Yamada & Akaori, 2009) à des approches presque¹ écologiques (Guichon & Tellier, 2017). Une grande variété se retrouve aussi dans les objets d'études possibles, qui vont de la présence en ligne (Satar, 2010), aux compétences pédagogiques (Hampel & Stickler, 2012 ; Guichon & Tellier 2017), à la gestion de l'interaction (Cappellini, 2013). Dans cet article, nous nous concentrerons sur l'étude des compétences techno-sémio-pédagogiques dans une approche écologique.

¹ Nous insérons le « presque » car la seule présence d'un enregistrement, duquel pour des raisons éthiques les sujets sont forcément au courant, modifie l'environnement en le rendant différent de l'environnement ou des phénomènes que l'on voudrait étudier.

2.2. Éléments d'un paradigme écologique et apports de l'oculométrie

2.2.1. Éléments d'un paradigme écologique

Différents chercheurs rattachent le paradigme écologique à un ensemble d'approches et d'auteurs de la deuxième partie du XXe siècle. Ainsi, Cappellini (2014, §2.5) identifie trois auteurs ayant donné leur empreinte à ce paradigme, d'emblée interdisciplinaire : Gregory Bateson (1977, 1980 [1972]), Urie Bronfenbrenner (1979) et James Gibson (1979). Françoise Blin (2016a) lie ce paradigme également au travail de Bronfenbrenner, mais aussi à la systémique (e.g. Larsen-Freeman & Cameron, 2008) et à la théorie de l'activité de matrice socioculturelle (Engstrom, 1988)². Dans cet article, nous retenons quatre éléments du paradigme écologique : les validités écologique et phénoménologique, les rôles induits par toute situation, la relation à l'environnement, le concept d'affordance.

Par validité écologique, Bronfenbrenner (1979) désigne le fait que toute étude écologique devrait se concentrer sur des données provenant d'environnements les plus proches possibles, voire identiques, à ceux où le phénomène étudié se produit. Cela place le paradigme écologique aux antipodes du paradigme expérimental, où au contraire le chercheur reproduit un phénomène dans un contexte contrôlé (par exemple le laboratoire) afin de pouvoir manipuler les différentes variables. Par validité phénoménologique, Bronfenbrenner indique que les analyses doivent prendre en compte le point de vue des sujets sur le phénomène étudié, ce qui le rapproche de la perspective émique privilégiée en sociologie compréhensive et en ethnométhodologie.

Le même auteur note que toute situation induit des rôles, qui, nous ajoutons, sont coconstruits et qui évoluent dans l'interaction. En ce sens, l'analyse conversationnelle peut offrir un cadre conceptuel et des outils pour étudier ces rôles, notamment avec un focus sur les catégorisations (Sacks, 1972 ; Mondada, 1999) que l'on retrouve dans les positionnements discursifs (Cappellini & Rivens Mompean, 2015).

Dans la relation à l'environnement, plusieurs points sont à prendre en compte. Avant tout, la relation à l'environnement pose la question de la perception. Dans ce sens, Gibson (1979) souligne que la perception des éléments d'un environnement, loin de reproduire l'objectivité de cet environnement, est un phénomène dialectique entre un sujet, les éléments de l'environnement et ce qu'ils permettent ou contraignent par rapport aux actions du sujet (*infra*). Dans le cadre d'une interaction verbale multimodale, cela se rattache à la co-construction du contexte tel qu'il est défini dans la théorie de la pertinence (Sperber & Wilson, 1986), autrement dit l'ensemble d'éléments qui sont rendus mutuellement manifestes dans l'interaction et leur relation de pertinence à l'(inter)action³. L'on voit donc que dans un paradigme écologique, la relation à l'environnement ne peut pas faire abstraction des actions des sujets, liés à des processus sémiotiques d'interprétation et de (co-)construction du sens (Van Lier, 2004). Cela signifie que du point de vue de l'analyste, l'environnement pourra être considéré comme un réservoir de ressources sémiotiques relevant de différents régimes de sens (Bezemer & Kress, 2016) mobilisées, et à analyser, comme un tout

² Voir néanmoins le débat dans les volumes 2 et 3 de *Language and Sociocultural Theory* pour les difficiles rapports entre systémique et théorie socioculturelle.

³ Nous renvoyons à Cappellini & Combe (2017) pour une discussion plus détaillée de ce point.

intégré selon le principe d'intégration des ressources sémiotiques (Thibault & Baldry, 2006).

Dans ce cadre, la notion d'affordance va désigner le rapport dialogique entre un acteur et les éléments de l'environnement qui permettent et/ou contraignent son action. Particulièrement pertinente pour nous est la conceptualisation d'affordance opérée par Blin (2016b). Nous nous inscrivons dans ce que Blin appelle une approche post-cognitive et phénoménologique de l'étude des affordances, ce qui signifie que nous concevons les affordances comme émergentes dans l'(inter)action d'un sujet avec un artefact et à étudier dans la perspective de la validité phénoménologique. Au-delà du catalogage et des différences entre affordances séquentielles et emboîtées, *handling affordances* et *effecter affordances* (ibid.), nous retenons surtout qu'il existe quatre types d'affordances interreliées : technologiques, sociales, linguistiques (et plus largement sémiotiques, cf. Bezemer & Kress, 2016) et éducatives. Pour illustrer un exemple d'interaction entre ces types, citons Blin (2016a, p. 49).

Failure to perceive and enact these technological affordances may constrain the realisation of longer-term educational affordances, which in turn may impact on the emergence of language affordances in unpredictable ways.

2.2.2. L'oculométrie dans une perspective écologique

Comme spécifié plus haut, l'oculométrie permet de recueillir des données concernant les mouvements oculaires⁴. L'hypothèse subjacente à tout travail se basant sur ce type de données est que les changements d'attention focale, autrement dit le « point » où une personne pose son regard, permettent de formuler des hypothèses concernant la cognition. Il est néanmoins important de souligner que les données d'oculométrie, comme toute autre donnée, ne fournissent que des traces de la cognition et non une observation directe de celle-ci.

Dans cet article, nous soutenons que l'oculométrie peut jouer un rôle dans l'étude des compétences techno-sémio-pédagogiques (Guichon, 2012) et plus largement de la compétence multimodale (Hauck, 2010). Les compétences techno-sémio-pédagogiques sont définies comme :

des connaissances et des savoir-faire relatifs :

- *aux outils de communication à disposition (forum, wiki, dispositifs de visioconférence, etc.) qui sont les plus appropriés aux objectifs d'une séquence didactique donnée ;*
- *à la prise en compte des modalités appropriées (écrit, oral, vidéo, ou une combinaison) pour telle activité et pour le développement de telles compétences langagières ;*
- *à la gestion pédagogique des activités d'apprentissage avec et autour des outils de CMO [communication médiée/médiatisée par ordinateur] (planification, régulations pendant la tâche, évaluation des apprentissages)*
(Guichon, 2012b : 187)

Comme on peut le voir dans ces lignes, ainsi que dans la plupart des modèles disponibles (Hauck, 2010 ; Dooly, 2016 ; Hampel & Stickler, 2005, 2015), cette

⁴ Faute d'espace, pour une introduction nous renvoyons à O'Rourke (2012) et à Conklin & Pellicier-Sanchez (2018).

compétence inclut non seulement la capacité à utiliser efficacement les modes dans des stratégies de communication et d'enseignement, mais aussi la connaissance⁵ des modes sémiotiques disponibles. Cette connaissance a été étudiée à travers l'introspection rétrospective, notamment à travers les journaux d'apprentissage (Fuchs et al., 2012) et plus rarement par des entretiens d'autoconfrontation (Cohen, 2017), mais jamais dans l'(inter)action elle-même.

Dans ce cadre, notre hypothèse est que si l'apprenti-tuteur/ice établit une relation avec l'environnement dans un cycle d'action-perception-interprétation (Van Lier, 2004), alors le « balayage oculaire » de l'espace sémiotique multidimensionnel de l'écran peut nous informer sur la connaissance/conscience qu'il/elle a des éléments des interfaces et sur certaines opérations cognitives liées à l'attention à des éléments multimodaux de l'interaction. Autrement dit, l'oculométrie permettrait d'ajouter des données importantes concernant la perception visuelle dans la visioconférence, données à analyser en les croisant différents cadres méthodologiques existants, comme, pour nous, celui de l'approche interactionniste multimodale (Develotte et al. 2011 ; Guichon & Tellier, 2017).

Cette hypothèse étant donnée, dans la partie suivante de l'article nous présentons une étude exploratoire visant à tester la viabilité.

3. Un exemple d'analyse

3.1. Contexte

Le contexte que nous prenons en considération est celui d'un dispositif de télécollaboration basé sur le modèle du *Français en (Première) Ligne* par visioconférence (Develotte et al., 2008) pour l'apprentissage du chinois langue étrangère. Dans ce modèle, de futurs enseignants d'une langue élaborent des tâches d'apprentissage qui sont ensuite administrées en ligne à des apprenants. Dans notre cas, le dispositif incluait 6 apprentis enseignants de chinois langue étrangère de la Hong Kong Polytechnic University, inscrits en première année d'un *MA Teaching Chinese as a Foreign Language*. Ces étudiants sont principalement en provenance de la Chine continentale et de Hong Kong et ne parlent pas français. L'autre groupe est constitué de 12 étudiants en 2^e année de Licence en Langues Littératures et Civilisation Etrangères et Régionales, parcours chinois, à Aix Marseille Université. Le niveau en chinois de cette population oscille entre le A2 et le B1 selon le *Cadre Européen Commun de Référence pour les Langues* (Conseil de l'Europe, 2001).

Chaque tuteur chinois a rencontré par Skype à six reprises, pendant environ une heure à chaque fois, un binôme d'apprenants français pendant le deuxième semestre de l'année universitaire 2016-17. Les thèmes des conversations étaient établis par les tuteurs en fonction d'une analyse des besoins du binôme d'interlocuteurs et en essayant de prendre en compte la progression du manuel utilisé dans la Licence de chinois. Chaque groupe choisissait le créneau de chaque séance et le lieu où la faire, sauf pour ceux qui ont participé à la recherche, qui étaient contraints de se rendre dans une salle de l'université équipée avec du matériel d'oculométrie.

⁵ Hauck (2010) parle de *awareness*.

3.2. Données et méthode

3.2.1. Recueil des données et pré-traitement

Les enregistrements ont été effectués du côté des tuteurs chinois. Un formulaire de consentement éclairé a été présenté et fait signer tant aux tuteurs (en anglais) qu'aux apprenants (en français). Le dispositif de recueil des données comprenait également des questionnaires après les échanges. Trois tuteurs ont enregistré un nombre variable de séances. Pour cet article, nous nous concentrons sur une séance d'une tutrice avec deux apprenantes. La séance dure 54 minutes.

Les données d'oculométrie ont été enregistrées avec du matériel Tobii X300 paramétré pour une captation à 120Hz. Concrètement, ce matériel se présente comme un écran d'ordinateur équipé d'une barre pour la captation infrarouge sur le bas. La capture d'écran dynamique et l'enregistrement audio de la session ont été réalisés en se basant sur le protocole de recueil de Cappellini (2014) affiné en concertation avec le Centre d'Expérimentation sur la Parole (CEP) du Laboratoire Parole & Langage.

Pour le premier traitement des données, nous avons utilisé le logiciel Tobii Studio afin d'établir des zones d'intérêt spécifique dynamiques. Cela nous a permis de comptabiliser automatiquement les moments et le nombre de fois où la tutrice a regardé soit le visage de l'une ou l'autre des apprenantes, soit son image dans l'interface Skype, soit l'espace de clavardage. Pendant la séance, il n'y a pas eu d'ouverture de navigateur internet. Grâce au soutien du Centre de Ressources Expérimentales du Labex *Brain and Language Research Institute*, les données relatives aux fixations dans les zones d'intérêt spécifique ont été exportées et importées dans le Eudyco Linguistic Annotator (ELAN) développé par le Max Planck Institute (Sloetjes & Wittenburg, 2008), outil qui a servi de base pour les analyses.

Enfin, la session a été transcrite et alignée dans ELAN en utilisant la convention de transcription ICOR (Groupe ICOR, 2006) telle qu'elle a été adaptée par Cappellini (2014) au chinois langue étrangère. Dans ELAN, nous avons ouvert une ligne (*tier*) pour le verbal de chacune des interlocutrices et pour l'écrit de clavardage des deux côtés de l'interaction. Les autres lignes d'annotation sont présentées ci-dessous dans le cadre des méthodes d'analyse.

3.2.2. Méthodes d'analyses

Pour les analyses, en suivant Cappellini (2016), nous nous sommes concentrés sur les actions et la perception de la tutrice pendant différents types de séquences latérales d'étayage, dont nous prenons en compte dans cet article les séquences potentiellement acquisitionnelles lexicales (De Pietro et al., 1989 : dorénavant SPAlex). Très brièvement, dans une SPAlex l'apprenant, pendant son tour de parole, fait face à une panne lexicale et sollicite, implicitement ou explicitement, l'aide d'un interlocuteur pour la résoudre. Suite à la phase de sollicitation, dans la phase de production de donnée, une ou plusieurs solutions sont proposées. Enfin, une phase, facultative, de (re)prise consiste en la répétition de l'item lexical, voire à son introduction dans l'énoncé de départ⁶.

⁶ Voir Matthey (1996) pour les différents types de prise.

Dans les analyses, nous avons adopté le cadre général d'une double analyse comme dans l'étude de Cappellini et Azaoui (2017). Dans un premier temps, chacun des deux auteurs a procédé indépendamment à l'identification et à la description multimodale des procédés à l'œuvre dans les instances de SPAlex. Pour ce faire, nous avons employé le schéma d'annotation de Wigham (2017) avec quelques aménagements. Ainsi, outre aux lignes pour le verbal (ci-dessus), nous avons ouvert une ligne (*tier*) pour chaque interlocutrice concernant :

- les gestes. A la typologie de McNeill (1992) enrichie des emblèmes (Kendon, 1982) utilisée par Wigham (2017), nous avons ajouté les gestes interactionnels de Bavelas et al. (1995) typiques de la visioconférence (Cappellini, 2014b ; Cappellini & Azaoui, 2017) ;
- les mouvements relatifs à la proxémique, notamment par rapport à la distance de l'écran/webcam ;
- les mouvements de la tête (rotatif, latéral, sagittal) ;
- des actions.

Nous n'avons pas annoté le regard, puisque nous avons intégré les données d'oculométrie à cet effet. Ensuite, nous avons comparé nos analyses et discuté les cas de désaccord (16 discussions sur 35 séquences latérales identifiées⁷), ce qui nous a permis d'une part d'affiner nos critères d'identification et de nous accorder pour les analyses suivantes.

Dans un deuxième temps, nous avons formulé indépendamment de manière inductive des observations concernant les procédés récurrents et leur possible signification. Ensuite, nous avons à nouveau comparé nos analyses. Dans ce cas, la comparaison ne visait pas la résolution de désaccords éventuels, mais l'enrichissement mutuel des analyses.

3.3. Analyses

3.3.1. Analyse globale

Du point de vue quantitatif, l'analyse par ELAN a donné lieu à la production de 1442 annotations manuelles concernant la multimodalité des échanges, dont 1312 portant sur les tours de parole. L'analyse a mis en lumière 35 séquences latérales, dont 16 SPAlex. L'oculométrie a permis d'annoter 10 979 fixations dans les zones d'intérêt spécifique de la part de la tutrice au cours de la séance, distribuées dans les quatre zones comme il suit :

	Nombre de fixations
Visage étudiante 1	3450
Visage étudiante 2	5951
Encadré tutrice	1393
Clavardage	185

Tableau 1 Distribution des fixations dans les Zones d'Intérêt Spécifique

⁷ Parmi ces 35 séquences, 9 sont des séquences conversationnelles explicatives (Gulich, 1990) et 10 des séquences d'évaluation normative (Py, 2000), que nous ne prenons pas en compte dans cet article.

Le nombre de fixations sur les images des trois interlocutrices est à mettre en relation avec le nombre de tours de parole de chacune. Ainsi, il n'est pas étonnant de noter que la tutrice regarde davantage l'étudiante 2 que l'étudiante 1, puisque l'étudiante 2 prend la parole 587 fois contre 264 pour l'étudiante 1. Ce qui retient notre attention, c'est le nombre assez élevé de fixations de la tutrice sur sa propre image, correspondant à environ 12,7 % des fixations. Nous verrons plus loin à quels moments cela correspond dans les SPAlex.

3.3.2. Séquences potentiellement acquisitionnelles lexicales

En ce qui concerne les 16 SPAlex identifiées, l'exemple suivant présente plusieurs caractéristiques récurrentes des procédés employés. Nous reproduisons la transcription du verbal et renvoyons le lecteur au visionnage de la vidéo pour la multimodalité⁸. La ligne 'TC' donne une indication temporelle relative à l'annotation permettant d'apprécier les chevauchements des tours de parole. Une traduction en français est fournie entre parenthèses. L'image suivante permet d'avoir un aperçu de l'annotation dans ELAN.

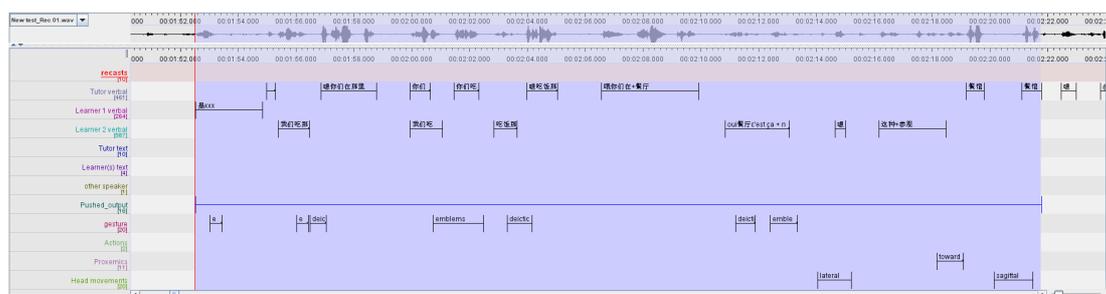


Image 1 Annotation d'une SPAlex dans ELAN

Cette séquence latérale arrive en début de conversation. Les interlocutrices françaises se trouvent dans un restaurant universitaire et elles essaient de l'expliquer à la tutrice chinoise.

Learner 2 verbal 我们吃那里 (où nous mangeons)

TC 00:01:55.380 - 00:01:56.460

Tutor verbal 嗯你们在+那里++你们在- (euh vous êtes + là ++ vous êtes -)

TC 00:01:56.870 - 00:02:00.640

Learner 2 verbal 我们吃 (nous mangeons)

TC 00:01:59.940 - 00:02:01.075

Tutor verbal 你们吃- (vous mangez -)

TC 00:02:01.475 - 00:02:02.325

⁸ Vidéo accessible à l'adresse <https://amupod.univ-amu.fr/video/2432-telecollaboration-polyu-amu-mars-2017/>

Learner 2 verbal 吃饭那里 (l'endroit pour manger)

TC 00:02:02.840 - 00:02:03.640

Tutor verbal 嗯吃饭那里 (euh l'endroit pour manger)

TC 00:02:03.985 - 00:02:05.055

Tutor verbal 哦你们在+餐厅 (oh vous êtes + au restaurant)

TC 00:02:06.565 - 00:02:09.935

Learner 2 verbal oui 餐厅 (restaurant) *c'est ça + non non non*

TC 00:02:10.845 - 00:02:13.065

Learner 2 verbal 嗯 (euh)

TC 00:02:14.650 - 00:02:15.010

Learner 2 verbal 这种+餐馆 (ce type + restaurant⁹)

TC 00:02:16.175 - 00:02:18.485

Tutor verbal 餐馆 (restaurant)

TC 00:02:19.185 - 00:02:19.815

Head movements sagittal

TC 00:02:20.154 - 00:02:21.494

Tutor verbal 餐馆 (restaurant)

TC 00:02:21.105 - 00:02:21.785

En début de séquence, l'étudiante 1 semble avoir l'intention de prendre la parole mais elle s'arrête. On entrevoit aussi un geste qui semblerait être un emblème (Kendon, 1982) pour « manger ». L'étudiante 1 se désiste et ensuite se tourne vers l'étudiante 2, avec un aparté inaudible dans l'enregistrement. Il est intéressant de noter que si dans un premier temps le regard de la tutrice est fixé sur le visage de l'étudiante 1, lorsque cette dernière se tourne vers l'étudiante 2, le regard de la tutrice suit la même trajectoire et va se poser sur le visage de l'étudiante 2. L'étudiante 2 se penche ensuite en avant, et en revenant elle énonce « 我们吃那里 (où nous mangeons) », en accompagnant l'énoncé de deux gestes, un premier emblème pour « manger » en concomitance de « nous mangeons » et un deuxième geste, déictique, avec l'index de la main droite pointé vers le bas, pour signifier « ici » en concomitance de « où » (dans cette construction, le chinois positionne l'adverbe de lieu en fin de phrase). La tutrice ne focalise pas son regard sur ces gestes.

Suite à cet énoncé, la tutrice reprend la parole et reformule, probablement en croyant répéter, en disant « 嗯你们在+那里++你们在- (euh vous êtes + là ++ vous êtes -) ». Le tempo de la formulation laisserait penser à une correction de la prononciation, puisque la tutrice reprend la même structure de la phrase et fait une

⁹ Le mot 餐馆 est synonyme de 餐厅, la seule différence étant le registre de langue, le premier étant plus oral et le deuxième plus formel.

brève pause intra-tour après le 在¹⁰, qui remplace le mot 吃 (manger). En ce qui concerne le regard, la tutrice opère trois fixations sur sa propre image en concomitances du segment « 在+那里 ». Il s'ensuit ce qui semblerait être un point de transition potentiel, où la tutrice laisse l'espace pour une reprise de parole de la part des apprenantes. Faute de prise de parole, on note deux actions simultanées : la première est que les apprenantes, qui avaient les yeux baissés, modifient la direction de leur regard et le pointent vers leur écran. En même temps, la tutrice reprend la parole pour répéter son énoncé. Pendant son tour de parole, l'étudiante 2 opère un chevauchement en répétant 吃, en se penchant vers l'écran et en reproduisant l'emblème pour « manger » de manière plus ostentatoire. Dans la suite, l'incompréhension se dissipe et nous assistons à nouveau à une répétition de ce que les apprenantes disent, de la part de la tutrice, avec des fixations sur l'encadré contenant sa propre image.

Une fois la phase de sollicitation/intercompréhension (Cappellini & Pescheux, 2015) résolue, la tutrice propose une donnée : « 哦你们在+餐厅 (oh vous êtes + au restaurant) ». Cette donnée est initialement acceptée par l'étudiante 2, qui ensuite revient sur sa décision et, en combinant une mimique avec des mouvements latéraux de la tête et des mouvements de la main seulement partiellement visibles, montre des hésitations, signalées également ensuite par le paraverbal « euh ». L'étudiante 2 consulte ensuite son portable, probablement ouvert sur un dictionnaire, et elle propose « 餐馆 » à deux reprises, suivie de l'étudiante 1. La tutrice répète alors le mot et les apprenantes font oui de la tête, ce qui clôt la séquence.

Au-delà du fait que cet exemple est ce que nous avons appelé une SPAlex non prototypique (Cappellini & Pescheux, 2015), cet extrait montre plusieurs éléments récurrents. Le premier élément est le fait que dans ces séquences, le regard de la tutrice semble suivre la prise de parole, y compris quand c'est elle qui la prend pour offrir une donnée ou pour répéter les propositions des apprenants. Le deuxième élément, est la présence constante d'une proxémique de l'écran qui revient à des gestes interactionnels pour prendre et donner la parole, surtout de la part des apprenantes. Un troisième élément est la présence importante de gestes de la part des apprenantes, mais sans qu'il y ait fixation (avec une seule exception). Les gestes des apprenantes ne rentrent donc pas dans la focale, tout en étant perçus par une vision périphérique. Enfin, dans les phases de clôture, il y a peu de reprises des données mais plutôt des acquiescements des apprenantes. La proxémique de l'écran souligne aussi les fins de séquence latérale, avec les apprenantes s'éloignant de l'écran.

4. Conclusions

Cet article s'est proposé de prendre en compte le rôle que l'oculométrie peut jouer dans le cadre d'une approche écologique. En ce qui concerne l'étude exploratoire de la réalisation des SPAlex pendant une session de tutorat en ligne par visioconférence, au-delà de phénomènes déjà observés, évoqués ci-dessus, dans la littérature adoptant une approche multimodale de ce type d'activité pédagogique (voir le recueil de

¹⁰ En chinois, le positionnement statique dans l'espace se produit avec le mot 在, qui peut être traduit en français par 'être', au sens d'être quelque part.

Guichon & Tellier, 2017), l'emploi de l'oculométrie a permis de mettre en lumière principalement deux phénomènes. Le premier est qu'en dépit d'une production gestuelle assez importante de la part des apprenantes, ces gestes ne font, à une exception près, jamais l'objet de fixations, ce qui nous amène à affirmer que les gestes sont perçus dans l'attention périphérique. Le deuxième phénomène est la fixation quasi systématique de la tutrice sur sa propre image lors de la phase de donnée des SPAlex. Si ce deuxième phénomène devait s'avérer de portée générale et non une caractéristique propre à cette tutrice, il soulèverait tant des questions théoriques – pourquoi le tuteur se regarde-t-il pendant qu'il propose une donnée ? – que des pistes praxéologiques pour la formation de formateurs – comment pourrait-on utiliser la focalisation sur sa propre image pendant l'interaction pour accompagner la conscientisation de ses stratégies ?

Les observations faites dans l'analyse nous amènent à formuler des hypothèses concernant les compétences techno-sémio-pédagogiques de l'apprenti-tutrice. Pendant les échanges, la tutrice arrive à gérer les tours de parole ainsi que la phase de donnée en s'appuyant sur les modes verbal, proxémique et mimo-gestuel, considérés comme des complexes de signes (Bezemer & Kress, 2016). La tutrice montre en particulier une compétence techno-sémio-pédagogique par l'utilisation de ces indices interactionnels dans la décision de produire une reformulation à la fin du tour de parole de l'apprenante, donc sans l'interrompre. En outre, nous sommes interpellés par le fait que la tutrice focalise son regard sur sa propre image lors des reformulations, ce qui pourrait signifier une attention portée à la manière de se présenter dans l'interaction. Reste à vérifier le caractère conscient, voire stratégique, de ces dynamiques. Ainsi, dans des études à venir, nous pourrions interroger les tuteurs premièrement pour vérifier s'ils sont conscients de poser leur regard sur leur propre image (si cela se vérifiait) à des moments spécifiques des séquences latérales, d'autre part, en cas de réponse positive, pour vérifier si cette dynamique est la manifestation d'une stratégie où s'il s'agit d'un procédé interactionnel sans but pédagogique.

Au-delà des résultats de l'étude exploratoire, qui nécessitent une validation par une analyse de corpus d'analyse plus larges, du point de vue méthodologique et épistémologique nous espérons avoir réussi à montrer comment l'oculométrie, qui est largement mobilisée dans des paradigmes expérimentaux, peut être mobilisée dans un paradigme socioculturel, notamment celui écologique. De ce point de vue, outre l'articulation conceptuelle, cet article, bien que par une étude exploratoire, a montré la validité pratique de notre hypothèse.

Parmi les pistes d'évolution possible, outre le déploiement de cette méthodologie sur des corpus plus importants, nous comptons nous appuyer sur la technologie du *face-tracking* afin de rendre automatique l'identification de zones d'intérêt spécifique concernant les visages des interlocuteurs, ce qui éviterait le travail chronophage et fastidieux de modification manuelle de ces zones tout au long de l'enregistrement. Enfin, nous comptons croiser cette méthodologie avec celle de l'autoconfrontation (Stickler & Shi, 2016) ainsi qu'avec des post-tests pour évaluer si la fixation sur certains éléments de la multimodalité de la visioconférence joue un rôle sur l'acquisition à court terme et à moyen terme d'éléments lexicaux.

5. Remerciements

Ce travail n'aurait pas été possible sans l'aide précieuse d'Antonio Serrato et d'Alain Ghio du Centre d'Expérimentation sur la Parole du Laboratoire Parole &

Langage (UMR 7309), de Christelle Zielinski du Centre de Recherches Expérimentales du Labex *Brain and Language Research Institute*, et de Fabrice Cauchard de la plateforme H2C2 de la Maison de la Recherche de l'UFR ALLSH à Aix-Marseille Université. Merci également aux étudiant.e.s ayant accepté de participer à cette recherche. Qu'ils/elles soient tou.te.s ici chaleureusement remercié.e.s.

6. Bibliographie

- Bateson, G. (1977) [1972]. *Vers une écologie de l'esprit 1*. Paris : Seuil.
- Bateson, G. (1980) [1972]. *Vers une écologie de l'esprit 2*. Paris : Seuil.
- Bavelas, J., Chovil, N., Coates, L. & Roe, L. (1995). Gestures specialized for dialogue. *Personality and social psychology bulletin*, 21(4), 394-405
- Bezemer, J. & Kress, G. (2016). *Multimodality, Learning and Communication: A social semiotic frame*. Londres : Routledge.
- Blin, F. (2016a). Towards an 'ecological' CALL theory: theoretical perspectives and their instantiation in CALL research and practice. In F. Farr & L. Murray (eds.). *The Routledge handbook of language learning and technology* (pp. 39-54). Londres : Routledge.
- Blin, F. (2016b). The theory of affordances. In C. Caws & M.-J. Hamel (eds.). *Language-Learner computer interactions* (pp. 41-64). Amsterdam : John Benjamins.
- Bronfenbrenner, U. (1979). *The Ecology of Human Development. Experiments by nature and design*. Cambridge MA : Harvard University Press.
- Cappellini, M. (2013). Co-construction des routines d'étayage dans un tandem franco-chinois par visioconférence. In C. Dejean-Thircuir, F. Mangenot, E. Nissen, T. Soubrié, (eds.). *Actes du colloque Echanger Pour Apprendre en Ligne 2013*. En ligne : http://epal.u-grenoble3.fr/dossier/06_act/pdf/epal2013-cappellini.pdf
- Cappellini, M. (2014). *Modélisation systémique des étayages dans un environnement télé-tandem pour le français et le chinois langues étrangères. Une étude interactionniste et écologique du soutien au développement de la compétence de communication*. Université Lille 3 SHS. Unpublished PhD thesis. <https://hal.archives-ouvertes.fr/tel-01392190>
- Cappellini, M. (2014b). Affordances et positionnements d'expertise culturelle dans un tandem par visioconférence. In I. Colon de Carvajal & M. Ollagnier-Beldame (eds.). *Actes du colloque Interactions Multimodales Par Ecran*. En ligne : <http://impec.ens-lyon.fr/actes-du-colloque-impec-2014-en-ligne-208378.kjsp>
- Cappellini, M. (2016). Roles and scaffolding in teletandem interactions - A study of the relations between the sociocultural and the language learning dimensions in a French-Chinese teletandem. *Innovation in Language Learning and Teaching*, 10(1), 6-20.
- Cappellini, M. & Azaoui, B. (2017) Sequences of normative evaluation in different pedagogical settings through desktop videoconference. *Language Learning in Higher Education*, 7(1), 55-80. <https://doi.org/10.1515/cercles-2017-0002>
- Cappellini, M. & Combe, C. (2017). Analyser des compétences techno-sémio-pédagogiques d'apprentis tuteurs dans différents environnements numériques : résultats d'une étude exploratoire. *Alsic*, 20. DOI : 10.4000/alsic.3186
- Cappellini, M. & Pescheux, M. (2015). La gestion des normes langagières dans un tandem franco-chinois par visioconférence. *Le bulletin suisse de linguistique*

- appliquée VALS ASLA, numéro spécial, tome 3*, pp. 171-187.
http://doc.rero.ch/record/11876/files/bulletin_vals_asla_2015_special_3.pdf
- Cappellini, M. & Rivens Mompean, A. (2015). Role taking for the teletandem pair involved in multimodal online conversation: some considerations from data to develop counselling practices. *Language Learning in Higher Education*, 5(1), 243-264.
- Cohen, C. (2017). Former à l'enseignement en ligne. In N. Guichon & M. Tellier (eds.). *Enseigner l'oral en ligne : Une perspective multimodale* (pp. 218-242). Paris : Didier.
- Conklin, K. & Pellicier-Sanchez, A. (2018). *Eye-tracking: a guide for applied linguistics research*. Cambridge : Cambridge University Press.
- Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues*, Paris : Didier.
- De Pietro, J.-F., Matthey, M. & Py, B. (1989). Acquisition et contrat didactique : les séquences potentiellement acquisitionnelles dans la conversation exolingue. In D. Weil & H. Fugier (eds.). *Actes du troisième colloque régional de linguistique* (pp. 99-124). Strasbourg : Université des sciences humaines et Université Louis Pasteur.
- Develotte, C., Guichon, N. & Kern, R. (2008). « Allo Berkeley ? Ici Lyon... Vous nous voyez bien ? » Etude d'un dispositif de formation en ligne synchrone franco-américain à travers le discours de ses usagers. *Alsic*, 11(2), 129-156.
- Develotte, C., Kern, R. & Lamy, M.-N. (eds.) (2011). *Décrire la conversation en ligne. Le face à face distanciel*. Lyon : ENS Editions.
- Dooley, M. (2016). 'Please remove your avatar from my personal space': Competences of the telecollaborative efficient person. In R. O'Dowd & T. Lewis (eds.). *Online Intercultural Exchange. Policy, pedagogy, practice* (pp. 192-208). London : Routledge.
- Engestrom, Y. (1988). The cultural-historical theory of activity and the study of political repression. *International journal of mental health*, 17(4), 29-41.
- Fuchs, C., Hauck, M. & Müller-Hartmann, A. (2012). Promoting learner autonomy through multiliteracy skills development in cross-institutional exchanges. *Language Learning & Technology*, 16(3), 82-102.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Londres : Lawrence Erlbaum Associates.
- Groupe ICOR (2006). *Convention ICOR*. En ligne : http://icar.univ-lyon2.fr/projets/corinte/documents/2013_Conv_ICOR_250313.pdf
- Guichon, N. (2012). *Vers l'intégration des TIC dans l'enseignement des langues*. Paris : Didier.
- Guichon, N. & Tellier M. (eds) (2017). *Enseigner à l'oral en ligne: une approche multimodale des interactions pédagogiques*. Paris : Didier.
- Gulich, E. (1990). Pour une ethnométhodologie linguistique. Description des séquences conversationnelles explicatives. In M. Charolles, S. Fisher & J. Jayez (éds.). *Le discours. Représentations et interprétations* (pp. 71-109). Nancy : Presses Universitaires de Nancy.
- Hampel, R. & Stickler, U. (2005). New skills for new classrooms: training tutors to teach languages online. *Computer-Assisted Language Learning*, 18(4), 311-326.
- Hampel, R. & Stickler, U. (2012). The use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL Journal*, 24(2), 116-137.

- Hampel R. & Stickler, U. (2015) (eds). *Developing online language teaching*. New York : Palgrave Macmillan.
- Hauck, M (2010). Telecollaboration: At the interface between multimedial and intercultural communicative competence. In S. Guth & F. Helm (eds.). *Telecollaboration 2.0* (pp. 219-244). Oxford : Peter Lang.
- Jefferson, G. (1972). Side Sequences. In D. Sudnow (éd.). *Studies in Social Interaction* (pp. 294-338). New York : The Free Press.
- Jewitt, C., Bezemer, J. & O'Hallaran, K. (2016). *Introducing multimodality*. London : Routledge.
- Kendon, A. (1982). Coordination of action and framing in face-to-face interaction. *Interaction rhythms: Periodicity in communicative behavior*, 351-363.
- Larsen-Freeman, D. & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford : Oxford University Press.
- Liddicoat, A. J. (2011). Enacting participation: hybrid modalities in online video conversation. In C. Develotte, R. Kern & M.-N. Lamy (eds.). *Décrire la conversation en ligne. Le face à face distanciel* (pp. 51-70). Lyon : ENS Editions.
- Mangenot, F. (2013). Les échanges en ligne comme secteur de pratiques et de recherches en ALAO : quelles problématiques, quelles évolutions ? *Les Cahiers de l'ILOB*, 5, 3-21.
- Matthey, M. (1996). *Apprentissage d'une langue et interaction verbale. Sollicitation, transmission et construction des connaissances linguistiques en situation exolingue*. Bern : Peter Lang.
- McNeill, D. (1992). *Hand & Mind. What gesture reveals about thought*. Chicago : University of Chicago Press.
- Michel, M. & Smith, B. (2018). Measuring lexical alignment during L2 chat interaction: An eye-tracking study. In S. M. Gass, P. Spinner & J. Behney (eds.). *Salience in second language acquisition* (pp. 244-268). London : Routledge.
- Mondada, L. (1999). L'accomplissement de l'« étrangeté » dans et par l'interaction : procédures de catégorisations des locuteurs. *Langages*, 134, 20-34.
- O'Rourke, B. (2012). Using eye-tracking to investigate gaze behavior in synchronous computer-mediated communication for language learning. In M. Dooly & R. O'Dowd (eds.). *Researching Online Foreign Language Interaction and Exchange. Theories, Methods and Challenges* (pp. 305-342). Bern : Peter Lang.
- Py, B. (2000). La construction interactive de la norme comme pratique et comme représentation. *Aile*, 12, 77-97.
- Sacks, H. (1972). An Initial Investigation of the Usability of Conversational Data for Doing Sociology. In D. Sudnow (ed.). *Studies in Social Interaction* (pp. 31-74). New York : The Free Press.
- Satar, M.H. (2010). *Social presence in online multimodal communication: A framework to analyse online interactions between language learners*. Thèse de doctorat soutenue à l'Open University (Royaume Uni). Inédite.
- Shi, L, Stickler, U. & Lloyd, M. E. (2017). The interplay between attention, experience and skills in online language teaching. *Language Learning in Higher Education*, 7(1), 205-238.
- Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. www.mpi.nl/publications/escidoc-60774

- Sperber, D. & Wilson, D. (1986). *Relevance. Communication and Cognition*. Oxford : Blackwell Publishing.
- Stickler, U. & Shi, L. (2017). Eyetracking methodology in SCMC: A new tool for empowering learning and teaching. *ReCALL*, 29(2), 160-177.
- Thibault, P. J. & Baldry, A. (2006). *Multimodal transcription and text analysis*. Sheffield : Equinox Publishing.
- Van Lier, L. (2004). *The Ecology and Semiotics of Language Learning: a Sociocultural Perspective*. Dordrecht : Kluwer Academic Publishers.
- Wigham, C. (2017). A multimodal analysis of lexical explanation sequences in webconferencing-supported language teaching. *Language Learning in Higher Education*, 7(1). 81-108.
- Yamada, M. & Akahori, K. (2009). Awareness and performance through self- and partner's image in videoconferencing. *Calico Journal*, 27(1), 1-25.